UNIVERSITY
OF SOUTHERN
QUEENSLAND
AUSTRALIA

DISCRIMINATION OF WHEAT CROWN ROT UTILISING WAVELET

BASED MODELS IN THE NIR SPECTRUM

A Thesis submitted by

Jacob Humpal, BSc

For the award of

Doctor of Philosophy

2020

ABSTRACT

Crown rot is a stubble-borne disease of significant economic importance to the wheat industry, with no rapid detection method and human scoring only possible late in disease development after visual symptoms appear. This research aimed to develop non-destructive plant sensing tools to improve crown rot screening and accelerate development of resistant and tolerant germplasm.

Machine learning models were developed for the discrimination and quantification of *Fusarium pseudograminearum* induced crown rot from contact near-infrared sensor data (900–1700 nm) in three glasshouse and two field trials, and near-infrared camera data (900–1700 nm) in one glasshouse trial. Contact sensor data was modeled with principal component analysis (PCA) and the discrete wavelet transform (DWT), and the impact of sensing location (i.e. plant part), timing of sensing and different training splits were compared. Specific models were generated by grouping sensor data into weekly intervals for each trial. Generalised models were generated by combining data across multiple trial sites and temporal windows (weekly, three-weekly and whole-season).

DWT achieved higher crown rot detection accuracy than PCA in 67% of test cases for contact sensing when training on 20% of the disease data for specific models, with maximum average accuracies of 70.5%. Both DWT and PCA performed best in 50% of test cases when trained on the 80-20% split, with maximum accuracies of 75.7% for DWT and 76.9% for PCA. PCA was more accurate than DWT in a majority of individual test cases across both data splits for generalised models, with maximum accuracies of 69.8%. PCA utilised fewer overall features successfully in both specific and generalised models on both training splits.

Significant differences in accuracy between sensing dates, sensing locations and interaction were determined. The differences were examined using DWT and six machine learning methods for generalised models, with an 80-20% train-test split. Individual readings and combinations of tiller, centre and flag leaf were evaluated. The highest performing combination of measurements

achieved an average of 77.9% accuracy across all five trials using the contact sensor.

Phenotyping capability was examined using a multilayer perceptron preprocessed with DWT and PCA to develop models to quantify disease severity. Six severity scales based on human scoring of visual symptoms were developed and tested on contact sensor data from glasshouse and field trials. Both PCA and DWT performed similarly with individual models obtaining mean model accuracy ranging from 32% to 96%. Models performed best in field environments.

Near-infrared image data was collected in a glasshouse trial across four weeks of early crown rot infection, using narrow bandpass filters centered at five wavebands identified using the contact sensor. Single and multi-input convolutional neural networks were created for discrimination and quantification of crown rot infection. The discrimination model achieved average accuracies of 53–100%, with highest average accuracies obtained in weeks 2–4. The quantification model achieved average accuracies of 73% when trained on combined data across all weeks. Developed models were successfully ported onto a mobile development board for real-time detection applications.

It is concluded that successful detection and quantification of crown rot was achieved using both contact and camera-based near-infrared sensing. These findings are the initial steps in developing a high-throughput phenotyping system to provide wheat breeders with new tools and methods for crown rot resistance breeding. Further work should evaluate developed models on a wide range of germplasm and extend to real-time model execution.

CERTIFICATION OF THESIS

This thesis is the work of Jacob Humpal except where otherwise acknowledged, with the majority of the authorship of the papers presented as a Thesis by Publication undertaken by the Student. The work is original and has not previously been submitted for any other award, except where acknowledged.

Principal Supervisor: Dr. Cheryl McCarthy

Associate Supervisor: Dr. Cassandra Percy

Associate Supervisor: Prof. J. Alex Thomasson

Student and supervisors' signatures of endorsement are held at the University.

STATEMENT OF CONTRIBUTION

The following detail is the agreed share of contribution for candidate and co-authors in the presented publications in this thesis:

Chapter 3: Machine learning for the detection of crown rot, *Fusarium pseudograminearum*, Part A: PCA and discrete wavelet approaches. **Humpal J**., McCarthy C., Percy C , & Thomasson AJ. (2020a). To be submitted to **Computers and Electronics in Agriculture.**

Jacob Humpal (JH) contributed 70% towards writing and drafting the manuscript, data collection, model development and analysis. Cassandra Percy (CP) contributed 15% towards trial establishment and inoculum preparation, concept and design, revision of the manuscript and final editorial input. Cheryl McCarthy (CM) contributed 10% towards the concept, drafting and final important editorial input. J. Alex Thomasson (JAT) contributing 5% towards the concept, revision of the manuscript and final important editorial input.

Chapter 4: Machine learning for the detection of crown rot, *Fusarium pseudograminearum*, Part B: generalised temporal models for PCA and discrete wavelet approaches. **Humpal J**., McCarthy C., Percy C., & Thomasson AJ. (2020b). To be submitted to **Computers and Electronics in Agriculture.**

JH contributed 70% towards writing and drafting the manuscript, data collection, model development and analysis. CP contributed 15% towards trial establishment and inoculum preparation, concept and design, revision of the manuscript and final editorial input. CM contributed 10% towards the concept, drafting and final important editorial input. JAT contributing 5% towards the concept, revision of the manuscript and final important editorial input.

Chapter 5: Machine learning for the detection of crown rot, *Fusarium pseudograminearum*, Part C: sensing location and timing. **Humpal J**., McCarthy C., Percy C., & Thomasson AJ. (2020c). To be submitted to **Computers and Electronics in Agriculture.**

JH contributed 70% towards writing and drafting the manuscript, data collection, model development and analysis. CP contributed 15% towards trial establishment and inoculum preparation, concept and design, revision of the manuscript and final editorial input. CM contributed 10% towards the concept, drafting and final important

editorial input. JAT contributing 5% towards the concept, revision of the manuscript and final important editorial input.

Chapter 6: Development of multilayer perceptron models for high-throughput phenotyping of crown rot, *Fusarium pseudograminearum*. **Humpal J**., Percy C., McCarthy C., & Thomasson AJ. (2020d). To be submitted to **Computers and Electronics in Agriculture.**

JH contributed 70% towards writing and drafting the manuscript, data collection, model development and analysis. CP contributed 15% towards trial establishment and inoculum preparation, concept and design, revision of the manuscript and final editorial input. CM contributed 10% towards the concept, drafting and final important editorial input. JAT contributing 5% towards the concept, revision of the manuscript and final important editorial input.

Chapter 7: CNNs for crown rot discrimination in wheat, Part A: detection. **Humpal J**., McCarthy C., Percy C., & Thomasson AJ. (2020e). To be submitted to **Computers and Electronics in Agriculture.**

JH contributed 70% towards writing and drafting the manuscript, data collection, model development and analysis. CP contributed 15% towards trial establishment and inoculum preparation, concept and design, revision of the manuscript and final editorial input. CM contributed 10% towards the concept, drafting and final important editorial input. JAT contributing 5% towards the concept, revision of the manuscript and final important editorial input.

Chapter 8: CNNs for crown rot discrimination in wheat, Part B: quantification. **Humpal J**., Percy C., McCarthy C., & Thomasson AJ. (2020f). To be submitted to **Computers and Electronics in Agriculture.**

JH contributed 70% towards writing and drafting the manuscript, data collection, model development and analysis. CP contributed 15% towards trial establishment and inoculum preparation, concept and design, revision of the manuscript and final editorial input. CM contributed 10% towards the concept, drafting and final important editorial input. JAT contributing 5% towards the concept, revision of the manuscript and final important editorial input.

ACKNOWLEDGMENT

LIST OF CONFERENCES ATTENDED AND CONFERENCE PAPERS

Humpal, J, McCarthy, C, Percy, C & Thomasson, JA 2018, 'Machine vision learning for crop disease and quality parameters'. 21st Symposium on Precision Agriculture in Australia, Adelaide, Australia, September 2018. (Speaker)

Humpal, J, McCarthy, C, Percy, C & Thomasson, JA 2019, 'Towards real-time crown rot disease discrimination by NIR models'. APPS Seminar Series, Brisbane, Australia, March 2019. (Speaker)

Humpal, J, McCarthy, C, Percy, C & Thomasson, JA 2020, 'NIR sensing and machine learning to rapidly assess crown rot in wheat'. GRDC Updates, Goondiwindi, Australia, March 2020. (Speaker)

Humpal, J, McCarthy, C, Percy, C & Thomasson, JA 2020, 'Detection of crown rot in wheat utilising near-infrared spectroscopy: towards remote and robotic sensing.' vol. 11414, SPIE, Online, April 2020. (Speaker).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

LIST OF UNITS, ACRONYMS AND ABBREVIATIONS

ANN: Artificial Neural Network

ANOVA: Analysis of Variance

ARI: Anthocyanin Reflectance Index

ARVI: Atmospherically Resistant Vegetation Index

C: Celsius

C-MOS: Silicon Complementary Metal-Oxide Semiconductor

CAE: Centre for Agricultural Engineering

CCH: Centre for Crop Health

Ch-FI: Chlorophyll Fluorescence Imaging

cm: Centimetre

CNN: Convolutional Neural Network

CR: Crown Rot

CSV: Comma Separated Value

DTC: Decision Tree Classifier

DWT: Discrete Wavelet Transform

Fc: *Fusarium culmorum*

Fg: *Fusarium graminearum*

Fp: *Fusarium pseudograminearum*

GPS: Global Positioning System

GRDC: Grains Research and Development Corporation

HSD: Honestly Significant Difference

HTTP: High-Throughput Phenotyping

InGaAs: Indium Gallium Arsenide

IR: Infrared

IRT: Infrared Thermography

KNN: K-Nearest Neighbor

LDA: Linear Discriminant Analysis

LED: Light Emitting Diode

m: Metre

MLP: Multi-Layer Perceptron

MTD: Maximum Temperature Difference

MTMF: Mixture Tuned Matched Filtering

N: Nitrogen

NDVI: Normalised Difference Vegetation Index

NIR: Near Infrared

nm: Nanometre

NN: Neural Network

PCA: Principal Component Analysis

pix: Pixel

PRI: Photochemical Reflectance Index

QDAF: Queensland Department of Agriculture and Fisheries

RFC: Random Forest Classifier

RGB: Red, Green, Blue

RTP: Research Training Program

SAM: Spectral Angle Mapper

SAVI: Soil-Adjusted Vegetation Index

SE: Standard Error

SFFS: Sequential Floating Forward Method

SIF: Solar Induced Fluorescence

SIPI: Structural Independent Pigment Index

SVC: Support Vector Classifier

SVI: Spectral Vegetation Index

SVM: Support Vector Machine

SWIR: Shortwave Infrared

UAV: Unmanned Aerial Vehicle

USQ: University of Southern Queensland

UV: Ultraviolet

VARI: Visible Atmospherically Resistant Index

VIS: Visible

CHAPTER 1: INTRODUCTION TO CROWN ROT

Crown rot (CR), caused by *Fusarium spp.* is an economically important crop disease in winter cereals, both in Australia and worldwide (Klein et al. 1991; Backhouse & Burgess 2002; Smiley et al. 2005; Alahmad et al. 2018). Increasing adoption of conservation agriculture has increased CR prevalence, due to inoculum preservation in trash from previous seasons (Simpfendorfer et al. 2019). CR is difficult to detect, with assessment relying on correlations with visible symptoms such as whiteheads and stem browning which indicate disease prevalence (Klein et al. 1991). Machine vision systems have advanced rapidly in recent years and have potential to detect and quantify crop stress both early and accurately. Recent literature has shown rapid development in technologies around biotic and abiotic crop stress detection, crop monitoring and weed detection (Long et al. 2016; Pinto et al. 2016; Thomas et al. 2017; Elvanidi et al. 2018; Lottes et al. 2018). New advancements in sensor technology have succeeded in reducing sensor size and cost while increasing sensitivity and resolution (Mahlein et al. 2012; Mahlein 2016; Bogue 2017). Traditional and novel data analysis techniques have been reported, with various approaches in machine learning and techniques adapted from robotics and engineering. Successful applications of these techniques have shown high stress detection accuracy is possible with varying spectral resolutions, showing promise for multi-resolution disease detection. This introduction gives an overview of *Fusarium spp.*, cereal CR and leads into a literature review examining existing machine vision applications in precision agriculture and detailing potential for technology adaptation and adoption for use in CR detection and quantification.

1.1 *FUSARIUM* CROWN ROT

CR is a stubble-borne, fungal disease of cereals caused by a related group of *Fusarium spp.* primarily, *F. pseudograminearum* (*Fp*) and *F. culmorum* (*Fc*) with minor occurrence of *F. graminearum* (*Fg*) in Australia, (Parry et al. 1994). CR has been reported in all wheat growing regions in Australia and internationally, predominantly caused by *Fp* in arid and semi-arid climates, with *Fc* also capable of infection in temperate environments (Leslie & Summerell 2006; Alahmad et al. 2018). Severe CR infection of wheat can lead to greater than 50% yield loss under conditions suitable for disease development in Australia (Klein et al. 1991), and has been reported to limit

grain production in North America, Africa, the middle east and China (Smiley et al. 2005; Lamprecht et al. 2006; Saremi et al. 2011; Alahmad et al. 2018). In Australian cropping systems, all three above species occur (Fig. 1) with *Fp* being more common than *Fc* and generally more aggressive than the closely related *Fg* (Burgess et al. 1996; Chakraborty et al. 2010).

Fig. 1: Distribution of recorded (O) CR incidence and suitable locations (+) in a 0.5° by 0.5° grid for *(a) F. graminearum, (b) F. pseudograminearum* and *(c) F. culmorum* Adapted from: (Backhouse & Burgess 2002).

1.1.1 ECONOMIC IMPACT

Wheat disease costs Australia $913 million in lost revenue annually or $76.64 per ha, and was last reported to be responsible for an average loss of 19.5% of the total value of the crop (Murray & Brennan 2009). CR ranks fourth for current wheat yield loss associated with disease impact in Australia, behind yellow spot (*Pyrenophora tritici-repentis*), stripe rust (*Puccinia striiformis* f.sp. *tritici*) and Septoria nodorum (*Parastagonospora nodorum* (syn. *Stagonospora nodorum*)) (Murray & Brennan 2009). The greatest yield losses from CR have been reported in durum wheat (*Triticum turgidum* var. *durum*), followed by bread wheat (*Triticum aestivum*) and lastly, barley

(*Hordeum vulgare*) (Hollaway et al. 2013), with other grains crops not being as significant in commodity size or reported losses.

The Australian Grains Research and Development Corporation (GRDC) invests in and manages research in the Australian grains industry including the management of CR, and divides the country into three primary growing regions; the Northern region, the Southern region and the Western region (Fig. 2). Estimated potential yield loss from CR has increased over time in Australia, from 3.0% in 1998 to 22.2% in 2008, across the Northern region, 0.4 to 10.5%, across the Southern region and 0.1 to 1.5% across the Western region (Simpfendorfer et al. 2019). CR is estimated to cost the Australian wheat sector \$79 million per year (\$6.63 / ha) in lost revenue and has potential to cost as much as \$434 million (\$36.44 / ha) (Murray & Brennan 2009). Potential cost is estimated as the combination of yield (1) and quality (2) losses. Potential yield loss is calculated as:

$$L_p = Y_0(100 - K_p)/100 \qquad (1)$$

where Y_0 is the disease-free yield, L_p is the potential yield loss in tonnes / hectare without control and K_p is the percentage potential loss from the disease (Murray & Brennan 2009).

Potential quality loss is calculated as:

$$Q_p = D_p Y_p \qquad (2)$$

where Q_p is the value of current quality losses and D_p is the price discount in dollars per tonne (Murray & Brennan 2009). These estimations assume no interaction between diseases.

Losses in barley were calculated from the equations above and were reported at 9.5, 2.5 and 2.6% of yield in the Northern, Southern and Western regions, respectively from *Fp* (Murray & Brennan 2010). However, potential losses are as high as 34.3, 10.8 and 13.8% of total yield in years when disease develops (Murray & Brennan 2010). The overall yield impact of CR induced by each of the pathogens *Fc* and *Fp* is similar on a per-plant basis (Hollaway et al. 2013).
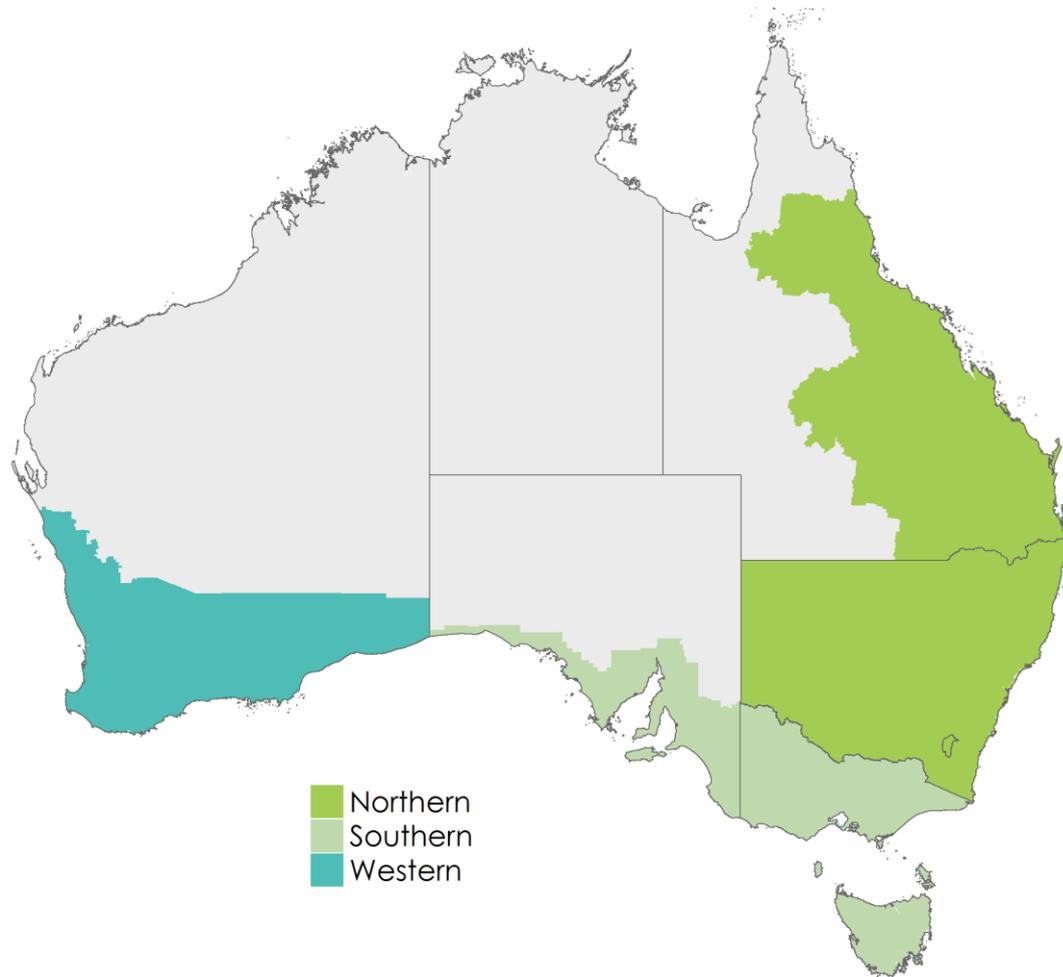
Fig. 2: Australian grains growing regions (GRDC).

1.1.2 DISEASE CYCLE

Understanding the CR disease cycle is necessary to identify potential opportunities for novel automated sensor detection. *Fp* infections in wheat are monocyclic during the growing season, motivating the classification of CR as a polyetic disease, a disease influenced directly by incidence in the preceding season (Backhouse 2006). The primary inoculation source for CR infection in wheat is fungal hyphae fragments in grass or cereal debris that come into contact with the sub-crown internode, crown, roots or stem (Burgess et al. 1993). After contact, the fungus begins to colonise the plant tissue from the local area of infection, moving to the stem, leaf sheaths and crown tissue (Klein et al. 1988). CR infection can occur at any stage in plant development, dependent upon moisture and contact with host debris. However, as CR is progressive, early infections often result in greater disease development in conducive seasons. This is an important consideration for machine-vision system development, as it indicates that early detection is preferable if possible. The lifecycle of *Fc* is not fully understood.

However, CR disease caused by *Fc* has been reported to progress in a similar way to disease caused by *Fp* (Knight et al. 2017). In wheat an ABC transporter protein is required for development of disease from *Fc* which may be important to consider when envisioning possible novel CR detection systems, as it is hypothesised that this protein plays a role in conveying resistance to a potential antifungal compound that wheat synthesises in response to infection by removing the compound from infected cells (Skov et al. 2004). This molecular response could potentially be used to detect CR before visible symptoms become apparent using spectral responses indicative of molecular changes brought upon by infection with a *Fusarium* pathogen. Molecular changes have been previously reported to be distinguishable between healthy and *Fusarium* damaged wheat kernels using near-infrared spectroscopy (Peiris et al. 2009).

While there is no documented role of asexual macroconidia or sexual ascospores as a source of primary inoculum, or otherwise, macroconidia can be used to quickly determine whether the source of infection is *Fc*, as its macroconidia are generally less elongate than those of *Fg* or *Fp* (Leslie & Summerell 2006; Alahmad et al. 2018). However, *Fg* and *Fp* are morphologically identical and, as such, are often confused (Fig. 3) (Leslie & Summerell 2006).
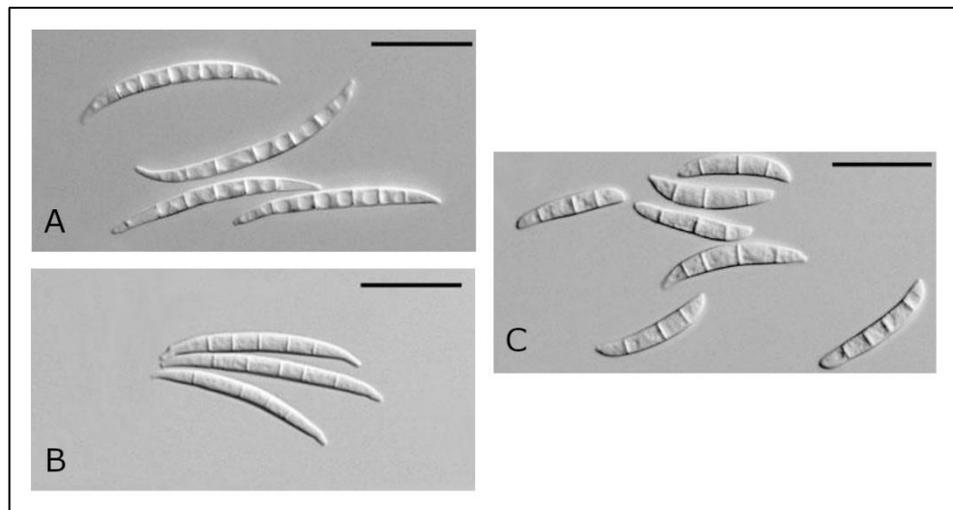
Fig. 3: Macroconidia. (a) *F. graminearum*, (b) *F. pseudograminearum* and (c) *F. culmorum*. Scale bar = 25 µm. Adapted from: (Backhouse & Burgess 2002).

1.1.3 IDENTIFICATION OF SIGNS AND SYMPTOMS

In seedlings, severe early CR infection can result in pre-emergence necrosis or seedling blight (Percy et al. 2012). Seedlings begin to form necrotic lesions basally, with lesions appearing on the coleoptile, subcrown internode and basal leaf sheaths (Purss 1969).

Traditionally, CR is identified between flowering and maturity by the presence of stem browning and the appearance of whiteheads, dead grain heads caused by early senescing culms which can be identified by the premature loss of colour of inflorescence (Klein et al. 1991; Hollaway et al. 2013; Knight et al. 2017; Knight et al. 2020). CR lesions can be identified by their honey-brown colour, often in association with whiteheads, with reports out of Victoria, Australia of 2–3% of heads affected on average (Hollaway & Exell 2010).

Disease predictive modeling has been used to facilitate CR management options. Backhouse reported on a study of long-term trials in Moree, NSW and Billa Billa, QLD, Australia, that CR infection rates were positively correlated to yield and rainfall in the previous season and negatively correlated to rainfall in the summer fallows (2006). Backhouse noted that the negative correlation in the summer fallows may be due to increased decomposition brought on by the increased rainfall, effectively removing inoculum from the system. Using this approach, a CR forecasting model was developed (Backhouse 2006). It has been reported more recently with a twelve year survey of 957 wheat crops in western Victoria, Australia that disease severity of CR induced by *Fc* and *Fp* was negatively associated with rainfall (Hollaway & Exell 2010). However, this was only reported across two regions, Western District and Wimmera. As CR is residue-borne, spatiotemporal modeling of disease dispersal and advancement may be possible, particularly in no-till farming systems, while other dispersal methods (i.e. ascospores, conidia) may be important for modeling movement in low incidence scenarios (Backhouse 2014).

Incorporating the results of quantitative fungal DNA soil tests like *PREDICTA® B* (Ophel-Keller et al. 2008) into predictive models may be beneficial in making decisions that impact the effectiveness of any disease control measures considered. However, this technology is most effectively used to measure pathogen levels in the

soil prior to planting and further benefit may come from developing machine vision solutions for quantification in-season.

1.1.4   CONTROLLING THE DISEASE

Rotating into a non-host crop has been shown repeatedly to be effective at reducing CR incidence (Burgess et al. 1996; Lamprecht et al. 2006; Evans et al. 2010). However, several years are required in a non-host crop to effectively eliminate the pathogen (Wildermuth et al. 1997). Where prolonged rotation with a non-host crop is not feasible, stubble burning and soil solarisation have been demonstrated to reduce inoculum levels. (Summerell et al. 1989; Burgess et al. 1996; Saremi et al. 2011). Another promising option for managing CR inoculum levels in the paddock is inter-row sowing. As CR develops through infection with *Fusarium spp.* that requires close contact between the host crop and previous stubble remains, inter-row sowing can reduce paddock CR build up by providing a buffer between the inoculum from the previous crop and a new crop (Simpfendorfer et al. 2012). However, inter-row sowing is only feasible in the short term and if CR is allowed to develop, few control options currently exist.

In-season CR control is currently not commercially available, however targeted fungicide approaches have shown potential for yield benefits. Improvements of 5–15% in retained yield were reported across 22 sites when fungicide was applied at the base of plant tillers. This was achieved using inter-row droppers with angled nozzles to target the pathogen more effectively than foliar applications (Simpfendorfer, cited in. Simpfendorfer et al. 2019). This approach is likely not widely adopted due to the cost to benefit risks involved for a minor (5–15%) yield improvement. As new fungicide modes of action are developed, in-season treatment may become more beneficial and cost-effective. A novel treatment for the reduction of CR inoculum levels is the use of microwave radiation (Petronaitis et al. 2018). Success in reducing inoculum recovery to 0% was established in the laboratory, however this method has not been established for use in the field and the practicality of such a method has not been determined. Incorporating real-time machine vision solutions into current and emerging integrated CR control strategies could potentially reduce the cost of treatment application by targeting control to only affected areas of the paddock, increasing the cost-effectiveness of treatments.

With current control options focused on pre-season management, the development of resistance and tolerance in high yielding germplasm is important to reduce the impact of CR as it becomes more prevalent in Australia and international cereal growing regions. Currently, breeding programs need to screen high volumes of germplasm at maturity in order to obtain yield data for tolerance measures (Kelly et al. 2016; Forknall et al. 2019). A machine-vision high-throughput phenotyping (HTP) platform could increase germplasm development efficiency over traditional methods and will be discussed in depth in the following section.

1.1.5   DEVELOPING RESISTANCE AND TOLERANCE

Breeding programs must effectively and efficiently screen for new material that is both resistant and tolerant to CR. Tolerance is the ability of a plant to limit yield loss while under disease pressure, whereas resistance is the ability of a plant to limit the infection of the pathogen causing the disease (Forknall et al. 2019). Both resistance and tolerance are important in breeding programs, as the final goal is not only to limit infection but to maximise yield, in the presence of disease.

Several recent developments have increased efficiency of CR resistance and tolerance breeding. Novel screening bioassays have been developed to select lines with resistance to CR more rapidly than in field programs (Wallwork et al. 2004; Mitter et al. 2006). Further developments have been made to optimise CR scoring between genotypes, with a study by Percy et al. (2012) examining CR scoring versus fungal colonisation indicating that visible symptoms are closely correlated to the spread of the fungus. This study further demonstrated that differences between genotypes could be more accurately assessed using the most recently expanded leaf sheaths, as the fungus appears to colonise these tissues both vertically and laterally (Percy et al. 2012). This research has been extended to improve sampling strategies in the selection of resistance in the field (Macdonald et al. 2016).

However, screening for resistance and tolerance to CR remains time and capital consuming as well as labour intensive, as germplasm must be grown to maturity before being harvested, stored, sampled, cleaned and then examined by trained assessors. Further, human plant disease assessment has been reported to be significantly variable, both between assessors and between plant material scored by individual assessors, as such, tools to increase effectiveness and decrease variability should be examined

(Nutter Jr et al. 1993; Newton & Hackett 1994; Nutter 1997). One such approach is the use of machine-vision and machine-learning technologies in the development of an HTP methodology to assist breeders in effective germplasm delivery.

The development of a machine-vision based HTP methodology for resistance and tolerance would allow for only germplasm with the most effective defense mechanisms against or tolerance of CR to move forward in breeding programs. This would free capital and resources for investment into additional material, increasing breeding productivity and delivering effective lines to growers quickly.

Improvements in screening have been beneficial in producing new germplasm more quickly than ever before (Wallwork et al. 2004; Mitter et al. 2006; Percy et al. 2012; Macdonald et al. 2016). Incorporating machine sensing technologies into these pipelines could assist breeders in further increasing productivity while reducing costs. Machine vision developments for crop disease detection are detailed in the literature review in Chapter 2.

2.1 FORMAT OF THESIS

The following chapters enclosed in this thesis are presented as a series of papers evolving in complexity of analysis of wheat CR detection and quantification. This presentation was selected as papers are intended for submission. Due to this formatting decision some sections may be repeated to differing degrees across chapters, including experimental design, materials and methods. A glossary containing domain specific terms is presented in Appendix A.

2.2 THESIS OUTLINE

Chapter 1 introduces CR and provides an overview of CR biology, impact and issues. CR control measures are discussed. The format of the thesis and the thesis outline is provided.

Chapter 2 provides a literature review around existing spectral sensing technologies and data analysis algorithms. An emphasis is placed on plant disease detection, specifically around early disease detection and machine learning-based analysis approaches.

Chapter 3 outlines the development of specific models for CR detection from individual datasets. Datasets incorporated three glasshouse and two field trials,

spanning two years. An overview of the trials used for each chapter is included in Appendix B. The impact of traditional 80-20% train-test splits and the relatively smaller 20-80% train-test splits for machine learning training was evaluated to determine if benefit could be derived from smaller splits. Additional tables and figures for Chapters 3, 4 and 6 are included in Appendix C. An overview of the developed code for Chapters 3–8 is provided in Appendix D. Code files are available upon request.

Chapter 4 outlines the development of general models for CR detection from combined data (i.e. data from multiple locations and timepoints). Datasets incorporated three glasshouse and two field trials, spanning two years. The impact of traditional 80-20% train-test splits and small 20-80% train-test splits was evaluated to determine if benefit could be derived from smaller splits when developing generalised models.

Chapter 5 evaluates the impact of plant feature location (i.e. tiller, centre, flag, head) in relation to the sensor on CR detection. Temporal impacts on CR detection were also evaluated.

Chapter 6 outlines the development of models for CR quantification. A multilayer perceptron artificial neural network is trained on six rating systems to evaluate the potential for CR quantification for use in high-throughput phenotyping of CR. Datasets incorporated three glasshouse and two field trials, spanning two years.

Chapter 7 outlines the development of camera-based models for CR detection. Multi-input convolutional neural networks were developed for the detection of CR using five bands in the near-infrared spectrum. A glasshouse trial in 2019 was used to develop the models.

Chapter 8 outlines the development of camera-based models for CR quantification. Multi-input convolutional neural networks were developed for the quantification of CR using combinations of five bands in the near-infrared spectrum to enable image-based high-throughput phenotyping of CR. A glasshouse trial in 2019 was used to develop the models.

Chapter 9 provides a discussion of the conclusions drawn from each of the previous chapters. Chapter 9 also presents a discussion around the opportunities for future research in this domain.

CHAPTER 2: LITERATURE REVIEW

Historically, plant disease identification has been the responsibility of human assessors. Yet, human data collection is known to show variability, particularly between assessors, and therefore, accuracy can vary significantly (Nutter Jr et al. 1993; Newton & Hackett 1994). The inter-assessor variation commonly seen in agriculture is likely due to imperfect scoring systems that leave ratings open to interpretation. This makes it highly unlikely that two assessors will score a plant identically with high repeatability. However, there are indications for the use of multiple sensor arrays to distinguish disease. If used in combination with novel processing algorithms, machine vision has the potential to improve upon and assist traditional human visual assessment.

This literature review aims to review the state of current technologies available not only to detect and discriminate crop disease, but to phenotype disease impact by successfully quantifying crop disease. Further, this review aims to expand the focus on foliar disease detection and quantification, to examine how previous research across crop disease domains can inform a disease detection and quantification methodology for visibly asymptomatic or low-symptomatic crop disease evaluation, such diseases including soil and stubble-borne fungal pathogens. Specifically, the review will outline current sensing technologies using both reflected and emitted waves in the electromagnetic spectrum. It will cover available sensors and tested approaches for disease discrimination and phenotyping, as well as, compare and evaluate the literature around traditional and machine-learning based data analysis approaches for evaluating crop disease data obtained from these sensors. Data analysis techniques for both multispectral and hyperspectral data will be discussed. Ultimately, this review aims to provide considerations for the development and deployment of a successful machine-vision system for the detection and quantification of visibly asymptomatic plant disease. Such a system should ideally provide early, accurate, repeatable disease detection while remaining cost effective.

Remote sensing is "the use of reflected and emitted energy to measure the physical properties of distant objects and their surroundings" (Moore 1979). Remote sensing began before World War I, with the advent of photography and has evolved to include what today are considered machine vision systems (Moore 1979). Since that time,

remote sensing has evolved to encompass the visible, red, green and blue (RGB) wavelengths to the ultraviolet, near-infrared, shortwave-infrared and thermal wavelengths, with sensor technologies evolving into multispectral and hyperspectral systems.

In the following sections, the reflectance-based techniques available in the traditional visible spectrum, the near-infrared (encompassing the shortwave-infrared), ultraviolet, and the thermal infrared spectrum will be discussed in relation to potential use in the detection of CR disease caused by *Fusarium spp*. Additionally, the use of combinations of wavebands from these portions of the spectrum will be discussed in relation to hyperspectral and multispectral sensing technologies. Traditional and wavelet analysis techniques for feature extraction, data compression and the general utilisation of spectral data will be evaluated. Finally, the significance of this study will be discussed, and the aims and objectives of the project stated.

1.1 ELECTROMAGNETIC SENSING

Ultraviolet, visible, near-infrared, shortwave infrared and thermal portions of the electromagnetic spectrum have been reported to be related to plant health and physiology. Healthy plant canopies have been reported to have a low reflectance in the visible spectrum, due to radiation absorption by photosynthetic molecules. This reflectance increases dramatically in the near-infrared range, caused by light scattering at the air-cell barrier of the internal tissue. When reflectance in the shortwave infrared range is observed, values decrease, primarily due to absorption by water and organic molecules (Woolley 1971; Jacquemoud & Ustin 2001).

1.1.1 RGB SENSING

The most rudimentary remote sensing techniques rely on traditional grayscale or red-green-blue photography (RGB), and have shown promise for detection of visible plant disease symptoms in conjunction with increasingly robust analytical platforms (Camargo & Smith 2009). These combinations of complementary technologies and techniques have been shown to be useful in a variety of potentially commercial applications, from outperforming experts in lesion estimation and counting to calculating necrotic/chlorotic leaf area (Bock et al. 2008). Further, RGB sensors are cost effective. Additional applications include selecting weeds for automatic spot spraying applications, using colour thresholding and texture analysis (Rees et al. 2009)

and distinguishing diseased tissue from cell-phone images, through extraction of texture-based features and gradient magnitude patterns (Neumann et al. 2014). Using RGB sensors in concert can expand potential applications. Utilising data fusion techniques, incorporating two 5 megapixel RGB cameras and a daylight artificial illumination source, tulip bulbs infected with Tulip Breaking Virus have been sorted automatically, with high accuracy (Polder et al. 2014). Although RGB sensing has been successfully applied to various agricultural applications, it is unlikely to be suitable for visibly asymptomatic or pre-symptomatic disease detection as applications in disease discrimination rely on changes in pigment levels, changes synonymous with visible disease symptoms. Additionally, many soil and stubble-borne diseases can exhibit symptoms that are difficult to detect through the crop canopy, often requiring removal from the soil for assessment.

1.1.2 ULTRAVIOLET

Ultraviolet (UV) light waves are shorter than visible light waves, with wavelengths below 400 nm. Plants have varying ability to utilise UV light, usually correlated to their natural environment, with plants at higher altitudes often utilising UV light differently, having evolved to survive with less atmospheric protection. One constant is that higher the carotenoid/chlorophyll-*a* ratio of the plant tissue, the lower the UV reflectance and the higher the absorption of that tissue (Filella & Peñuelas 1999). It has been hypothesised that the hyperspectral carotenoid/chlorophyll-*a* ratio could be a viable indicator of nutritional value of tall fescue (*Festuca arundinacea)* (Yang et al. 2010). As chlorophyll-*a* is detectable in the visible spectrum, plant disease with few or without visible symptoms may be more difficult to detect using any technique incorporating visible light. This could provide a barrier to early detection.

While it has previously been reported that leaf reflectance in the UV spectrum is low and generally uniform (Knipling, 1970), it has more recently been demonstrated that glaucous (waxy) leaf surface waxes are very effective reflectors of both UV and longer wavelength light and exhibit significantly higher reflectance when compared to hairy and hairless leaves (Holmes & Keiller 2002). This waxy epidermis has also been reported to be positively correlated with increased exposure to UV light (Kakani et al. 2003, Steinmüller & Tevini, 1985). As most UV photons are intercepted by the leaf epidermis (Rundel, 1983), and the epidermis is a likely entry-point for disease, UV reflectance may be a viable indicator of plant disease.

However, the impact of disease on these pigments, measured in the UV spectrum, has not been deeply investigated, so the impact upon UV reflectance should be considered in future disease discrimination and quantification work.

1.1.3 THERMAL INFRARED

New advances in machine sensing and analysis have allowed thermal imaging to be used in agricultural industries. One such application is the visualisation of stomatal closure, an indicator of drought, using infrared thermography (IRT) (Jones et al. 2002). Drought phenotyping using thermal infrared as an indicator for water stress has been reported (Buitrago et al. 2016, Jones et al. 2009). Thermal infrared spectroscopy has also been reported as an important indicator of plant transpiration (Hou et al. 2019, Tian et al. 2014). However, most current data indicates that IRT, alone, is not enough to distinguish a localised rise in temperature and associate it with a stressor at the canopy level (Oerke et al. 2006; Lenthe et al. 2007). However, on a microscopic scale, Oerke was able to detect an increase in the maximum temperature difference (MTD) in leaves infected with apple scab, allowing for disease quantification, indicating potential for use in the early stages of resistance breeding pipelines. Oerke (2011) also reported differences in scab intensities as the disease progressed but indicated that infrared thermography must be used in conjunction with other data, if disease identification is desired. Nevertheless, MTD may show potential for early screening of resistant cultivars.

It has been reported that thermal infrared sensing can be used to discriminate between a variety of leaf chemical components both in deciduous tree and agricultural plant species (Riberro da Luz & Crowley, 2007). A thermal infrared sensor with a range of 2,500–15,380 nm was used to successfully discriminate differences between natural background and leaves and to identify cellulose, cutic, silica, xylan and oleanolic acid, important leaf chemical components. It was suggested that for successful identification at this scale a thermal infrared sensor needs to have a high signal-to-noise ratio and a small instantaneous field of view.

1.1.4 CHLOROPHYLL FLUORESCENCE IMAGING

A novel machine vision approach is chlorophyll fluorescence imaging (Ch-FI) which has been shown to reveal cell death, with greater contrast and earlier than other systems (Chaerle et al. 2004; Chaerle et al. 2007) and to correctly identify resistance to downy

mildew (*Bremia lactucae*) in lettuce cultivars (Bauriegel & Herppich 2014). However, the high energy source required to excite the tissue indicates this technique is unlikely to be practically successful in field or airborne applications.

An emerging approach in Ch-FI is passive solar induced chlorophyll fluorescence imaging (SIF). SIF is an indicator of the photosynthetic process itself and not of a chlorophyll response (Meroni et al. 2009) and does not require the high energy source of traditional Ch-FI. This could be useful in crop disease detection as a fluorescence response could potentially be measured before a visible stress indicator presents itself (e.g. chlorophyll reduction) enabling earlier stress detection than reflectance alone (Entcheva et al. 2008). It has been reported that fluorescence is positively correlated (declines) with photosynthesis in high stress conditions (Van der Tol et al. 2009).

The fluorescence (F) signal can be interpreted using either radiance or reflectance-based measurements (Meroni et al. 2009). Radiance measurements can be recorded in either physical or auxiliary units and exploit Fraunhofer lines (Meroni et al. 2009). Fraunhofer lines are narrow dark lines of the solar and atmospheric spectrum where irradiance is strongly reduced. The three main Fraunhofer lines exploited for use in SIF are the hydrogen (H) absorption line, centered at 656.4 nm and two oxygen (O_2) absorption lines, centered at 687.0 and 760.4 nm (Meroni et al. 2009). Reflectance-based approaches for measuring SIF use reflectance indices related to F but cannot report physical or auxiliary units. These indices use relationships focusing on the effect of F on the red-edge area of the spectrum (650–800 nm). Most of these indices require 2–3 narrow channels to develop a relationship to F, instead of exploiting Fraunhofer lines (Meroni et al. 2009).

Very little work has been reported in successful detection of crop disease using SIF. However, several parties have reported success in water stress detection using both ground-based (Stellmes et al. 2007, Evain et al. 2002, Mcfarlane et al. 1980) and airborne systems (Zarco-Tejada et al. 2009). Further, success has been reported in ground-based detection of herbicide stress (Carter et al. 2004, 1996) and airborne nitrogen (N) stress (Corp et al. 2006). When using a fusion of SIF and reflectance spectroscopy for the detection of wheat stripe rust, a correlation of determination of 0.89 was reported (Jing & Bai 2019). Many of the studies using SIF for stress detection are in the early stages and are often limited to the capability of a particular method or sensor (Meroni et al. 2009).

1.1.5 MULTISPECTRAL

Multispectral imaging takes an array of readings of light at multiple bands of wavelengths across the electromagnetic spectrum. Typically, these sensors record reflected light at the pixel level in 3 to 15 separate bands. Multispectral sensors show promise in distinguishing stressors by analysing differences in reflectance patterns at multiple wavelengths. Several research parties have observed unique reflectance patterns in different model systems (Steddom et al. 2005; Franke & Menz 2007; Polder et al. 2014). However, Carter and Knapp (2001) reported that relying on chlorophyll concentrations alone to screen between plant stresses would be difficult and unlikely to be plausible. Though, they did notice significant changes in reflectance in the 548–599 nm and 701–723 nm ranges, when observing leaf optical chlorophyll stress responses in five plant species (sweetgum (*Liquidambar styraciflua* L.), red maple (*Acer rubrum* L.), wild grape (*Vitis rotundifolia* Michx.), switchcane (*Arundinaria gigantea* (Walter) Muhl.), and longleaf pine (*Pinus palustris* Miller)) and stress responses simulated in vitro by adding chlorophyll *a* + *b* solutions of sequentially decreasing concentrations to fiberglass filter pads, indicating these ranges may be closely related to senescence responses in plant physiological systems.

In 2005, Steddom observed that when individual wavebands and vegetative indices calculated from those wavebands, obtained from a multispectral radiometer were compared with visual disease estimates from human assessors, precision was improved for all indices and wavebands, versus human discrimination, when assessing the severity of *Cercospora beticola* (leaf spot) on sugar beet. Steddom reports the improvement is likely due to visual estimates of disease severity often being overestimated when infections are small, in addition to the learning curve required for human assessment of crop disease to become repeatably accurate (Sherwood et al. 1983; Nutter 1997; Steddom et al. 2005). Additional success of multispectral systems has been reported in the classification of Tulip Breaking Virus, using multiple bands in the visible spectrum, with results reported to be comparable to human experts (Polder et al. 2014).

There is evidence for potential in-field applications utilising multispectral sensors derived from hyperspectral sensing systems (Bravo et al. 2003; Huang et al. 2007). Bravo evaluated yellow rust (*Puccinia striiformis*) of wheat at boom-height (~1 m), in the field, using hyperspectral data in the visible spectrum and reported that

classification error was as low as 4%, down from 12%, when sensing was combined with irradiance normalisation and a spatially moving average. Using quadratic discrimination based on the Bayesian decision rule where an object is classified dependent on which class maximises the posterior probability to select bands, the group was able to replace the spectrum with four selected bands, with no loss of performance (Bravo et al. 2003). This derivation of bands was replicated in 2011 where it was indicated that detection of powdery mildew (*Blumeria graminis* f. sp. *hordei*) in wheat was possible using 13 extracted bands (Mewes et al. 2011). Retaining disease detection and quantification assessment accuracy, while reducing the total number of features examined allows for decreased processing time and lower sensor cost as hyperspectral bands of interest can be replicated in a multispectral system, using spaced bands at specific wavelengths. This derivation of important hyperspectral features from diseased crops for the use in low-cost multispectral sensors for disease detection and quantification will begin to be evaluated in this PhD thesis and should be further developed in future studies.

Further to ground and airborne-based sensing approaches, multispectral imaging lends itself to space-based approaches due to the limited number of bands needed for sensing. Fewer bands equate to fewer required sensors, allowing smaller imaging devices. Smaller devices are important in space-based systems due to cost limitations with sending larger payloads into space. Space-based multispectral imaging was included in some of the earliest space-based imaging platforms, such as the Landsat program (Williams et al. 2006). However, like most evolving technologies, there are challenges when using these technologies at this scale. One study used QuickBird, a now decommissioned imaging satellite, and calculated the normalised difference vegetation index (NDVI) to detect early infections of powdery mildew (*Blumeria graminis*) and leaf rust (*Puccinia recondita*) in winter wheat (Franke & Menz 2007). However, the researchers were only able to determine moderate suitability for the early detection of crop disease using this satellite system, as the images captured had a low spatial resolution of approximately 2.4 m. Garcia-Ruiz found that diseased versus healthy plant classification accuracy was greater in high-resolution, drone data (5.45 cm/pix), in comparison to lower resolution, aircraft data (0.5 m/pix) (Garcia-Ruiz et al. 2013). Therefore, to readily detect early crop disease infestations, the spatial resolution likely must be closer to the individual plant level. Further, attempting to

discriminate between plant stressors at low resolution would be considerably difficult using vegetation indices as they are seldom stress specific.

1.1.6 HYPERSPECTRAL

Hyperspectral imaging aims to resolve the fine spectral variations that may be indicative of stress (i.e. disease detection). These variations may not be resolved when using fewer, broad bands, as with multispectral imaging. To accomplish this hyperspectral sensors obtain hundreds of largely contiguous wavebands, while multispectral sensors record fewer spaced bands; this distinction often means hyperspectral system data has significantly higher volume than multispectral sensing system data (Hagen & Kudenov 2013). The amount of data generated by hyperspectral systems can cause post-processing to be a computational challenge. Hyperspectral systems typically employ visible to near-infrared (VIS-NIR) spectral bands, therefore hyperspectral literature primarily comprises visible and NIR sensing. Visible and near-infrared hyperspectral applications and computational considerations will be examined in this review.

The near-infrared (NIR) spectrum (roughly 750–2500 nm) includes wavelengths outside of the range visible to the human eye. Reflectance in the NIR spectrum has been reported to correlate to plant starches, proteins and water content, among other molecules (Kumar et al. 2002). This correlation makes the NIR spectrum a good candidate for the detection and discrimination of plant stresses, particularly early in the disease cycle, before visible symptoms begin to appear or in cases where the stresses exhibit few visible symptoms, are altogether visibly asymptomatic or are difficult to detect through the crop canopy. It was demonstrated in 2017 that disease symptoms of powdery mildew (*Blumeria graminis* f. sp. *hordei*) on barley could be detected two days before symptoms appeared visually, using VIS-NIR reflectance wavelengths recorded with a hyperspectral sensor (Thomas et al. 2017). Pre-symptomatic detection may be important in developing detection mechanisms for soil or stubble-borne diseases, such as CR, which oftentimes exhibit few visual symptoms. Additionally, this relationship between NIR reflectance and molecular composition allows for applications such as nutrient measurement and plant tissue quality analysis, such as cotton fibre quality assessment to be performed using spectroscopy, complementing genetic techniques (Chen et al. 2002; Sui et al. 2008).

Hyperspectral line scanners have allowed for accurate leaf transmittance, reflectance and disease mapping. However, the high cost associated with these sensors has prohibited adoption (Bergsträsser et al. 2015). Recent research comparing a hyperspectral line scanner to a novel multispectral camera has determined that narrowband LEDs and silicon complementary metal-oxide semiconductor (C-MOS) imaging detectors could be used for disease detection of powdery mildew (*Blumeria graminis* f. sp. *hordei*) in barley and *Cercospora* leaf spot (*Cercospora beticola*) in sugar beet, in place of line scanners, cutting the prohibitive costs associated with these devices (Grieve et al. 2015). However, in addition to requiring close proximity (1–2 m), the authors report a scan time for the new system of "typically…less than a minute for a 3 component SVI" (spectral vegetation index), which would require significant post-processing to achieve sufficient overlay of individual LED cycles in the field. Moreover, each additional SVI component would increase full-cycle time. While cost-effective, these novel sensing systems need to be further developed to replace current field devices.

Disease detection with hyperspectral sensors has been successful in laboratory, glasshouse and field systems. In 2007 it was reported that the use of multiple narrow bands at close proximity, in the NIR-SWIR range (1350–1750 nm and 2200–2500 nm), were capable of the discrimination of infection of apple scab (*Venturia inaequalis*) from healthy leaves, immediately after infection (*c*-values > 0.8) (Delalieux et al. 2007). Bands in the VIS spectrum (650–700 nm) became more important after infection was fully established at three weeks (*c*-values > 0.8), indicating the importance of temporal patterning in crop disease detection (Delalieux et al. 2007). In addition, Mahlein et al. (2010; 2013) were able to detect symptoms of three different diseases, Cercospora leaf spot (*Cercospora beticola*), powdery mildew (*Erysiphe betae*) and rust (*Uromyces betae*) on sugar beet, with results suggesting potential to differentiate between diseases by using combinations of two or more spectral vegetation indices.

Fewer studies have been completed using airborne hyperspectral data. However, potential exists for the use of spectral vegetation indices in airborne applications. Huang saw potential in using the PRI (photochemical reflectance index) on airborne hyperspectral data in order to discriminate wheat infected with yellow rust (*Biotroph Puccinia striiformis*) under field conditions (Huang et al. 2007). In 2011, the influence

of resolution on hyperspectral data was evaluated. Images of an infection of powdery mildew (*Blumeria graminis*) in wheat were obtained using an aerial hyperspectral sensor, the Integrated Spectronics, HyMap™ (450–2500 nm, 5 m resolution). Hyperspectral data was processed using Spectral Angle Mapper (SAM) and Support Vector Machine (SVM) classifiers. It was concluded that adequate fungal detection was possible using 13 of 109 possible wavebands, showing potential for the detection of crop disease using models incorporating few wavebands (Mewes et al. 2011). The large amount of data acquired, and the cost associated with many of these sensors may be prohibitory for widespread technology adoption. To reduce the burden of processing hyperspectral data, transformation techniques should be investigated.

2.1 DISEASE DISCRIMINATION – DATA TRANSFORMATIONS

Successful machine vision systems rely on more than high resolution optics. The analysis of data, particularly regarding feature and model selection can impact the discriminatory capability of a system. For instance, hyperspectral imaging is likely important for the resolution it provides for crop disease detection. However, data processing is often burdensome. Data transformation techniques such as the discrete wavelet transform (DWT) and principal component analysis (PCA) may enable the practical use of hyperspectral data. In the following sections several data transformation techniques will be reviewed with a focus on detection system optimisation, evaluating feature selection and data reduction methods, including spectral vegetation indices, principal component analysis and wavelet transformations.

2.1.1   SPECTRAL VEGETATION INDICES

Post-processing algorithms for multispectral and hyperspectral data analysis greatly affect the way data are viewed and interpreted. Some of the most commonly used are spectral vegetation indices (SVIs). The use of SVIs may enable the resolution of crop disease features that cannot be detected at the resolution available from individual bands. SVIs can provide additional information to what is seen using individual bands by determining relationships between bands. Further, SVIs may enable the use of satellite data for stress detection. Widely used SVIs are NDVI (normalised difference vegetation index), ARVI (atmospherically resistant vegetation index), SAVI (soil-adjusted vegetation index), and VARI (visible atmospherically resistant index) with new indices consistently being developed and older indices improved (Gitelson et al. 2002) (Table 1). Different SVIs are developed to discriminate different materials, such

as, soil from foliage or water stressed foliage from hydrated foliage. This discrimination is achieved using large differences in specific spectral bands between materials. These differences are used to create indices which can be used to quantify the variation between the materials.

Table 1: Spectral Vegetation Indices

Index	Equation	Reference
Anthocyanin Reflectance Index	$ARI = \left(\dfrac{1}{R550}\right) - \left(\dfrac{1}{R700}\right)$	(Gitelson et al. 2001)
Atmospherically Resistant Vegetation Index	$ARVI = (-0.18 + 1.17)\, NDVI$	(Kaufman & Tanre 1992)
Normalised Difference Vegetation Index	$NDVI = \dfrac{R800 - R670}{R800 + R670}$	(Rouse Jr et al. 1974)
Physiological Reflectance Index	$PRI = \dfrac{p531 - p570}{p531 + p570}$	(Gamon et al. 1992)
Soil-Adjusted Vegetation Index	$SAVI = \dfrac{(1+L)(NIR - Red)}{(NIR + Red + L)}$	(Huete 1988)
Structure Independent Pigment Index	$SIPI = \dfrac{(R800 - R445)}{(R800 + R680)}$	(Penuelas et al. 1995)
Visible Atmospherically Resistant Index	$VARI = \dfrac{Green - Red}{Green + Red - Blue}$	(Gitelson et al. 2002)

Mahlein (2013) developed specific spectral indices from three bands, one disease specific band and two bands used to create a normalised reflectance difference, which can detect and classify leaf spot (*Cercospora beticola*), beet rust (*Uromyces betae*) and powdery mildew (*Erysiphe betae*), in sugar beet. However, more work is needed to determine the potential for using such a system to determine differences between diseases with unknown indices. It is likely that utilising combinations of input data will allow for stronger discrimination and classification of stresses. Mahlein also investigated disease discrimination using combinations of indices (2010). She found that it is possible to use scatter matrices of multiple spectral vegetation indices, to determine unique cluster orientations within a matrix, indicative of unique disease signatures, but commented that the potential may be better realised when differentiating between abiotic and biotic stresses, versus exclusively biotic stresses.

2.1.2 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is an adaptive data reduction technique used to reduce the data dimensionality of large datasets, while minimising information loss. PCA creates a series of new uncorrelated features, or principal components, that maximise variance contained in fewer features than the original dataset (Jolliffe & Cadima 2016). Several research groups have shown accurate discrimination of both weeds in crop stands and crop disease, using PCA in combination with machine learning methods (López-Granados et al. 2008; Suzuki et al. 2008; Liu et al. 2010). It has been presented that PCA used in conjunction with linear discriminant analysis (LDA) or neural networks (NNs) outperforms equivalent models using raw spectral data (Suzuki et al. 2008). Further, PCA has been reported to increase the classification accuracy of decision tree classifiers when used as a data reduction technique in preprocessing (Nasution et al. 2018). PCA has also been shown to be useful in further reducing spectral features derived from the discrete wavelet transform (DWT) for the detection of visible crop fungal infections in an RGB image set (2616 images) (Pujari et al. 2013a). However, PCA has no regard for data labels as it creates new independent variables, which may impact patterns potentially important for classification (Behmann et al. 2015).

2.1.3 WAVELET ANALYSIS

The DWT (discrete wavelet transform) is a signal processing technique which allows for unique signal features to be identified at both local and global scales. The DWT is important for crop disease detection in relation to reflectance signal processing as it allows the resolution of crop disease features and plant responses at multiple scales. Further, this technique does not require additional signal smoothing as traditional techniques like PCA do. DWT analysis has been shown to outperform traditional data analysis techniques when evaluating hyperspectral data (Bruce & Li 2001; Bruce et al. 2002; Koger et al. 2003; Ge & Thomasson 2006; Ge et al. 2007). Further, the DWT has been reported to provide several advantages over traditional feature selection techniques, including increased computational efficiency, multi-resolution capability and superior signal smoothing capability (Bruce et al. 2002; Ge & Thomasson 2006; Ge et al. 2007). The wavelet approach (1) allows feature scaling to determine the amount of input data that is included in each feature:

$$\omega_s f(\lambda) = f(\lambda) * \psi_s (\lambda) = f(\lambda) * \frac{1}{s} \psi \left(\frac{\lambda}{s}\right) \qquad (1)$$

(Bruce & Li 2001)

where $f(\lambda)$ is the hyperspectral signature, s is the scale and $\psi s\,(\lambda)$ is the scaled wavelet function. This presents an advantage over traditional methods which require an operator derived from the smoothing operation (Bruce et al. 2002). The DWT can systematically vary the operator or waveband scale.

When using the DWT in conjunction with the dyadic filter bank algorithm (Burrus et al. 1998), the DWT has been reported to be more computationally efficient than traditional data feature selection methods, which require a separate smoothing operation (Bruce et al. 2002; Ge et al. 2007). Further, the DWT has been reported to exhibit superior signal smoothing capability over both Savitzky-Golay and Fourier filtering (Ge & Thomasson 2006). The DWT has also been reported to use less regressors, or features, than traditional techniques (Ge & Thomasson 2006; Ge et al. 2007).

Bruce (2002) reported that often larger scale operators were more useful in classification than smaller operators derived from high resolution data. Therefore, potential exists for the development of disease specific remote sensors using wider wavebands determined by the DWT, for in-field disease discrimination.

One important consideration when developing a wavelet-based feature selection methodology is the mother wavelet employed. As the Ricker (Ricker 1943), Daubechies (Daubechies 1988) and Morlet (Grossmann et al. 1990) families of wavelets are closer in shape to the Gaussian signal shape observed in spectral reflectance curves than the widely used Haar (Haar 1909) wavelet, they should be considered for use in crop disease detection methodologies (Ge et al. 2007). Further, the use of Daubechies mother wavelets has been reported to increase the classification accuracy of hyperspectral reflectance signatures in agricultural systems (Bruce et al. 2002; Koger et al. 2003).

The DWT has been reported to outperform classifiers incorporating full original reflectance signatures or features derived from original signatures, selected by traditional feature selection techniques (Bruce et al. 2002; Kempeneers et al. 2005; West et al. 2007). The DWT was reported to outperform the fast-Fourier transform (FFT) and the discrete cosine transform (DCT) by approximately 20% when classifying vegetation (Bruce et al. 2002).

Additional successful implementations of the DWT in agriculture include fungal and nitrogen stress detection (Kempeneers et al. 2005; West et al. 2007; Pujari et al. 2013a; Shi et al. 2017; Shi et al. 2018). Kempeneers et al. (2005) used reflectance values from 420–2467 nm and DWT analysis to detect fungal and nitrogen related fruit tree stress. Sequential floating forward selection (SFFS) was used to extract features from wavelet transformed hyperspectral data. The team found that features extracted from wavelet transformed data performed better than features extracted from original reflectance features alone. However, the large original dataset required may limit the potential for this approach to be used in commercial sensors. West et al. (2007) utilised the DWT as a pre-processing step for a data fusion system. The goal was to explore the possibility of feature extraction and dimensionality reduction of soybean rust disease detection over a two-week period using a multiclassifier system. The system was derived from multiple wavelet decompositions at different scales. As training data are generally not available in quantities sufficient for all possible true features at full scale, the DWT was seen as an ideal candidate for reducing dimensionality.

Further, West found that data from combined dates had less discriminatory power than when dates were analysed individually. It was found that as time progressed the accuracy increased as the disease became more widespread. This is likely to be consistent in other disease systems where diseases exhibit clear visible symptoms. The multiclassifier system performed with 89–90% accuracy, while LDA produced an accuracy range from 60–90%. The potential for wavelet analysis to reduce disease discrimination error is exhibited.

Recently research has indicated that wavelet-based feature selection can be used to successfully discriminate other fungal diseases including yellow rust (*Puccinia striiformis*) and powdery mildew (*Blumeria graminis*) in winter wheat (Shi et al. 2017; Shi et al. 2018). Additionally, wavelets have been used to successfully discriminate and extract fungal disease regions from images of multiple crops for successful implementation in neural network classification, with accuracies of 86.48% (Pujari et al. 2013a).

However, the DWT carries disadvantages as well. The DWT is not shift-invariant, meaning a shift in the input signal is not directly equal to a shift in the output wavelet coefficients (Bradley 2003, Bruce et al. 2002). In this instance, a shift in wavelength in the input signal (e.g. from 700 to 720 nm) does not necessarily equate to the same

magnitude of shift (e.g. 20 nm) in the output coefficients. To overcome this weakness, a continuous or redundant wavelet transform could be used (Ge et al. 2007). However, this would result in the loss of the computational efficiency provided by the DWT.

Overall, the advantages in spectral band selection, number of features, computational efficiency and original curve representation outweigh the limitations of shift-variance. The DWT is a strong candidate for use in preprocessing hyperspectral data for crop disease detection and quantification and should be considered for the development of a wheat CR discrimination methodology.

2.2 DISEASE DISCRIMINATION – MACHINE LEARNING

Machine learning algorithms are a set of statistical computer algorithms that can interpret data through experience with the data, or without being specifically programmed (Mitchell 1997). In the following sections, multiple machine learning approaches that have been successfully applied in agriculture will be examined, with an emphasis on crop disease discrimination.

2.2.1 NEURAL NETWORKS

Neural networks (NNs) consist of layers of interconnected nodes which are inspired by the neural pathways that make up an animal's biological nervous system. Each node acts as an individual neuron that can both transmit and receive signals from other nodes (Fig. 1). These machine learning methods interpret patterns from training data without being specifically programmed. NNs are constructed from three types of layers; an input layer, hidden layers and an output layer. The input layer consists of the initial data provided to the NN. Hidden layers are the intermediate layers of a NN, where the patterns are discovered. The output layer of a NN is the layer that returns the result for the given inputs. For each node an activation function is defined. The activation function determines if the node should be turned 'on' or 'off'. NNs have been reported to be useful in plant discrimination, weed detection and plant disease identification (Moshou et al. 2002; Moshou et al. 2004; Burks et al. 2005; Moshou et al. 2005; Karimi et al. 2006; López-Granados et al. 2008; Suzuki et al. 2008; Wang et al. 2008; Wu et al. 2008; Arribas et al. 2011).
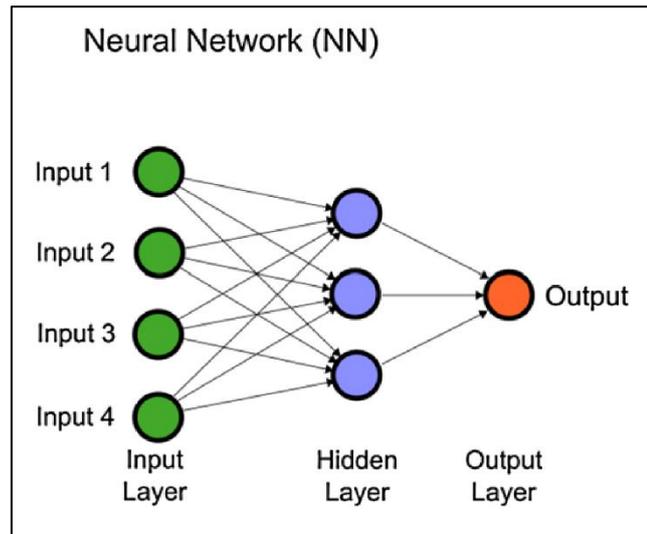
Fig. 1: A simple neural network (Behmann et al. 2015).

Several studies have reported success in detecting crop disease using NNs (Moshou et al. 2004; Moshou et al. 2005; Wang et al. 2008; Wu et al. 2008; Liu et al. 2010; Moshou et al. 2011). Fungal disease of rice (*Oryza sativa* L.) was discriminated in a laboratory setting using hyperspectral data (350–2500 nm) processed with PCA and a NN, with 91.6–100% accuracy (Liu et al. 2010). Additionally, both late blight (*Phytophthora infestans*) of tomato and fruit grey mold (*Botrytis cinereas*) of eggplant have been discriminated successfully using backpropagating NNs (BP-NN) (Wang et al. 2008; Wu et al. 2008). Further, in 2004 Moshou reported being able to detect yellow rust (*Puccinia striiformis* f.sp. *Tritici*) of wheat using a NN, from spray boom height, with 95–99% accuracy. Moshou again reported yellow rust detection capability in 2005, with 94.5% accuracy, using data fusion and a self-organising neural network, reducing discrimination error from 16.5% to 1% and showing real-time detection capability. This work was translated into an equipment-based sensing system that allowed real-time detection of yellow rust, combining GPS, hyperspectral and multispectral data inputs with a NN (Moshou et al. 2011). This outcome implies potential for disease specific sensors to be developed for other diseases (i.e. CR).

2.2.2 LOGISTIC REGRESSION

Logistic regression utilises a logistic function to model a binary dependent variable (0/1) (Fig. 2).
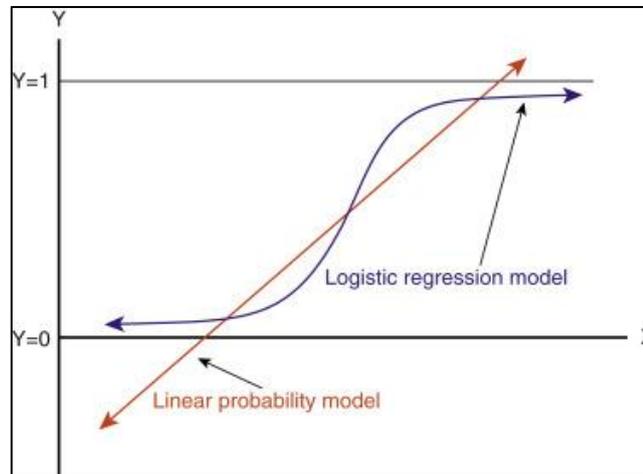
Fig. 2: Logistic regression model and linear probability comparison (Liu 2018).

Multinomial logistic regression analysis has been used to predict yellow mosaic virus (*Geminiviridae spp.*) in black gram (*Vigna mungo*) with 68.75% accuracy (Prabhakar et al. 2013). Additional work comparing k-nearest neighbors, support vector machines, naïve-Bayes and logistic regression for texture-based detection of three foliar diseases (powdery mildew, leaf spot, rust) on sunflower has been undertaken (Pinto et al. 2016). It was reported that logistic regression outperformed the other machine learning methods used with an average accuracy across diseases of 92.57%. However, logistic regression has not been investigated extensively for use in crop disease detection. Although, some use has been reported in developing disease forecast models based on weather patterns (Henderson et al. 2007).

2.2.3 DECISION TREE CLASSIFIER

Decision tree classifiers (DTCs) create models that predict target outputs from several input variables. Each tree "branch" or unique pathway through the classifier, represents a combination of features that lead to a specific target outcome class, conventionally called a "leaf"" (Fig. 3).
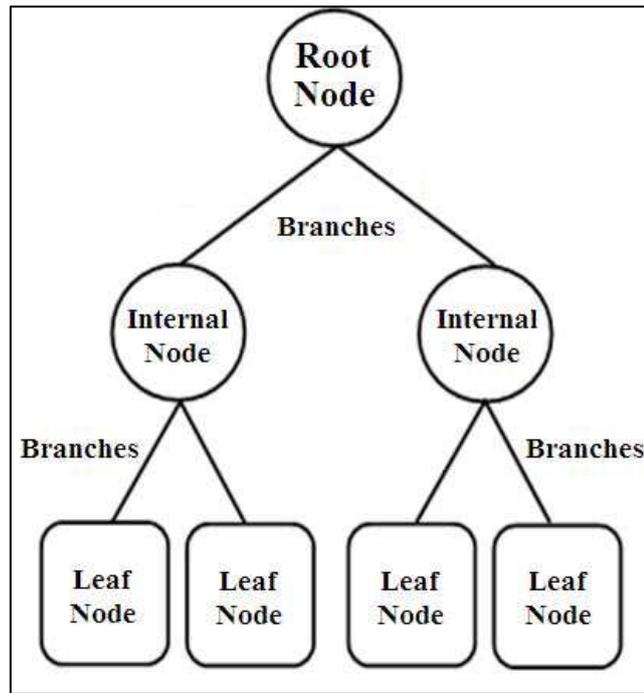
Fig. 3: Example of a decision tree classifier with branches representing feature combinations and leaves representing possible outcome classes (Sá et al. 2016).

DTCs are capable of disease detection and forecasting, discrimination of weeds from crop stands and recognition of nitrogen (N) status (Goel et al. 2003; Yang et al. 2003; Yang et al. 2004; Franke & Menz 2007; Cintra et al. 2011). DTCs have been used in conjunction with mixture tuned matched filtering (MTMF) and NDVI to detect fungal infections of powdery mildew (*Blumeria graminis*) and leaf rust (*Puccinia recondita*) in winter wheat from aerial hyperspectral sensors, with accuracies varying temporally from 56.8–88.6% (Franke & Menz 2007). Novel uses of DTC in agriculture include residue level estimation, tillage application prediction and disease forecasting (Yang et al. 2003; Cintra et al. 2011).

2.2.4 RANDOM FOREST CLASSIFIER

Random forest classifiers (RFCs) consist of an ensemble of decision tree classifiers where each individual tree returns a class prediction and the class with the majority vote becomes the predicted output. RFCs have been used to detect both gray leaf spot (*Cercospora spp.*) and Phaeosphaeria leaf spot (*Phaeosphaeria maydis*) in maize (Adam et al. 2017; Dhau et al. 2018).

2.2.5 SUPPORT VECTOR MACHINE

Support vector machines (SVM) are primarily a supervised learning approach which constructs a single or a set of hyper-planes in a high or infinite dimensional space. Separation between classes is based upon the division of the data by the plane (Fig. 4). SVM accommodates non-linear boundary definition by the use of the "kernel-trick", to resolve hyperplanes by increasing the dimensionality (Hofmann, 2006).
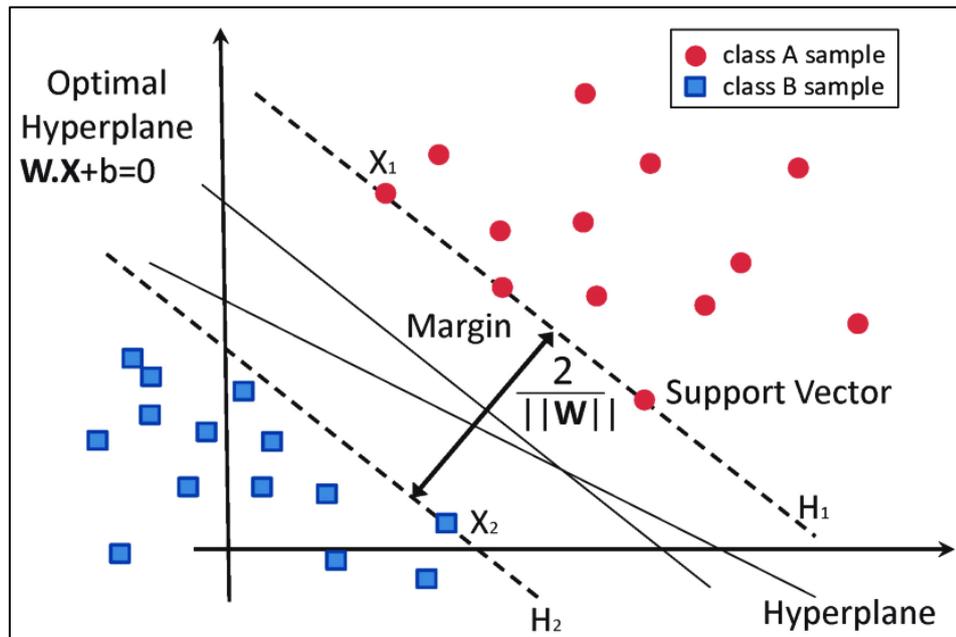
Fig. 4: Support vector machine, associated hyper-planes and classes (García-Gonzalo et al. 2016)

Several research parties have reported success in using SVMs to detect various crop stresses, from N deficiency to weed detection to pre-symptomatic crop disease identification (Karimi et al. 2006; Rumpf et al. 2010; Ahmed et al. 2012; Rumpf et al. 2012). Rumpf (2010) reported that by using multiple SVIs as inputs for a SVM, three sugar beet diseases could be discriminated (Cercospora leaf spot (*Cercospora beticola*), sugar beet rust (*Uromyces betae*), powdery mildew (*Erysiphe betae*)), with accuracies up to 97%. Of interest when detecting low-level or visibly asymptomatic disease, they reported the ability to detect pre-symptomatic disease with 65–90% accuracy, dependent upon timepoint. This suggests that temporal features and pre-symptomatic signatures should be considered when developing a crop disease detection framework.

2.2.6 K-NEAREST NEIGHBOR

K-nearest neighbor (KNN) classification is a non-generalised machine learning method. No internal model is created, but rather the method assigns classification based on a majority vote of the nearest points (Fig. 5).
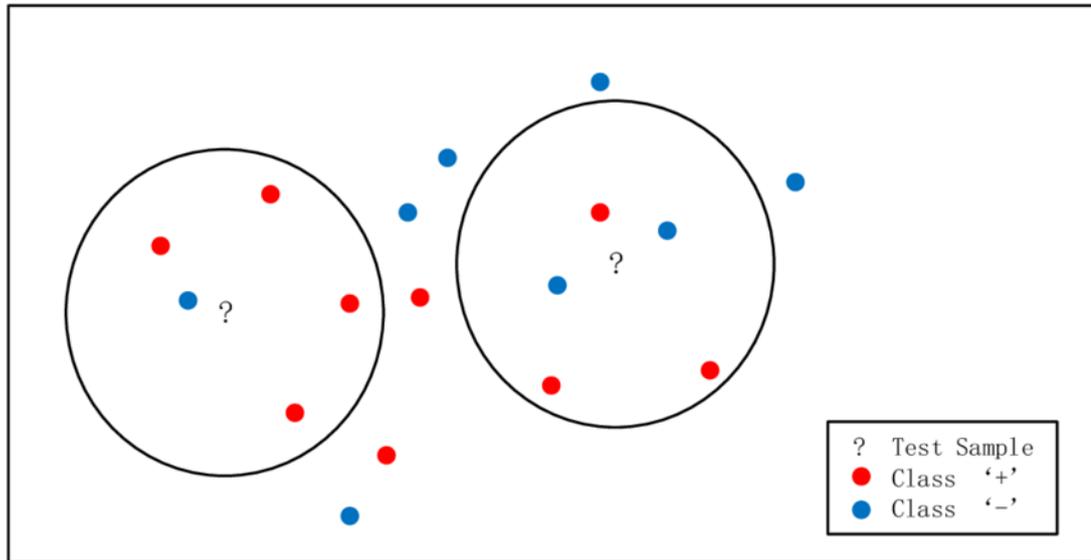
Fig. 5: Example k-nearest neighbor classifier, separating classes ('+' and '-') for two unknown samples (Cheng et al. 2014).

KNNs have been used to discriminate plant diseases with visible foliar symptoms (Pujari et al. 2013b; Zhang et al. 2015; Joshi & Jadhav 2016; Pinto et al. 2016; Prasad et al. 2016). Additionally, KNN classification has been shown to be beneficial in disease severity classification, with various fungal infections classified into normal (no disease), partial, and severe classes in mango, pomegranate and grape (Pujari et al. 2013b). Reported accuracies achieved were greater than 94%, indicating potential for not only disease detection and identification, but additionally disease quantification, suggesting phenotyping potential. The ability to accurately quantify disease severity is important in developing sensing technologies to aid plant breeders in developing resistant germplasm. However, more work should be conducted to determine use in discriminating non-visible spectral features.

2.3 DISEASE DISCRIMINATION – OVERFITTING

A system with high accuracy and high spectral resolution, may perform differently in differing environments, for example, a glasshouse versus a field environment. This is often due to data overfitting. Overfitting occurs when modeling approaches are more complex than required to solve a general problem, such as using a non-linear approach

like an artificial neural network (ANN) with a dataset that is linear, or including gratuitous features which can lead to over specificity (Hawkins 2004). However, there are many techniques for crop disease discrimination, ranging from basic waveband ratios, to more complex machine learning algorithms to provide optimal disease detection capability for a given disease problem while minimising overfitting errors.

3 CONCLUSIONS

Many successful machine-vision disease detection platforms have been developed in glasshouse and laboratory environments (West et al. 2007; Wu et al. 2008; Mahlein et al. 2010; Mahlein et al. 2013; Thomas et al. 2017). However, when transferring these systems into field environments, additional considerations must be in place. These considerations include determining the optimal balance between adequate resolution and the ability to scale the system for detection at an operational level (Franke & Menz 2007; Garcia-Ruiz et al. 2013). Additionally, energy requirements must be considered. Certain technologies are better suited to a controlled laboratory or glasshouse environment, such as, chlorophyll-fluorescence imaging, which takes a large amount of energy to operate (Chaerle et al. 2004; Chaerle et al. 2007; Bauriegel & Herppich 2014). Another consideration is the amount of data processing required for the application. Field experiments and commercial applications often introduce more data than glasshouse or laboratory experiments. This can affect the choice in processing algorithms, often dependent on whether real-time detection or quantification is desirable or plausible.

A majority of successful crop disease detection systems rely on bands within the VIS-NIR range. Bands within the NIR spectrum have been reported to correlate to plant structural and functional molecules (starches, proteins, water, etc.) (Kumar et al. 2002). These correlations make these wavebands good candidates for the detection and quantification of visibly asymptomatic plant diseases or plant diseases with few symptoms or where detection is desired early in the disease cycle. Visibly asymptomatic detection capability has been reported in barley, with powdery mildew (*Blumeria graminis* f. sp. *hordei*) detection occurring two days before visible symptoms appeared (Thomas et al. 2017). The thermal infrared spectrum has also been investigated for crop disease detection potential, but was reported to not be viable without additional sensors (e.g. multispectral, hyperspectral) (Oerke et al. 2006; Lenthe et al. 2007). Another potential spectrum of interest in disease detection and

quantification is the ultraviolet spectrum. Literature indicates that a relationship between bands in the UV spectrum relating to carotenoid levels and chlorophyll in the visible spectrum could potentially be used for disease detection (Penuelas et al. 1995; Yang et al. 2010). However, as this relationship relies on bands within the visible spectrum, it is likely not ideal for detection of many soil and stubble-borne pathogens, which may exhibit few visible symptoms or where visible symptoms are difficult to detect.

By reducing the number of wavebands required for disease discrimination, processing time can be improved while also reducing overfitting errors (Bravo et al. 2003; Hawkins 2004; Mewes et al. 2011). Successful multispectral systems have been developed from hyperspectral feature derivation (Bravo et al. 2003; Mewes et al. 2011). Further to the reduction of processing time and overfitting errors, fewer features equates to less expensive multispectral systems.

While hyperspectral imaging is likely important for crop disease detection, data processing is often burdensome. Data transformation techniques such as the DWT and PCA have the potential to enable the practical use of hyperspectral data. Both PCA and DWTs have been reported to be successful in feature selection and data reduction for crop sensing systems (Bruce & Li 2001; Bruce et al. 2002; Koger et al. 2003; Ge & Thomasson 2006; Ge et al. 2007; Suzuki et al. 2008). The principal components in PCA are uncorrelated, maximising variance between components or features (Jolliffe & Cadima 2016). However, PCA also has no regard for data labels, which may make it miss important feature patterns (Behmann et al. 2015). DWTs are advantageous in the regard that they can produce features at both local and global scales (Bruce & Li 2001). These larger features can be used to reduce the impact of noise across environments, particularly in field data. Further, DWTs have been reported to repeatedly outperform traditional data reduction techniques (Bruce & Li 2001; Bruce et al. 2002; Koger et al. 2003; Ge & Thomasson 2006; Ge et al. 2007). The ability to scale machine vision systems to the operation and task required is important for precision agriculture. Advances in and novel uses of pre-processing techniques, such as DWTs show potential for accurate scaling of vision systems to various agriculture applications (West et al. 2007; Sui et al. 2008). Feature selection and data reduction is only the first step in successful algorithm development.

Many successful crop disease detection systems rely on SVIs. However, these indices are often not stress or disease specific. While it has been reported that combinations of SVIs can be used to represent unique disease signatures, literature suggests that this approach is likely to be more successful in differentiating between biotic and abiotic stresses (Mahlein et al. 2010). However, success has been reported in creating disease specific SVIs (Mahlein et al. 2013). Further, the use of SVIs in successful airborne field applications has been reported (Huang et al. 2007). While developing specific SVIs for visibly asymptomatic disease is promising, disease signatures that are non-linear may be better explained with other processing algorithms.

Machine learning is beginning to move to the forefront of crop disease literature (Wang et al. 2008; Wu et al. 2008; Liu et al. 2010; Moshou et al. 2011; Heim et al. 2018; Heim et al. 2019). Additionally, artificial neural networks have been reported to improve disease detection capability when used in conjunction with PCA (Liu et al. 2010). ANNs have been used to successfully develop a real-time, in-field disease detection platform for monitoring yellow rust in wheat (Moshou et al. 2011). However, logistic regression, support vector machines, decision tree, random forest and k-nearest neighbor classifiers have also all been reported to have potential in crop disease discrimination (Franke & Menz 2007; Rumpf et al. 2010; Prabhakar et al. 2013; Pujari et al. 2013a; Pujari et al. 2013b; Pinto et al. 2016; Adam et al. 2017; Dhau et al. 2018). Tree-based classifiers must be considered carefully, however, as they can easily fall to the problem of overfitting (Hawkins 2004).

In order to develop a successful CR detection and quantification framework that can be further refined into a high throughput phenotyping (HTP) methodology, several points should be considered. Firstly, the system must be able to accurately detect low-symptomatic stubble-borne pathogens. With correlation to plant molecular structure, the NIR spectrum seems the ideal candidate to examine any low-symptomatic crop disease. Secondly, the devised system should be able to detect infection early, in order to give breeding companies as much time as possible to focus breeding efforts on promising lines. Thirdly, the system should be able to discriminate disease, whether or not the relationship between disease and reflectance is linear, as such, non-linear methods should be investigated. Finally, developed algorithms should be as efficient as possible, meaning both PCA and DWT approaches should be evaluated for overall system accuracy and number of derived features. The fewer features needed to retain

accuracy, the lower the cost of developed multispectral system, potentially enabling real-time sensing applications.

Rapid and effective crop stress assessment is vital for making informed decisions in commercial farming operations and plant breeding programs. Inexpensive high-resolution, machine vision systems have the potential to enable precision stress management both at the individual plant and paddock levels. Multispectral and hyperspectral imaging systems in conjunction with DWT-based feature selection and machine learning techniques may contribute to plant breeding programs through high-throughput phenotyping, early disease detection and management and quality monitoring. The DWT presents advantages over traditional hyperspectral preprocessing techniques including improvements in spectral band selection, number of features, computational efficiency and original curve representation which outweigh any limitations introduced by shift-variance.

4.1 FOCUS AND AIMS OF THE THESIS

4.1.1 PROJECT AIM

The primary aim of this thesis is to discover the feasibility of using a DWT analysis approach to detect and discriminate CR in wheat and phenotype variety resistance in the near infrared spectrum (900–1700 nm). Specifically, using the multi-resolution capability of DWT models, the goal is the discrimination of stresses within a scale that can be generalised to diagnose crop stress conditions from a multispectral imaging sensor utilising machine learning techniques. Additionally, the goal is to provide evidence that differentiation can be determined with similar accuracy and fewer features than traditional analysis techniques (PCA). The use of fewer features potentially allows for the model to have benefits in multiple environments by reducing overfitting of models to spectral features not directly associated with wheat CR.

4.1.2 RESEARCH QUESTION

Can a combination of wavelet analysis and machine learning techniques be used to analyse signatures in the near-infrared spectrum to develop a sensor-based framework for rapid CR disease discrimination or phenotyping in wheat?

4.1.3 RESEARCH OBJECTIVES

1. Investigate available wavelet-based techniques for discriminating crop disease in the literature.

2. Determine if a DWT model can outperform traditional analysis techniques (PCA) in the discrimination of CR of wheat in the near-infrared spectrum (900–1700 nm).

3. Develop models to identify whether bandwidths corresponding to available commercial near-infrared filter sizes can discriminate and phenotype CR of wheat.

4. Use commercially available filters on an imaging sensor to determine if the developed models can be utilised in an imaging system.

CHAPTER 3

MACHINE LEARNING FOR THE DETECTION OF CROWN ROT, *FUSARIUM PSEUDOGRAMINEARUM*, PART A: PCA AND DISCRETE WAVELET APPROACHES

In this study, principal component analysis (PCA) and the discrete wavelet transform (DWT) were compared for ability to increase machine learning model disease prediction accuracy across six machine learning methods (logistic regression, k-nearest neighbors, decision trees, extreme random forests, support vector machines, artificial neural networks), for the detection of *Fusarium pseudograminearum* induced crown rot. ANOVA F-values were used to select the optimal wavelet features. Two train-test data splits were evaluated for effectiveness in model development, 80-20% and 20-80%. Models were evaluated in three-week groupings, in addition to weekly glasshouse, field and combined groupings.

Humpal J., McCarthy C., Percy C., & Thomasson AJ. (2020a). Machine learning for the detection of crown rot, *Fusarium pseudograminearum*, Part A: PCA and discrete wavelet approaches. This chapter was prepared according to the instructions to authors given by **Computers and Electronics in Agriculture.**

Machine learning for the detection of crown rot, *Fusarium pseudograminearum*, Part A: PCA and discrete wavelet approaches

Jacob Humpal[*], Cheryl McCarthy[*], Cassandra Percy[**], J. Alex Thomasson[***]

[*] Centre for Agricultural Engineering, University of Southern Queensland, Toowoomba, Queensland, Australia

[**] Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia

[***] Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, Mississippi, United States

J. Humpal, C. McCarthy (*)

Centre for Agricultural Engineering, University of Southern Queensland,

Toowoomba, Qld 4350, Australia

*e-mail: cheryl.mccarthy@usq.edu.au

Ph: +61 7 4631 2297

Abstract- In this paper, the discrete wavelet transform (DWT) and traditional principal component analysis (PCA) are compared for use in crop disease model development from near-infrared hyperspectral data. The unique scaling ability of the discrete wavelet transform is discussed and applied to hyperspectral data obtained from five crown rot (CR) (*Fusarium pseudograminearum*) resistance trials, in bread wheat (*Triticum aestivum*). Data features are extracted using six supervised learning techniques and are scored using 5-fold cross-validation and F-scores. The results are compared to PCA evaluation on artificially derived wavebands of the same bandwidths as those obtained through the DWT. Prediction accuracy is evaluated using both a 20-80% train-test split and an 80-20% train-test split. The results show promise for automated classification of disease with poor visible markers, using both wavelet and PCA-derived models. The wavelet-based classification system outperforms the PCA-based system in 67% of 36 test cases when training on 20% of the disease data with both systems performing best in 50% of the 36 test cases on the 80-20% data split. PCA-based classification utilised fewer overall features in both training sets.

I. Introduction

CR occurs across the globe in many arid and semi-arid cropping regions (Kazan & Gardiner 2018). CR infection occurs when *F. pseudograminearum* hyphae fragments from grass and cereal residues in close contact with a host receive sufficient moisture for activation. The fungus colonises the local area of infection and proceeds to progress through the crown tissues, including leaf sheaths, sub crown internodes and stems (Fig. 1) (Percy et al. 2012; Alahmad et al. 2018). Severe CR can lead to necrosis of the stem, significantly limiting grain production. *F. pseudograminearum* can be responsible for a 50% or greater yield reduction in bread wheat (*Triticum aestivum*) grown under conducive conditions in Australia (Klein et al. 1991). The lack of readily discernible visible symptoms until late in the season with the extension of stem browning up the tillers and the appearance of white heads causes delays in production decision making. Improvements in productivity, specifically disease identification and decision making, will help growers remain profitable and operations sustainable.

Historically, plant disease identification and quantification have been left to human assessors. Human data collection is known to show variability, particularly between multiple raters (Nutter Jr et al. 1993; Newton & Hackett 1994). The inter-rater variation commonly seen in crop disease quantification is thought to be due to imperfect scoring systems that leave ratings open to interpretation (Nutter 1997), reducing the likelihood that two raters will score a plant identically with high repeatability. Traditional CR assessment involves removing the plant from the field, stripping the plant of the leaf sheaths around the lower internodes and scoring symptoms based on colour variation of the stem. This reliance on human visual colour assessment creates difficulties in CR quantification particularly in inter-rater consistency similar to the difficulties seen in other systems (Nutter Jr et al. 1993; Newton & Hackett 1994; Nutter 1997). Further, these human-based scoring systems are labour intensive. The use of sensors in combination with machine learning systems to distinguish disease offers opportunity to improve the assessment of CR. Machine vision has the potential to alleviate many of the difficulties of CR visual assessment, including accuracy, precision and efficiency around assigning germplasm resistance rankings. In this paper, non-visual host responses, indicative of molecular changes in the near-infrared spectrum are evaluated for potential to provide an automated sensing technology and reduce the reliance on manual scoring methods.
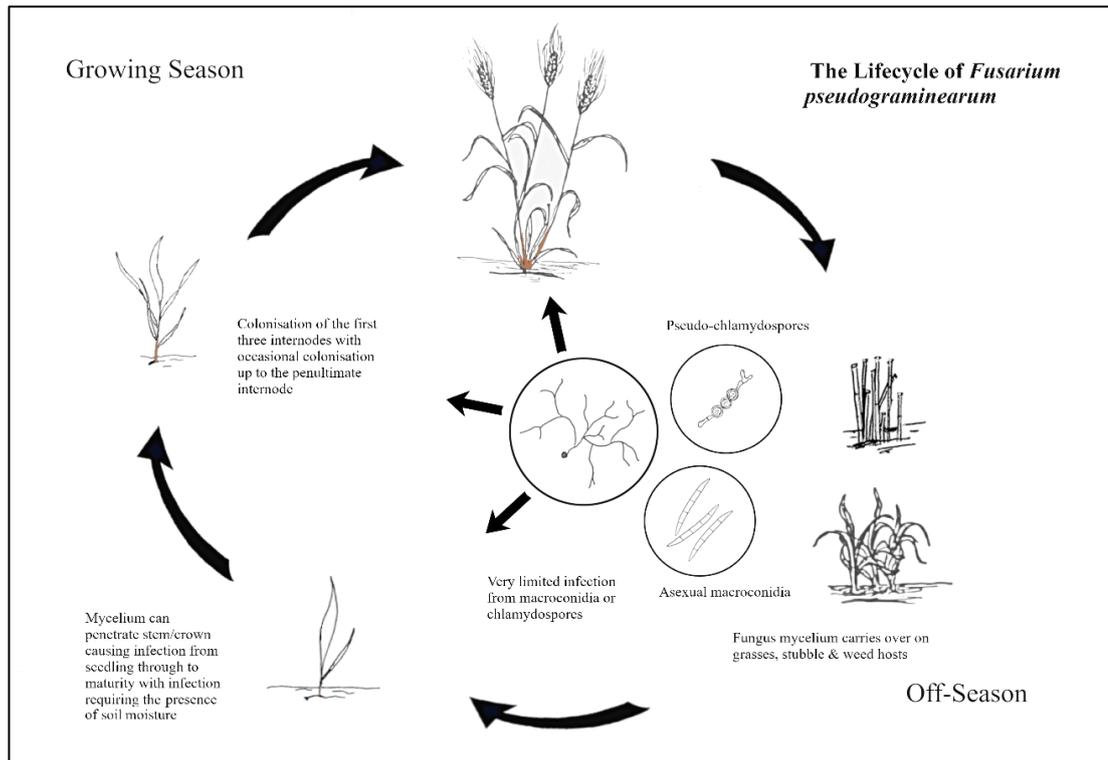
Fig. 1. The lifecycle of *F. pseudograminearum* including the growing season and off-season phases.

Spectral reflectance imaging refers to the "acquisition of a series of digital images at a number of different, well-defined optical wavelengths" (Klein et al. 2008), such that a contiguous spectral signal is obtained for each pixel within the image series. The light waves associated with the point on the electromagnetic spectrum vary in wavelength and are referred to as separate wavebands. Reflectance imaging has been reported to be useful in differentiating plant stressors in laboratory settings (Oerke et al. 2011; Mahlein et al. 2013; Grieve et al. 2015; Thomas et al. 2017), in addition to the identification of stresses in agricultural environments (Franke & Menz 2007; Huang et al. 2007; Garcia-Ruiz et al. 2013; Polder et al. 2014). A currently popular reflectance imaging approach is the use of hyperspectral sensors (Bergsträsser et al. 2015; Thomas et al. 2017; Heim et al. 2018). Hyperspectral sensors obtain hundreds of contiguous wavebands; however, the high-resolution data obtained from these sensors can impede interpretation accuracy by increasing the potential for over-fitting of developed models (Hawkins 2004). Traditional data analysis techniques cause derived disease models to be less robust across trials, due to the resolution of the data used in the creation of these models (Sankaran et al. 2010).

The discrete wavelet transform (DWT) is advantageous in respect to signal space-scale, as the DWT allows for recognition of features at both local and global spectral scales (Kempeneers et al. 2005; West et al. 2007). Therefore, a wavelet analysis approach would allow for multiple data resolutions to be used when creating models from the determined regressors, or independent variables, that can be used for the prediction of features within the data (Bruce et al. 2002). The multiple data resolutions generated using a wavelet analysis approach provide potential for novel application of the DWT to the detection of CR. Multiple resolution models would permit improved multispectral crop sensor development, as the determined features allow for in-field systems targeting specific wavebands to be developed (Kempeneers et al. 2005; West et al. 2007).

This paper is Part A of a series of papers that describe the development of disease detection models for the identification and quantification of CR in wheat. Part A investigates model development and compares the use of PCA and DWT approaches for preprocessing of spectral data before processing with machine learning algorithms on a weekly basis, with each week of each experiment treated as an independent dataset. The paper will further evaluate the effect of different train-test sizes, as smaller training sizes may be desirable to reduce computation costs but may limit the accuracy of the developed models. Part B compares performance of models developed on combined datasets to introduce increased variation into the models with the goal of increasing model robustness, while Part C investigates temporal and spatial considerations in developing disease models.

II. Materials and Methods

A. Glasshouse and field sites, plant cultivation and experimental design

Glasshouse trials were established at the Leslie Research Centre, Rockville, QLD in 2018 and at USQ's Centre for Crop Health, Darling Heights, QLD in 2019, in collaboration with researchers from the Centre for Crop Health. In the initial glasshouse trial conducted at the Leslie Research Centre, four wheat genotypes with varying resistances to CR (referred to as standards) were observed with positive or null inoculation with *F. pseudograminearum*. Twelve replicates for a total of 96 plants were observed using a near-infrared point sensor. The second glasshouse trial at the Leslie Research Centre consisted of the original four genotypes with the addition of

one standard genotype, for a total of five genotypes replicated six times, totaling 60 plants in this trial. The third glasshouse trial was conducted at the Centre for Crop Health glasshouses in 2019 and replicated the second glasshouse trial. All pots in each glasshouse trial were arranged in randomised block designs with plants watered to field capacity once per week, and temperature being maintained at 20–25° C to ensure minimal outside discrepancies between pots. *F. pseudograminearum* colonised wheat grain inoculum (Percy et al. 2012) was applied to individual coleoptiles of plants at the two-leaf stage. Plants were harvested and rated for disease at maturity.

An initial field trial was planted at the Tosari research station (-27.859964, 151.452766), south of Toowoomba, Queensland, in June of 2018. Inoculated and non-inoculated paired plots were replicated 3 times in a randomised strip plot design. Six plants from each plot were chosen at random in each of the five genotypes corresponding to the glasshouse trials. A replication of the Tosari field trial from 2018 was conducted in 2019. *F. pseudograminearum* colonised millet inoculum was applied in the furrow above the seed at planting to inoculated plots in both field trials. The observed plants were pulled at maturity and scored manually for the presence and severity of CR at the Centre for Crop Health. Trained assessors scored all of the material after harvest and did not report the presence of additional diseases or natural infection of crown rot in the non-inoculated plants. All glasshouse and field trials were statistically designed in collaboration with biostatisticians at Queensland Department of Agriculture and Fisheries (QDAF).

B. Hyperspectral reflectance measurements

Near-infrared measurements of components of the crop canopy were recorded in the 900–1700 nm range using the handheld DLP® NIRscan™ Nano, hyperspectral sensor (Texas Instruments, USA), for plants inoculated with *F. pseudograminearum* and those left as null pots/plots in all trials. The Nano is a point-based hyperspectral sensor with an Indium Gallium Arsenide (InGaAs) detector with a shape of 854×480, pixel pitch of 5.4 µm and a 10 nm spectral resolution. This module uses two tungsten lamps as an integrated light source and produces a signal to noise ratio of 6000:1. Measurements were taken once a week throughout the growing season from Zadoks (Z) 13, 21 (Zadoks et al. 1974), for nine weeks. Week 8 in the glasshouse 2 trial is missing due to technical issues at collection. The maximum distance between all other measurement dates was 8 days. Readings were collected on contact from the center of

the leaf of the newest emerged tiller, the centre-most leaf and the newest flag leaf. The spectral signatures were recorded and output from the device as CSV files. Calibration references were collected from 10% grey, 60% grey and 99% white reference Spectralon® panels. Measurements were corrected as described below:

$$Corrected\ Reflectance = Calibration\ Coefficient \left(\frac{Target\ Radiance}{Reference} \right)$$

(Labsphere 1998)

C. Data grouping

Data was grouped into twelve distinct groups, six for each of the 20-80% and 80-20%, train-test splits. Data was analysed per each week for both exclusively glasshouse and field data, denoted as 'Weekly Glasshouse' and 'Weekly Field', as well as, combined weekly data from all trials, 'Weekly – Overall'. Data from all trials combined was also split into three, tri-weekly groups, from the identification of Z 13, 21. These groups are denoted as 'Weeks 1–3', 'Weeks 4–6' and 'Weeks 7–9'. Data was grouped by site to allow comparison of weekly models developed using data from only field sites or only glasshouse sites. The 'Weekly – Overall' grouping allowed the comparison of these weekly models to models using combined data for each week. The separation of data into tri-weekly groups was used to examine the impact of time-window on the detection capability of the models.

D. Data analysis

All data analysis was conducted in the Python computing environment (Python version 3.6.8, Python Software Foundation 2019) primarily using the SciPy ecosystem (Jones et al. 2001) and the Scikit-learn library (Pedregosa et al. 2011). Additional libraries and packages have been cited where relevant in the following sections.

D.1 PCA approach

Raw spectra of adjacent wavebands were averaged to approximate the wider bandwidths produced in the wavelet approach. These spectra were standardised and scaled to 0–1 using the preprocessing package in the Python SciPy library, version 0.19.1 (Jones et al. 2001). Spectra were smoothed using the Savitzky-Golay filter (Savitzky & Golay 1964) with a window size of 5 bands and a second order

polynomial, using the savgol_filter in the Python library, SciPy (Jones et al. 2001). Principal component analysis (PCA) was performed on the transformed signals and constrained to 1 to 10 principal components using the decomposition package in the Python SciPy library (Jones et al. 2001). These principal components were used to determine the features of maximum importance within the input signal set. The PCA transformed dataset was used in the machine learning algorithms.

D.2 Wavelet approach

The wavelet approach (1) uses a scaled operator to determine regressors at multiple resolutions:

$$\omega_s f(\lambda) = f(\lambda) * \psi_s(\lambda) = f(\lambda) * \frac{1}{s} \psi\left(\frac{\lambda}{s}\right) \tag{1}$$

(Bruce & Li 2001)

where $f(\lambda)$ is the hyperspectral signature, s is the scale and $\psi_s(\lambda)$ is the scaled wavelet function. This presents an advantage over traditional methods which require an operator derived from the smoothing operation (Bruce et al. 2002). Raw spectra were inspected for outliers (z score > 3) and transformed using the discrete wavelet transform (DWT) with a Daubechies 2 wavelet from the PyWavelets library, version 1.0.3 (Lee et al. 2019). The Daubechies 2 wavelet is part of the Daubechies family of wavelets (Daubechies 1988) and is asymmetric and biorthogonal (Fig. 2).
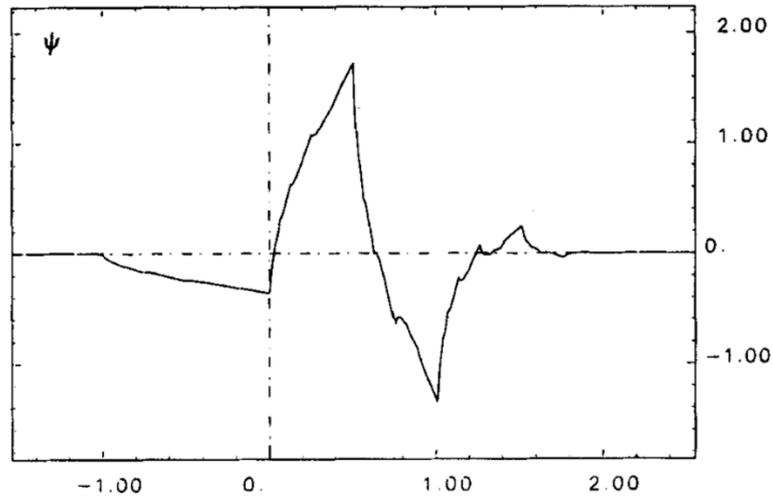
Fig. 2. The Daubechies 2 wavelet (Daubechies, 1988).

Resulting derived spectral bandwidths were 7.02, 14.04, 27.59, 53.33, 100.00, 200.00, 400.00 and 800.00 nm. Spectra were standardised and scaled to 0–1 using the preprocessing package in the Python SciPy library version 0.19.1 (Jones et al. 2001).

D.3 Feature selection

Wavebands with bandwidths less than 25 nm were discarded before feature selection in this study for both glasshouse and field trials, for the intended purpose of increasing model robustness across environments. Features of maximum importance represent key wavelengths in the hyperspectral signature that are highly correlated to differences between the signatures of inoculated and non-inoculated plants. Features for the PCA approach were selected based on the top 1–10 features of most importance in determining the principal components. Wavelet feature selection was determined using the SelectKBest function from the feature_selection package of the Python Scikit-learn library with a selection classifier of f_classif which selects features based upon computed k highest ANOVA F-values (Pedregosa et al. 2011). Models were built using the top k, 1–10 features selected based upon this function.

D.4 Machine learning techniques

Six machine learning techniques were evaluated for effectiveness in discriminating *F. pseudograminearum* induced CR: logistic regression classification, k-nearest neighbors, decision trees, extreme random forests, support vector machines and artificial neural networks, based on the original algorithms presented in Verhulst (1838), Fix (1951), Breiman et al. (1984), Geurts et al. (2006), Вапник and Червонекис (Vapnik and Chervonekis) (1964) and McCulloch and Pitts (1943). These methods are described in the following paragraphs.

Logistic regression is a statistical modelling approach that uses a logistic function to model a binary dependent variable, a common example being, pass/fail. Mathematically, a binary logistic model has a dependent variable with two possible values which are represented by a variable labeled as "0" or "1".

K-nearest neighbors classification is an instance-based or non-generalised learning method, meaning this method does not construct an internal model, but instead stores instances of the training data. Classification is assigned based on a majority vote of the nearest neighbors of each point.

Decision trees create a model that predicts the value of a target variable based on several input variables. In these models each branch represents combinations of features that lead to an outcome or class label, the leaves of the tree. An extension of decision tree classification systems are random forests and extreme random forests. Random forests are an ensemble of individual decision trees. In extreme random forests each individual tree is trained using the whole training dataset and the feature splits are randomised. Instead of computing the optimal decision point for each feature, a random point is selected. The decision point that yields the highest accuracy in predicting the class label is chosen to split the branch.

A support vector machine constructs a hyper-plane or set of hyper-planes in a high or infinite dimensional space, which can be used for classification or regression. Separation is achieved by the hyper-plane that has the largest distance to the nearest training data points.

Artificial neural networks (ANN) are classification systems that are inspired by neural networks that constitute animal brains. Specifically, the neurons in a biological nervous system. Each "neuron" can transmit to and receive signals from other neurons. ANNs "learn" to perform tasks by analysing training data without being programmed with specific rules. These networks automatically discover unique characteristics from the input training dataset.

D.5 Model evaluation and validation

Data for each of three glasshouse and two field trials was evaluated using the PCA and wavelet methods. Each of the resulting datasets were split into training and test sets of 20-80% and 80-20% respective splits. These splits were random independent validation splits performed using the train_test_split function in Scikit-learn (Pedregosa et al. 2011). The input feature sets were trained for each combination of hyperparameters (Appendix E.1) determined using the GridSearchCV function from Scikit-learn's model_selection package (Pedregosa et al. 2011). The GridSearchCV function performs an exhaustive search over specified parameter values for an estimator, optimised by 5-fold cross-validation. The resulting models were scored on accuracy and F_1/F-score. Additionally, the number of features included in the top performing models was considered. The equations used for accuracy, precision, recall and F_1 are reported below (2,3,4,5):

$$accuracy\left(y, \hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \cdot 1\left(\hat{y}_i = y_i\right) \qquad (2)$$

Where $\hat{y}_i$ is the predicted value of the $i^{\,th}$ sample and y_i is the corresponding true value

$$Precision = \frac{tp}{tp + fp} \qquad\qquad (3)$$

Where tp is number of true positive values and fp is number of false positive values

$$Recall = \frac{tp}{tp + fn} \qquad\qquad (4)$$

Where tp is number of true positive values and fn is number of false negative values

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \qquad (5)$$

III. RESULTS

A. Binary classification into inoculated and non-inoculated plants: accuracy

Data is presented in both tables and figures throughout the results to clearly show minimum-maximum value ranges in addition to highlighting differences between the datasets. The wavelet-based models returned higher maximum and minimum accuracies than PCA analysis in all scenarios under the 20-80% data split, while PCA-based models returned higher maximum accuracies across all scenarios under the 80-20% data split, although DWT-models retained higher minimums. From Table 1, wavelet models returned maximum accuracies of 59.10% to 70.48% under the 20-80% split, while PCA returned accuracies of 56.88% to 68.33%. Under the 80-20% data split, PCA returned higher maximum accuracies of 64.49% to 76.88%, with wavelet model maximums from 63.23% to 75.68% (Table 1). Wavelet models returned higher minimum accuracies across all scenarios under the 80-20% split (Table 1, Fig. 3, 4).

In the 20% training, 80% testing dataset, PCA analysis outperformed wavelet analysis in extreme random forest classification in all groupings. DWT analysis performed better than PCA in 100% of artificial neural network classification, support vector classification, logistic regression classification and k-nearest neighbor classification cases across all groupings (Fig. 3, Table 2). In the 80% training, 20% testing dataset, DWT analysis outperformed PCA analysis in all groupings in support vector

classification and k-nearest neighbor classification. PCA outperformed wavelet analysis in the tree-based classification methods, decision tree classification and extreme random forest classification, across all groupings (Fig. 4, Table 2). The highest average accuracy achieved was 76.88% using an extreme random forest classifier with the PCA approach on an 80-20% split of weekly glasshouse data, while the lowest average accuracy achieved was 50.31%, also using the PCA approach. The minimum accuracy was attained using the 'Weeks 7–9' temporal grouping on a 20-80% train-test split with a support vector classifier (Table 1, Fig. 3, 4).

Table 1: Minimum and maximum ranges of average classification accuracies (%) of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% or 80-20% train-test split.

Data split	Temporal grouping	DWT		PCA	
		Minimum	Maximum	Minimum	Maximum
20-80	Weeks 1–3 (n=1512)	56.19	66.80	54.60	64.86
	Weeks 4–6 (n=1512)	57.11	66.95	51.75	64.27
	Weeks 7–9 (n=1452)	53.07	63.12	50.31	61.79
	Weekly Glasshouse (n=1884)	56.79	70.48	52.75	68.33
	Weekly Field (n=2592)	53.57	59.10	51.35	56.88
	Weekly – Overall (n=4476)	55.48	65.82	52.18	63.51
80-20	Weeks 1–3 (n=1512)	62.69	72.66	58.09	72.81
	Weeks 4–6 (n=1512)	60.08	71.85	56.18	73.39
	Weeks 7–9 (n=1452)	56.41	66.78	51.48	69.23
	Weekly Glasshouse (n=1884)	62.55	75.68	58.38	76.88
	Weekly Field (n=2592)	55.80	63.23	50.51	64.49
	Weekly – Overall (n=4476)	59.79	70.59	55.16	71.81
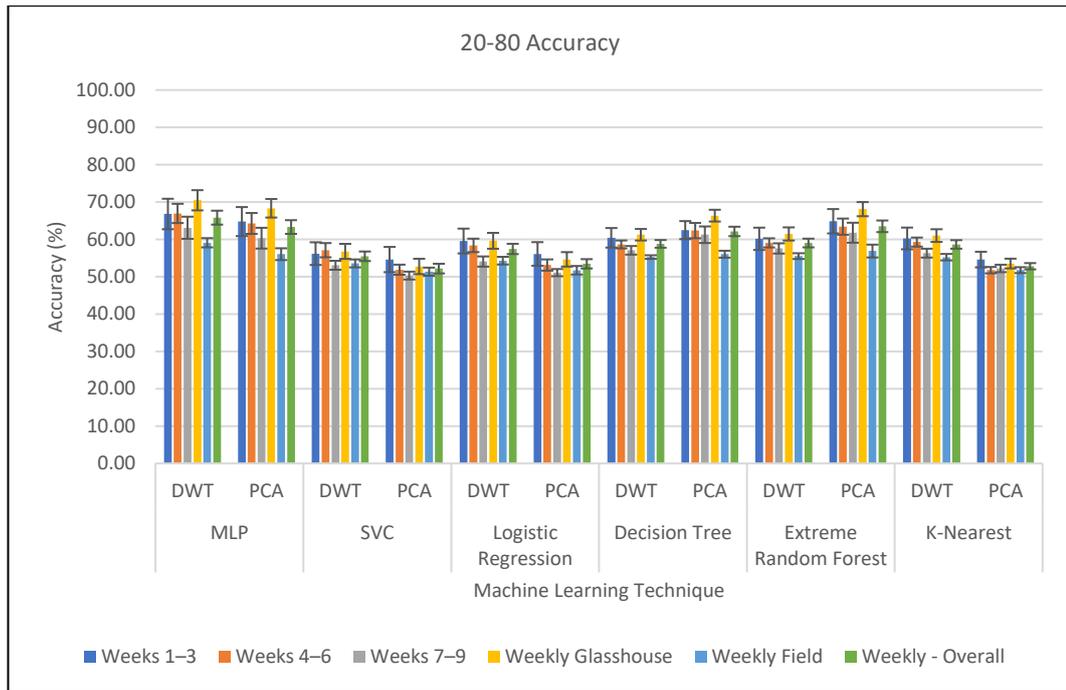
Fig. 3: Average classification accuracies (%) of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split.
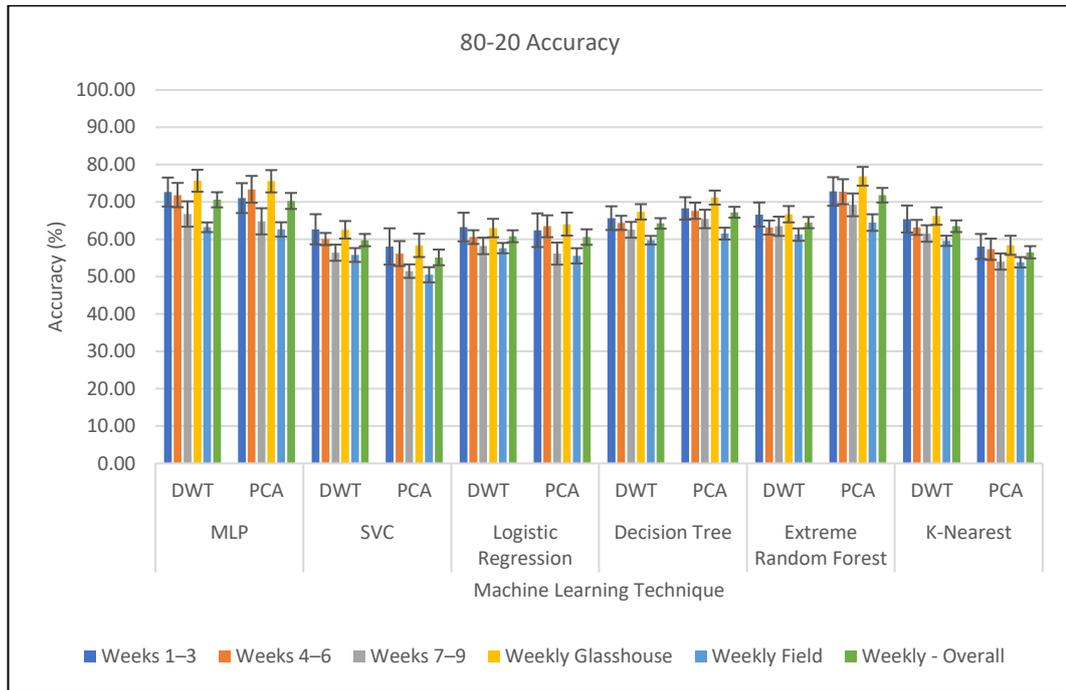
Fig. 4: Average classification accuracies (%) of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split.

Table 2: Percentage of instances in which each system (DWT or PCA) outperformed the others in CR detection based upon accuracy percentage for both a 20-80% and an 80-20% train-test split.

		MLP		SVC		Logistic Regression		Decision Tree		Extreme Random Forest		K-Nearest	
Data split	Temporal grouping	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
20-80	Weeks 1–3 (n=1512)	80.00	20.00	80.00	20.00	100.00	0.00	40.00	60.00	0.00	100.00	100.00	0.00
	Weeks 4–6 (n=1512)	80.00	20.00	80.00	20.00	80.00	20.00	20.00	80.00	20.00	80.00	100.00	0.00
	Weeks 7–9 (n=1452)	100.00	0.00	80.00	20.00	80.00	20.00	0.00	100.00	0.00	100.00	100.00	0.00
	Weekly Glasshouse (n=1884)	69.23	30.77	73.08	26.92	73.08	26.92	23.08	76.92	23.08	76.92	92.31	7.69
	Weekly Field (n=2592)	77.78	22.22	77.78	22.22	72.22	27.78	50.00	50.00	44.44	55.56	88.89	11.11
	Weekly – Overall (n=4476)	72.73	27.27	75.00	25.00	72.73	27.27	36.36	63.64	31.82	68.18	90.91	9.09
80-20	Weeks 1–3 (n=1512)	80.00	20.00	80.00	20.00	60.00	40.00	40.00	60.00	0.00	100.00	100.00	0.00
	Weeks 4–6 (n=1512)	20.00	80.00	60.00	40.00	40.00	60.00	0.00	100.00	20.00	80.00	80.00	20.00
	Weeks 7–9 (n=1452)	40.00	60.00	100.00	0.00	60.00	40.00	20.00	80.00	20.00	80.00	100.00	0.00
	Weekly Glasshouse (n=1884)	44.00	56.00	65.38	34.62	42.31	57.69	26.92	73.08	3.85	96.15	84.00	16.00
	Weekly Field (n=2592)	61.11	38.89	83.33	16.67	55.56	44.44	44.44	55.56	44.44	55.56	94.44	5.56
	Weekly – Overall (n=4476)	50.00	47.73	72.73	27.27	47.73	52.27	34.09	65.91	20.45	79.55	88.64	11.36

B. Binary classification into healthy and diseased plants: precision and recall

Wavelet-based models returned higher maximum F_1 scores than PCA analysis in all scenarios under the 20-80% data split (Table 3). DWT-based models returned higher minimum scores across all data splits and datasets, with the exception of the 'Weeks 7–9' grouping in the 20-80% data split. From Table 3, DWT models returned maximum F_1 scores of 0.6190 to 0.7276 under the 20-80% split, while PCA returned F_1 scores of 0.5646 to 0.6854. Under the 80-20% data split, PCA returned higher maximum F_1 scores of 0.6318 to 0.7606, with wavelet model maximums from 0.6689 to 0.7514 (Table 3). PCA-based models had a larger range between minimum and maximum F_1 scores across all scenarios and both data splits, with the exception of the 'Weeks 7–9' grouping in the 20–80% data split (Table 3, Fig. 5, 6). The highest average F_1 of 0.7606 was achieved using an extreme random forest classifier with the PCA approach on an 80-20% split of weekly glasshouse data, while the lowest average F_1 observed of 0.2809 also used the PCA approach. This is consistent with maximum and minimum average accuracies. The minimum F_1 score was achieved using weekly field data on a 20-80% train-test split with a support vector classifier (Table 3, Fig. 5, 6).

Table 3: Minimum and maximum ranges of average classification F_1* scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% or 80-20% train-test split. The models utilised were a multilayer perceptron (MLP), a support vector classifier (SVC), a logistic regression classifier, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier.

*F_1 - the harmonic mean of precision and recall

Data split	Temporal grouping	DWT		PCA	
		Minimum	Maximum	Minimum	Maximum
20-80	Weeks 1–3 (n=1512)	0.5463	0.6917	0.3222	0.6473
	Weeks 4–6 (n=1512)	0.5239	0.6971	0.3372	0.6406
	Weeks 7–9 (n=1452)	0.3445	0.6632	0.4132	0.6253
	Weekly Glasshouse (n=1884)	0.5082	0.7276	0.3976	0.6854
	Weekly Field (n=2592)	0.4069	0.6190	0.2809	0.5646
	Weekly – Overall (n=4476)	0.4667	0.6832	0.3498	0.6314
80-20	Weeks 1–3 (n=1512)	0.5386	0.7320	0.4454	0.7166
	Weeks 4–6 (n=1512)	0.5698	0.7195	0.4911	0.7270
	Weeks 7–9 (n=1452)	0.5748	0.7009	0.4194	0.6904
	Weekly Glasshouse (n=1884)	0.5535	0.7514	0.4467	0.7606
	Weekly Field (n=2592)	0.5614	0.6689	0.4573	0.6318
	Weekly – Overall (n=4476)	0.5573	0.7177	0.4510	0.7079
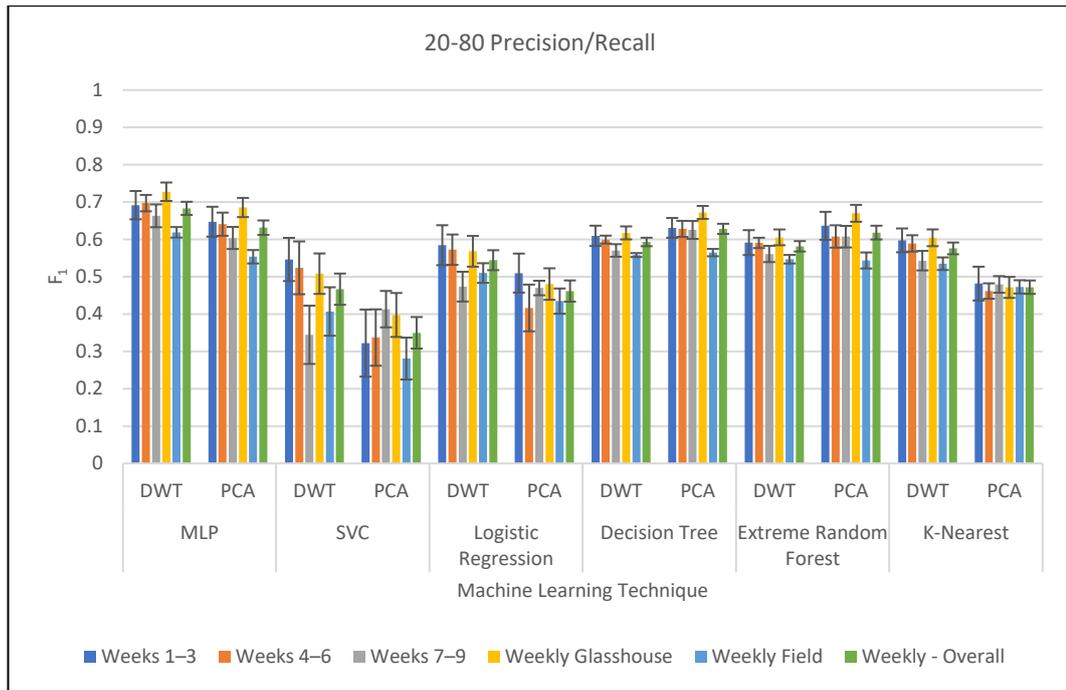
Fig. 5: Average F_1* scores of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split.

*F_1 - the harmonic mean of precision and recall
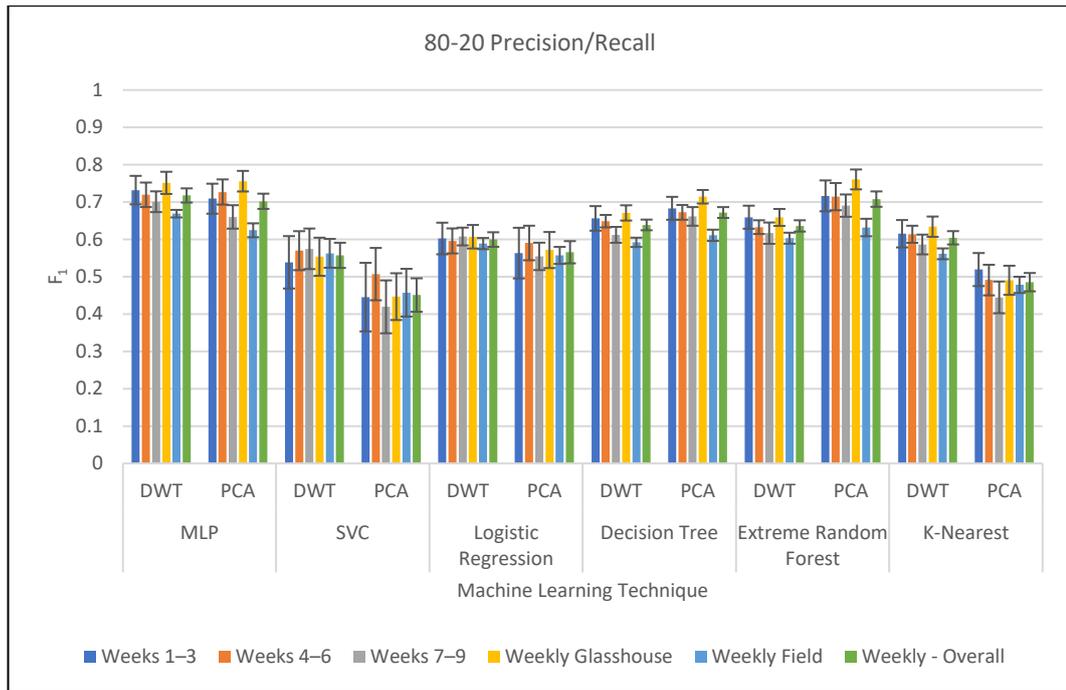
Fig. 6: Average F_1* scores of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split.

*F_1 - the harmonic mean of precision and recall

C. Binary classification into healthy and diseased plants: feature number

The number of features in the models with the highest classification accuracies were recorded. In the 20% training, 80% testing dataset, PCA analysis returned fewer features than wavelet analysis in 67% of test cases, with feature numbers ranging from 5–7. PCA outperformed wavelet analysis in artificial neural network classification, logistic regression classification, decision tree classification and extreme random forest classification, while wavelet analysis performed better in k-nearest neighbor classification. Both techniques performed equally in support vector classification (Table 4). In the 80% training, 20% testing dataset, PCA analysis returned fewer features in optimal models than wavelet analysis in 100% of test cases across all machine learning approaches tested, with total number of features ranging from 4–5 (Table 4).

Table 4. Average number of features chosen, with standard error, between 1 and 10, of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), on a 20-80% or 80-20% train-test split. Accuracies are reported as the best accuracy score on any given week from Zadoks (Z) 13, 21.

	MLP		SVC		Logistic Regression		Decision Tree		Extreme Random Forest		K-Nearest	
Data split	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
20-80	7 ± 0.41	5 ± 0.48	6 ± 0.44	6 ± 0.39	6 ± 0.49	5 ± 0.42	6 ± 0.45	5 ± 0.43	6 ± 0.50	5 ± 0.38	4 ± 0.48	7 ± 0.35
80-20	6 ± 0.40	5 ± 0.46	6 ± 0.44	4 ± 0.40	6 ± 0.44	5 ± 0.43	7 ± 0.38	5 ± 0.41	6 ± 0.42	4 ± 0.38	6 ± 0.41	5 ± 0.43

D. Temporal effects

Average CR detection accuracy and F_1 scores were calculated for each week, from three weeks post-inoculation, across all trials to determine whether a temporal patterning effect appeared as the season progressed. Accuracies and F_1 scores dropped between the 3rd and 4th week of measurement and again between the 7th and 8th week of measurement when trained on the 20-80% and the 80-20% data splits, with the 80-20% split drop in detection capability more pronounced than in the smaller training dataset (Table 5). From Table 5, in the 20-80% data split accuracies and F_1 scores dropped from 61.08 ± 1.411% (0.5778 ± 0.0262) in week 4 to 57.58% ± 1.012% (0.5218 ± 0.0261) in week 6. In the 80-20% data split accuracies and F_1 scores dropped from 68.85 ± 1.865% (0.6567 ± 0.0270) to 61.43% ± 1.563% (0.5833 ± 0.0225).

Table 5. Average accuracies and F_1, with standard errors of CR classification detection and corresponding growth stages (Zadoks et al. 1974), across all trial sites, machine learning models and pre-processing techniques for a 20-80% or 80-20% train-test split.

Data Split	Approximate Growth Stage	Weeks From Zadoks (Z) 13, 21	Accuracy	F_1
20-80	Tillering	1	59.78 ± 2.042	0.6103 ± 0.0235
		2	61.08 ± 1.411	0.5778 ± 0.0262
		3	59.34 ± 1.280	0.5241 ± 0.0273
		4	57.58 ± 1.012	0.5218 ± 0.0261
		5	60.73 ± 0.936	0.5878 ± 0.0203
	Stem Extension	6	58.25 ± 1.203	0.5565 ± 0.0238
		7	56.23 ± 1.125	0.5302 ± 0.0200
		8	55.02 ± 0.756	0.5240 ± 0.0182
	Boot	9	58.24 ± 1.126	0.5167 ± 0.0253
80-20	Tillering	1	64.06 ± 2.381	0.6134 ± 0.0300
		2	68.85 ± 1.865	0.6567 ± 0.0270
		3	63.87 ± 1.449	0.5901 ± 0.0235
		4	61.43 ± 1.563	0.5833 ± 0.0225
		5	67.64 ± 1.282	0.6746 ± 0.0165
	Stem Extension	6	64.46 ± 1.473	0.6128 ± 0.0219
		7	62.12 ± 1.666	0.6166 ± 0.0229
		8	58.00 ± 1.013	0.5686 ± 0.0185
	Boot	9	62.22 ± 1.447	0.5786 ± 0.0221

IV. DISCUSSION

A. Crown rot detection

Current CR detection methods are invasive and time-consuming, particularly under field conditions. A rapid, non-invasive approach is ideal to improve management decisions, reduce input costs to growers, decrease labour and production time and improve the accuracy of phenotyping for researchers and breeders. Near-infrared spectroscopy has the potential to detect crop disease before visible symptoms appear, increasing available decision time (Kumar et al. 2002).

Our results illustrate successful disease discrimination using machine learning and near-infrared sensing technologies to detect *F. pseudograminearum* induced CR early in the disease lifecycle, as soon as early tillering (Table 1, 5). The DWT-based classification models outperformed the PCA-based models in 67% of the 36 test cases when trained on the 20-80% data split, while both techniques performed best in 50% of the 36 test cases on the larger test data split (Table 2). PCA-based classification

utilised fewer overall features in both training sets. Accuracies and F_1 scores are consistently higher, across methods, in the second week of measurement and again in week five when compared to other timepoints. Results indicate that PCA has greater potential in differentiating between plus and minus inoculated plants when used in combination with tree-based machine learning algorithms (e.g. decision tree classifiers, extreme random forest classifiers), consistent with the literature (Nasution et al. 2018), however further work is needed to investigate this potential fully. The top performing model regardless of analysis approach was the artificial neural network, a MLP (multilayer perceptron) in these experiments. This suggests potential for further model refinement and development using deep learning algorithms.

While *F. pseudograminearum* induced CR detection was achieved using this approach, limitations have been identified that need to be addressed in future research. The models achieved accuracies on the order of 60%. This level of detection accuracy may be due to lower disease pressures at the time of detection, as infection is still in the early stages. This may result in less spectral variance between inoculated and non-inoculated plants. The models would likely improve with the input of more disease data and further work should investigate this. CR breeders indicate that detection accuracies over 30% would help advance breeding programs. The models were trained individually on each set of environmental conditions: each timepoint was treated as an individual dataset with models being optimised for that specific timepoint. Additional research needs to address whether a single developed model can be created for use across temporal and environmental situations. Generalised models will be developed and potential uses across these variable situations will be addressed in a Part B of this paper series. The outcomes of the research presented in this paper demonstrate strong capability of CR detection and warrant further investigation to develop models for disease tracing and spatial modelling in laboratory, glasshouse or field environments. The results indicate that accuracy sufficient for tracing ($> 50\%$) CR spread can be obtained early in the growing season with a time window to allow for data collection (3-weekly average). Accuracy of greater than 50% is expected to be necessary for tracing CR spread, based on expert pathologist consultation, and the results are successful in achieving the necessary accuracy. Spatial modeling using this approach would allow an increased understanding of the CR disease cycle, expanding research

opportunities and potentially increasing the effectiveness of novel CR control measures (e.g. precision spraying) (Simpfendorfer et al. 2019).

B. Temporal effects on disease detection

The results show CR detection ability, as well as a temporal patterning effect as the season progresses. Both accuracies and F_1 scores drop when trained on the 80-20% split around week 3 and week 4 of data collection (Table 5). Models trained on the 20-80% split exhibited less discrepancy in detection ability than models trained on a larger portion of the dataset. The unique patterning seen in model accuracies as the season progresses is likely indicative of changes in the underlying host-disease interaction. Further work in disease detection modelling has the potential to increase understanding of the biological relationships in the *F. pseudograminearum*-wheat disease cycle. Further research is necessary to develop an understanding of these temporal effects. Increasing input feature bandwidths may decrease the potential for environmental or site-specific features to be selected for disease detection.

C. Disease model specificity limitations

Reflectance-based crop disease approaches are currently not adequately robust to handle genotypic and environmental variability with success. Difficulties arise due to variation in spectral signatures caused by genotypic diversity and environmental conditions, particularly differences in canopy cover and structure (Jacquemoud et al. 2009). Additional issues can arise when trying to discriminate spectral signatures from those caused by other stresses, whether biotic or abiotic (Zhang et al. 2012). These problems have been discussed in (Anderegg et al. 2019) and described in detail in previous studies (Devadas et al. 2015; Zheng et al. 2019). Temporal features observed in this study (*Temporal effects on disease detection*) may be beneficial in excluding some of this genetic and environmental variability from datasets by producing signatures that are unique in both time and observable reflectance patterning. Anderegg et al. (2019) were able to show that temporal changes in spectral reflectance indicated both the presence of Septoria tritici blotch (*Zymoseptoria tritici*) in wheat, as well as disease severity, in addition to the separation of disease from physiological senescence, indicating that such an approach is a viable alternative to manual disease assessment.

The results of this study indicate that weeks 1–3, early tillering, comprise a window of increased CR detection capability (Tables 1, 3, 5; Fig. 3–6). Further, the results presented here indicate that taking a measurement for each week in this window increases overall detection capability, versus single weekly measurements. These results should inform further methodology development, particularly around high-throughput phenotyping methodology development for CR breeding programs.

D. Effects of training data

Our results indicate that when reducing the train-test split from 80-20% to 20-80% accuracies decrease, however, training time is also decreased. In these experiments, wavelet-based models performed better that PCA models when using a smaller train-test split. DWT-based models also produced a smaller overall range of accuracies than PCA-based models (Tables 1, 3). This indicates that wavelet approaches may be beneficial when working in agriculture where models cannot be trained on all possible outcomes and speed is a priority. Additionally, these results indicate that an 80-20% split may overfit the models by including too many specific signatures associated with a particular plant's disease cycle, as opposed to a robust model including only major indicators of the presence of *F. pseudograminearum*. Other training splits should be investigated to find a balance between model specificity and detection capability and minimise this potential risk.

E. Feature selection considerations

In our results the PCA-based models with top accuracies used less features, overall, than the wavelet-based models (Table 4). This may be a consideration for developing disease specific sensors, where less features could equate to lower cost and more technology adoption. This is a limitation of current imaging systems as each feature or band is associated with a particular filter on the system. Any additional feature required for detection increases the size, cost, weight and processing time of the system, which may impact potential uses, specifically when envisioning use in an aerial environment (e.g. UAV-based system).

V. CONCLUSION

In these experiments, we tested the ability of near-infrared spectroscopy, processed with PCA or DWT and machine learning methods, to detect *F. pseudograminearum* in bread wheat. Our results demonstrate potential in the proposed approach, for the

early detection of CR, additionally suggesting potential in limiting training dataset size in machine learning-based crop disease detection.

Our results indicate that wavelet-based models performed better when trained on less data (20-80%), while PCA performance increased when trained on the larger training dataset (80-20%). This suggests that wavelet-based models may be of more value when training sets are limited or when large-scale data acquisition is difficult. Further, smaller training datasets limit the amount of processing time required for detection and may reduce overfitting of models to the data. Wavelet-based models had higher minimum accuracies and less of an absolute difference between minimum and maximum accuracies than the PCA approach. While PCA demonstrated better ability to discriminate CR than wavelet-based approaches under the larger training data split, the larger differences in accuracies reported indicates that wavelet-based approaches may be more repeatable. However, results indicate that PCA is a better dimensionality reduction technique than DWT when used in conjunction with tree-based classifiers (decision tree classifiers, random forest classifiers).

Overall, our results indicate that accuracies obtained from models developed and tested on glasshouse data were the best. Early and late tillering (week 2 and 5) consistently returned the highest accuracies in CR discrimination, indicating temporally specific methodologies can increase CR detection capability. Further, it was discovered that a 3-weekly average, specifically weeks 1–3, returned higher overall accuracies than weekly (overall) models alone.

Experiments incorporating CR quantification and further model development, particularly with a focus on single model development across sites and environments should be tested. This is likely to increase model robustness as well as contribute to a deeper understanding of CR biology, specifically temporally. Finally, further work is needed to discover the degree to which these models can be utilised in image-based systems, potentially allowing for robotic or remote (e.g. UAV) approaches to CR detection and quantification.

REFERENCES

Alahmad, S, Simpfendorfer, S, Bentley, A & Hickey, L 2018, 'Crown rot of wheat in Australia: Fusarium pseudograminearum taxonomy, population biology and disease management', *Australasian Plant Pathology*, vol. 47, no. 3, pp. 285-99.

Anderegg, J, Hund, A, Karisto, P & Mikaberidze, A 2019, 'In-field detection and quantification of Septoria tritici blotch in diverse wheat germplasm using spectral-temporal features', *BioRxiv*, p. 664011.

Bergsträsser, S, Fanourakis, D, Schmittgen, S, Cendrero-Mateo, MP, Jansen, M, Scharr, H & Rascher, U 2015, 'HyperART: non-invasive quantification of leaf traits using hyperspectral absorption-reflectance-transmittance imaging', *Plant methods*, vol. 11, no. 1, p. 1.

Breiman, L, Friedman, J, Olshen, R & Stone, C 1984, 'Classification and regression trees. Wadsworth Int', Group, vol. 37, no. 15, pp. 237-51.

Bruce, LM & Li, J 2001, 'Wavelets for computationally efficient hyperspectral derivative analysis', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1540-6.

Bruce, LM, Koger, CH & Li, J 2002, 'Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331-8.

Daubechies, I 1988, 'Orthonormal bases of compactly supported wavelets', *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909-96.

Devadas, R, Lamb, D, Backhouse, D & Simpfendorfer, S 2015, 'Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat', *Precision agriculture*, vol. 16, no. 5, pp. 477-91.

Fix, E 1951, Discriminatory analysis: nonparametric discrimination, consistency properties, USAF school of Aviation Medicine.

Franke, J & Menz, G 2007, 'Multi-temporal wheat disease detection by multi-spectral remote sensing', *Precision agriculture*, vol. 8, no. 3, pp. 161-72.

Garcia-Ruiz, F, Sankaran, S, Maja, JM, Lee, WS, Rasmussen, J & Ehsani, R 2013, 'Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees', *Computers and electronics in agriculture*, vol. 91, pp. 106-15.

Geurts, P, Ernst, D & Wehenkel, L 2006, 'Extremely randomized trees', *Machine learning*, vol. 63, no. 1, pp. 3-42.

Grieve, B, Hammersley, S, Mahlein, A-K, Oerke, E-C & Goldbach, H 2015, 'Localized multispectral crop imaging sensors: Engineering & validation of a cost effective plant stress and disease sensor', in *Sensors Applications Symposium (SAS)*, 2015 IEEE, IEEE, pp. 1-6.

Hawkins, DM 2004, 'The problem of overfitting', *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 1-12.

Heim, R, Wright, I, Chang, HC, Carnegie, A, Pegg, G, Lancaster, E, Falster, D & Oldeland, J 2018, 'Detecting myrtle rust (Austropuccinia psidii) on lemon myrtle trees using spectral signatures and machine learning', *Plant pathology*, vol. 67, no. 5, pp. 1114-21.

Huang, W, Lamb, DW, Niu, Z, Zhang, Y, Liu, L & Wang, J 2007, 'Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging', *Precision agriculture*, vol. 8, no. 4-5, pp. 187-97.

Jacquemoud, S, Verhoef, W, Baret, F, Bacour, C, Zarco-Tejada, PJ, Asner, GP, François, C & Ustin, SL 2009, 'PROSPECT+ SAIL models: A review of use for vegetation characterization', *Remote Sensing of Environment*, vol. 113, pp. S56-S66.

Jones, E, Oliphant, T & Peterson, P 2001, 'SciPy: Open Source Scientific Tools for Python'.

Kazan, K & Gardiner, DM 2018, 'Fusarium crown rot caused by Fusarium pseudograminearum in cereal crops: recent progress and future prospects', *Molecular plant pathology*, vol. 19, no. 7, pp. 1547-62.

Kempeneers, P, De Backer, S, Debruyn, W, Coppin, P & Scheunders, P 2005, 'Generic wavelet-based hyperspectral classification applied to vegetation stress detection', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 610-4.

Klein, ME, Aalderink, BJ, Padoan, R, De Bruin, G & Steemers, TAG 2008, 'Quantitative Hyperspectral Reflectance Imaging', *Sensors* (Basel, Switzerland), vol. 8, no. 9, pp. 5576-618.

Klein, T, Burgess, L & Ellison, F 1991, 'The incidence and spatial patterns of wheat plants infected by Fusarium graminearum Group 1 and the effect of crown rot on yield', *Australian Journal of Agricultural Research*, vol. 42, no. 3, pp. 399-407.

Kumar, L, Schmidt, K, Dury, S & Skidmore, A 2002, 'Imaging spectrometry and vegetation science', in *Imaging spectrometry*, Springer, pp. 111-55.

Labsphere, 'Reflectance characteristics of Spectralon panels', Reflectance Calibration Laboratory, Sutton, NH, USA. (1998).

Lee, GR, Gommers, R, Waselewski, F, Wohlfahrt, K & O'Leary, A 2019, 'PyWavelets: A Python package for wavelet analysis', *Journal of Open Source Software*, vol. 4, no. 36, p. 1237.

Mahlein, A-K, Rumpf, T, Welke, P, Dehne, H-W, Plümer, L, Steiner, U & Oerke, E-C 2013, 'Development of spectral indices for detecting and identifying plant diseases', *Remote Sensing of Environment*, vol. 128, pp. 21-30.

McCulloch, WS & Pitts, W 1943, 'A logical calculus of the ideas immanent in nervous activity', *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115-33.

Nasution, M, Sitompul, O & Ramli, M 2018, 'PCA based feature reduction to improve the accuracy of decision tree c4. 5 classification', in *Journal of Physics: Conference Series*, IOP Publishing, p. 012058.

Newton, A & Hackett, C 1994, 'Subjective components of mildew assessment on spring barley', *European Journal of Plant Pathology*, vol. 100, no. 6, pp. 395-412.

Nutter, J 1997, 'Disease severity assessment training', *Exercises in plant disease epidemiology*, pp. 1-7.

Nutter Jr, F, Gleason, M, Jenco, J & Christians, N 1993, 'Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems', *Phytopathology*, vol. 83, no. 8, pp. 806-12.

Oerke, E-C, Fröhling, P & Steiner, U 2011, 'Thermographic assessment of scab disease on apple leaves', *Precision agriculture*, vol. 12, no. 5, pp. 699-715.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R & Dubourg, V 2011, 'Scikit-learn: Machine learning in Python', *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-30.

Percy, C, Wildermuth, G & Sutherland, M 2012, 'Symptom development proceeds at different rates in susceptible and partially resistant cereal seedlings infected with Fusarium pseudograminearum', *Australasian Plant Pathology*, vol. 41, no. 6, pp. 621-31.

Polder, G, van der Heijden, GW, van Doorn, J & Baltissen, TA 2014, 'Automatic detection of tulip breaking virus (TBV) in tulip fields using machine vision', *Biosystems Engineering*, vol. 117, pp. 35-42.

Python version 3.6.8, Python Software Foundation 2019.

Sankaran, S, Mishra, A, Ehsani, R & Davis, C 2010, 'A review of advanced techniques for detecting plant diseases', *Computers and electronics in agriculture*, vol. 72, no. 1, pp. 1-13.

Savitzky, A & Golay, MJ 1964, 'Smoothing and differentiation of data by simplified least squares procedures', *Analytical chemistry*, vol. 36, no. 8, pp. 1627-39.

Simpfendorfer, S, McKay, A & Ophel-Keller, K 2019, 'New approaches to crop disease management in conservation agriculture. In (Eds J Pratley and J Kirkegaard) "Australian Agriculture in 2020: From Conservation to

Automation" pp 173-188 (Agronomy Australia and Charles Sturt University: Wagga Wagga)', pp. 173-88.

Thomas, S, Wahabzada, M, Kuska, MT, Rascher, U & Mahlein, A-K 2017, 'Observation of plant–pathogen interaction by simultaneous hyperspectral imaging reflection and transmission measurements', *Functional Plant Biology*, vol. 44, no. 1, pp. 23-34.

Verhulst, P-F 1838, 'Notice sur la loi que la population suit dans son accroissement (Notice on the law of population growth)', *Corresp. Math. Phys*., vol. 10, pp. 113-26.

West, T, Prasad, S & Bruce, LM 2007, 'Multiclassifiers and decision fusion in the wavelet domain for exploitation of hyperspectral data', in *2007 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp. 4850-3.

Zadoks, JC, Chang, TT & Konzak, CF 1974, 'A decimal code for the growth stages of cereals', *Weed Research*, vol. 14, no. 6, pp. 415-21.

Zhang, J, Pu, R, Huang, W, Yuan, L, Luo, J & Wang, J 2012, 'Using in-situ hyperspectral data for detecting and discriminating yellow rust disease from nutrient stresses', *Field Crops Research*, vol. 134, pp. 165-74.

Zheng, Q, Huang, W, Cui, X, Dong, Y, Shi, Y, Ma, H & Liu, L 2019, 'Identification of Wheat Yellow Rust Using Optimal Three-Band Spectral Indices in Different Growth Stages', *Sensors*, vol. 19, no. 1, p. 35.

Вапник, В & Червонекис, А 1964, 'Об одном классе алгоритмов обучения распознаванию образов (On one class of pattern recognition learning algorithms)', *Автоматика и телемеханика*, vol. 25, no. 6, p. 937.

CHAPTER 4

MACHINE LEARNING FOR THE DETECTION OF CROWN ROT, *FUSARIUM PSEUDOGRAMINEARUM*, PART B: GENERALISED TEMPORAL MODELS FOR PCA AND DISCRETE WAVELET APPROACHES

In this study, an extension of Part A, principal component analysis (PCA) and the discrete wavelet transform (DWT) were compared for ability to increase machine learning model disease prediction accuracy across six machine learning methods (logistic regression classification, k-nearest neighbors, decision trees, extreme random forests, support vector machines, artificial neural networks), for the detection of *Fusarium pseudograminearum* induced crown rot, using models developed from combined data (generalised) across multiple trial sites and temporal windows. ANOVA F-values were used to select the optimal wavelet features. Two train-test data splits were evaluated for effectiveness in model development, 80-20% and 20-80%. Models were developed in three-week classes, weekly classes and an 'All' class consisting of combined data across the nine-week sensing window.

Machine learning for the detection of crown rot, *Fusarium pseudograminearum,* Part B: generalised temporal models for PCA and discrete wavelet approaches

Jacob Humpal[*], Cheryl McCarthy[*], Cassandra Percy[**], J. Alex Thomasson[***]

[*] Centre for Agricultural Engineering, University of Southern Queensland, Toowoomba, Queensland, Australia

[**] Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia

[***] Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, Mississippi, United States

J. Humpal, C. McCarthy (*)

Centre for Agricultural Engineering, University of Southern Queensland,

Toowoomba, Qld 4350, Australia

*e-mail: cheryl.mccarthy@usq.edu.au

Ph: +61 7 4631 2297

Abstract- In this paper, generalised models are developed from data obtained from multiple sites infected with crown rot (CR) (*Fusarium pseudograminearum*) at temporal scales of one, three and nine weeks. Models are developed on near-infrared hyperspectral data using the discrete wavelet transform (DWT) and traditional principal component analysis (PCA) and are compared for CR prediction capability. The discrete wavelet transform is applied to hyperspectral data obtained from five bread wheat (*Triticum aestivum*), resistance trials. Data features are extracted using six supervised learning techniques and are scored using 5-fold cross-validation and F-scores. The DWT results are compared to PCA evaluation at similar bandwidths. Prediction accuracy is assessed using a 20-80% and an 80-20% train-test split. The results show ability for development of models for the automated classification of disease with poor symptom visibility, using both DWT and PCA model approaches, across multiple sites, at variable temporal scales. The PCA-based system performed best in a majority of individual test cases, in terms of accuracy when training on both 20% and 80% of the disease data. PCA-based classification also utilised fewer overall features in both training sets.

I. Introduction

Wheat CR (*Fusarium pseudograminearum*) is a global concern, causing up to a 50% yield loss, or greater, under conducive conditions, particularly in arid and semi-arid regions (Klein et al. 1991; Kazan & Gardiner 2018). When *F. pseudograminearum* hyphae fragments contained in crop residues and soil receive sufficient moisture, the fungus is able to infect a susceptible host in close contact to the pathogen. The fungus begins by colonising the area immediately around the point of contact and proceeds upwards and inwards (Alahmad et al. 2018). With a lack of cost-effective fungicide or other in-season control options, severe infection can lead to necrosis of the stem, limiting production capability. The visible symptoms of CR include browning of the stem under the leaf sheaths and the appearance of white heads late in the season. These symptoms are often difficult to detect quickly and limit the ability of breeding programs to rapidly produce germplasm with improved resistance and tolerance to this disease, as rating is time consuming, often involving the removal of plant material from the field. A rapid, high-throughput phenotyping (HTP) system for CR resistance and tolerance would save breeders time and inputs, allowing growers access to new improved commercial cultivars faster.

A wavelet analysis approach could provide novel application for CR detection. The ability of the approach to return features of variable scales simplifies the development of multispectral sensors for infield CR detection and quantification, by returning narrow and wide-scale features or bands in the disease detection model. An accurate sensing system with minimal features keeps the cost of the system down while also increasing model robustness. This paper is Part B of a series of three papers that describe the development of crop disease models for the detection and quantification of CR in wheat. Part A demonstrated variation in the accuracy and F_1 scores of PCA and wavelet approaches for preprocessing of spectral data collected across 9 weeks from five experiments (Chapter 3). This paper, Part B, further explores improvements in accuracy and precision and recall by comparing the performance of models developed across different combinations of data with the aim to generalise CR phenotyping models, making developed models more robust across environments. Part

C expands upon this model development and these findings to investigate the potential impact of spatial and temporal information on CR detection.

The literature has consistently shown that using spectral reflectance to discriminate or phenotype disease to facilitate crop breeding programs is difficult due to the large amount of variation that arises from genotypic and environmental variation (Jacquemoud & Baret 1990; Jacquemoud & Ustin 2001; Delalieux et al. 2007; Jacquemoud et al. 2009; Anderegg et al. 2019; Zheng et al. 2019). This variation can occur due to differences in canopy reflectance and morphology brought on by genetic or environmental factors and can be further influenced by temporal considerations (e.g. growth stage) as described in detail in previous studies (Devadas et al. 2015; Alahmad et al. 2018; Zheng et al. 2019) and observed in Part A of this research (Chapter 3). The aim of this study is to determine if and to what degree models built upon combined near-infrared reflectance data from germplasm with diverse genetic backgrounds across multiple environments, both glasshouse and field, can discriminate CR without being specifically tailored to a single dataset. This paper will further evaluate the effect of different train-test sizes to determine whether there is a significant reduction in the accuracy of developed models. A reduction in the size of the train-test split could reduce computation costs.

II. Materials and Methods

A. Trial site, plant cultivation and experimental design

This study delivers part B of a series of objectives built around analysis of data from five CR trials. The methodologies are outlined in Part A (Chapter 3) of this series and are described in brief in the following section. Three glasshouse trials were conducted between 2018 and 2019 in Toowoomba, Queensland in collaboration with the University of Southern Queensland's Centre for Crop Health. In the first glasshouse trial, four standard genotypes of bread wheat were observed under null and positive *F. pseudograminearum* colonised grain inoculation. Twelve replicates, for 96 total plants were observed. In the second glasshouse trial an additional standard was added to the initial four, with six replicates of each treatment, a total of 60 plants observed. The third glasshouse trial was a replication of the second. All glasshouse trials were maintained between 20–25° C and watered to field capacity, weekly. Plants were

68

removed from the glasshouse and assessed on disease severity at physiological maturity.

Two field trials were conducted in 2018 and 2019 at the Tosari Research Station (-27.859964, 151.452766). These trials consisted of three replicates of paired, inoculated and non-inoculated plots arranged in a randomised strip plot design. Six plants were randomly selected from each of the plots of the five genotypes for NIR scanning each week. All plants from glasshouse and field trials were removed from the soil at maturity for manual disease scoring at the Centre for Crop Health. While natural infection with crown rot or other pathogens is possible, assessors reported no natural infection of material. All glasshouse and field trials were statistically designed in collaboration with biostatisticians at Queensland Department of Agriculture and Fisheries (QDAF).

B. Near infrared measurements

Near infrared measurements were collected using the handheld, DLP® NIRscan™ Nano (Texas Instruments, USA), for nine weeks, from when the Zadoks (Z) 13, 21 stage was identified. The Nano has an Indium Gallium Arsenide (InGaAs) detector with a shape of 854×480, pixel pitch of 5.4 µm and a 10 nm spectral resolution, with two tungsten lamps as an integrated light source. The signal to noise ratio of the Nano is 6000:1. Measurements were collected from the centre of the youngest tiller leaf, the youngest flag leaf and the leaf considered centre-most in the canopy profile, for both inoculated and non-inoculated plants. Measurements were saved as comma separated value (CSV) files for use in data analysis. All measurements were calibrated using 10% grey, 60% grey and 99% white reference Spectralon® panels. Measurements were corrected as described below:

$$Corrected\ Reflectance = Calibration\ Coeffient \left(\frac{Target\ Radiance}{Reference}\right)$$

(Labsphere 1998)

C. Data grouping

Data was grouped into a combined dataset, comprising all measurements across all trials, denoted as 'All', a weekly dataset ('Weekly'), comprising all measurements for a particular week across trials, and a '3-weekly group' dataset, comprising models built upon combined data from the first three, middle three or last three weeks, in the

nine week window, across trials. The '3-weekly group' dataset was used to examine the impact of larger time-windows on the detection capability of models developed on combined data from across trials and dates. Data groupings were replicated for both the 20-80% and 80-20% data splits.

D. Data analysis

All model analysis was performed using the Python computing environment (Python version 3.6.8, Python Software Foundation 2019). Data analysis primarily used the SciPy ecosystem (Jones et al. 2001) and the Scikit-learn library (Pedregosa et al. 2011). Additional libraries and packages have been cited where relevant throughout.

D.1 PCA approach

The PCA analysis was performed following the methodology outlined in Part A of this paper series (Chapter 3). To approximate the bandwidths produced in the wavelet approach, raw spectral data of adjacent wavebands were averaged. For implementation in machine learning algorithms the resulting averaged bands were standardised and scaled 0–1 using the Python SciPy library, version 0.19.1 (Jones et al. 2001). Spectra were then smoothed using the Savitzky-Golay filter (Savitzky & Golay 1964) with a window size of 5 bands and a second order polynomial, using the savgol_filter function in the Python library SciPy (Jones et al. 2001). Principal component analysis (PCA) was performed on the smoothed signals and constrained to a maximum of 10 principal components using the decomposition package in the Python SciPy library (Jones et al. 2001). These principal components were used to determine the most important signal features for disease detection within the input data set. The PCA transformed dataset was used with the machine learning algorithms.

D.2 Wavelet approach

The wavelet approach uses a scaled operator to return data features at both local and global resolutions, as described below (1):

$$\omega_s f(\lambda) = f(\lambda) * \psi_s(\lambda) = f(\lambda) * \frac{1}{s} \psi\left(\frac{\lambda}{s}\right) \tag{1}$$

(Bruce & Li 2001)

where $f(\lambda)$ is the near-infrared signal, s represents the scale and $\psi_s(\lambda)$ the scaled wavelet function. The discrete wavelet transform (DWT) is advantageous over traditional methods which often require an operator derived from the smoothing operation (Bruce et al. 2002). Near-infrared spectral signatures were examined for outliers (z score > 3) and transformed using the DWT with a Daubechies 2 wavelet from the PyWavelets library, version 1.0.3 (Lee et al. 2019). The Daubechies 2 wavelet is an asymmetric and biorthogonal wavelet from the Daubechies family of wavelets (Part A, Fig. 2) (Daubechies 1988).

Resulting spectral features were transformed to bandwidths of 7.02, 14.04, 27.59, 53.33, 100.00, 200.00, 400.00 and 800.00 nm. Resulting spectra were standardised and scaled to 0–1 in the Python SciPy library version 0.19.1 (Jones et al. 2001).

D.3 Feature selection

Wavebands with bandwidths less than 25 nm were discarded before feature selection to increase model robustness across environments, in this study both glasshouse and field. Signal features represent the wavelengths in the hyperspectral signature which were most highly correlated to the differences in disease presence. Features for the PCA approach were selected based on the features of most importance in determining the principal components. Wavelet feature selection was determined using the SelectKBest function from the feature_selection package of the Python Scikit-learn library, using the selection classifier f_classif function which computes the ANOVA F-value (Pedregosa et al. 2011). The top 1–10 features were selected based upon their determined k highest ANOVA values.

D.4 Machine learning techniques

Six machine learning techniques were evaluated for effectiveness in discriminating CR, *Fusarium pseudograminearum*: logistic regression, k-nearest neighbors, decision trees, extreme random forests, support vector machines and artificial neural networks. Logistic regression is a statistical model that uses a logistic function to model a binary dependent variable. K-nearest neighbors assigns classification based on a majority vote of the nearest neighbors compared to stored instances of training data. Decision trees create a model that predicts the value of a target variable based on several input variables. In these models each branch represents combinations of features that lead to an outcome or class label, the leaves of the tree. A support vector machine constructs

a hyper-plane or set of hyper-planes in a high or infinite dimensional space, where separation is achieved by the hyper-plane that has the largest distance to the nearest training data points. Artificial neural networks (ANN) are classification systems that analyse training data without being programmed with specific rules and can automatically discover unique characteristics from the input training data set.

D.5 Model evaluation and validation

Nine weeks for each of three glasshouse and two field trials were evaluated using the PCA and wavelet methods. Each of the resulting datasets were split into training and test sets of 20-80% and 80-20% respective splits. The input feature sets were trained for each combination of hyperparameters determined using the GridSearchCV function from Scikit-learn's model_selection package (Pedregosa et al. 2011). The GridSearchCV function performs an exhaustive search over specified parameter values for an estimator, optimised by 5-fold cross-validation. Each validation split used for scoring was a random independent validation split. These splits were obtained using the train_test_split function in Scikit-learn (Pedregosa et al. 2011). The resulting models were scored on accuracy and F_1 (F-score). Additionally, the number of features included in the top performing models was considered. The equations used for accuracy, precision, recall and F_1 are reported below (2,3,4,5):

$$accuracy\left(y, \hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \cdot 1\left(\hat{y}_i = y_i\right) \qquad (2)$$

Where $\hat{y}_i$ is the predicted value of the ith sample and y_i is the corresponding true value

$$Precision = \frac{tp}{tp + fp} \qquad (3)$$

Where tp is number of true positive values and fp is number of false positive values

$$Recall = \frac{tp}{tp + fn} \qquad (4)$$

Where tp is number of true positive values and fn is number of false negative values

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \qquad (5)$$

III.   RESULTS

A. Binary classification: accuracy

PCA-based models returned higher maximum accuracies across all data groupings and both data splits with accuracies ranging from 56.02–69.78% versus 53.71–58.43% with DWT. However, wavelet-based models had a narrower overall range in accuracies across all test cases of 9.22% versus 21.69% in the PCA-based models (Table 1, Fig. 1, 2).

When models were developed on data from all trial sites across the nine weeks, in both the 20% and 80% training datasets, PCA outperformed wavelet analysis in logistic regression classification, decision tree classification and extreme random forest classification. Wavelet analysis performed better in artificial neural network-based classification and support vector classification in both training datasets. Wavelet analysis performed better than PCA in k-nearest neighbor classification when trained on 20% of the dataset with PCA performing better when trained on 80% of the dataset (Fig. 1, 2, Table 2).

When examining models developed in three-week groupings, PCA analysis outperformed wavelet analysis in the 20% training dataset in support vector classification, logistic regression classification, decision tree classification and extreme random forest classification with wavelet analysis performing best in artificial neural network classification and k-nearest neighbor classification. In the 80% training dataset PCA outperformed wavelet analysis in artificial neural network classification, decision tree classification, extreme random forest classification and k-nearest neighbor classification, while wavelet analysis performed best in support vector and logistic regression classification (Fig. 1, 2, Table 2).

Models developed from all trial sites for individual weeks showed that in both the 20% and 80% training datasets, PCA outperformed wavelet analysis in artificial neural network classification, decision tree classification and extreme random forest classification. Wavelets performed best in both training datasets in support vector classification, logistic regression classification and k-nearest neighbor classification (Fig. 1, 2, Table 2). Top average accuracies of $69.78 \pm 0.72\%$ were recorded using the PCA approach on data from three-weekly groups trained on an 80-20% train-test split (Table 1).

Table 1: Minimum and maximum ranges of average classification accuracies of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% or 80-20% train-test split. Models were developed on data combined across all 5 experiments. Data is reported as 'All' data (weeks 1–9); 3-week groupings (weeks 1-3, 4-6, 7-9); and 'Weekly' (each individual week) from Z 13, 21.

Data split	Temporal grouping	DWT		PCA	
		Minimum	Maximum	Minimum	Maximum
20-80	All data (n=4476)	49.88	54.43	49.78	56.02
	3-weekly groups (n=1452–1512±SE)	49.21 ± 0.66	53.71 ± 1.11	49.15 ± 0.19	56.11 ± 1.27
	Weekly (n=444–504±SE)	50.01 ± 0.37	54.34 ± 0.95	49.53 ± 0.09	57.77 ± 0.76
80-20	All data (n=4476)	51.00	56.28	48.09	67.72
	3-weekly groups (n=1452–1512±SE)	50.51 ± 0.85	55.42 ± 1.32	48.76 ± 0.69	69.78 ± 0.72
	Weekly (n=444–504±SE)	51.34 ± 1.24	58.43 ± 1.26	49.91 ± 1.02	64.66 ± 1.30

Fig. 1: Average accuracies of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on 20% of input datasets. Models were developed on combined data from all trial sites from all weeks measured combined (weeks 1–9), from 3-week groupings and from combined data for each week from Zadoks (Z) 13, 21.

Fig. 2: Average accuracies of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on 80% of input datasets. Models were developed on combined data from all trial sites from all weeks measured combined (weeks 1–9), from 3-week groupings and from combined data for each week from Zadoks (Z) 13, 21.

Table 2: Percentage of instances in which each system (DWT or PCA) outperformed the others in CR detection, based upon accuracy percentage and trained on 20% or 80% of input data. Models were developed on data combined across all 5 experiments. Data is reported as 'All' data (weeks 1–9); 3-week groupings (weeks 1-3, 4-6, 7-9); and 'Weekly' (each individual week) from Z 13, 21.

Data split	Temporal grouping	MLP		SVC		Logistic Regression		Decision Tree		Extreme Random Forest		K-Nearest	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
20-80	All data (n=4476)	100.00	0.00	100.00	0.00	0.00	100.00	0.00	100.00	0.00	100.00	100.00	0.00
	3-weekly groups (n=1452–1512±SE)	66.67	33.33	33.33	66.67	33.33	66.67	0.00	100.00	0.00	100.00	100.00	0.00
	Weekly (n=444–504±SE)	44.44	55.56	66.67	33.33	55.56	44.44	0.00	100.00	22.22	77.78	100.00	0.00
80-20	All data (n=4476)	100.00	0.00	100.00	0.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00
	3-weekly groups (n=1452–1512±SE)	0.00	100.00	100.00	0.00	100.00	0.00	0.00	100.00	0.00	100.00	0.00	100.00
	Weekly (n=444–504±SE)	44.44	55.56	55.56	44.44	55.56	44.44	22.22	77.78	0.00	100.00	66.67	33.33

B. Binary classification: precision and recall

Wavelet models returned higher maximum F_1 scores than PCA analysis when trained on 3-weekly (0.5843 ± 0.0348) groups under the 20-80% data split and a higher minimums (0.5464) under models developed on all available data ('All') in the 80-20% split. PCA-based models returned higher maximum F_1 scores across all scenarios under the 80-20% data split and higher minimum scores under the 'All' category in the 20-80% data split (0.5199). PCA-based models returned scores with a smaller overall range than wavelet-based models when trained under the 80-20% data split (Fig. 3, 4, Table 3). Overall, the highest score was achieved ($F_1 = 0.6953$) with PCA analysis when trained on 3 weekly groups under an 80-20% data split.

Table 3: Minimum and maximum range of average F_1 scores of CR detection of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% or 80-20% train-test split. Models were developed on data combined across all 5 experiments. Data is reported as 'All' data (weeks 1–9); 3-week groupings (weeks 1-3, 4-6, 7-9); and 'Weekly' (each individual week) from Z 13, 21.

		DWT		PCA	
Data split	Temporal grouping	Minimum	Maximum	Minimum	Maximum
	All data (n=4476)	0.4854	0.6656	0.5199	0.6645
20-80	3-weekly groups (n=1452–1512±SE)	0.4090 ± 0.2068	0.5843 ± 0.0348	0.2533 ± 0.1267	0.5470 ± 0.0034
	Weekly (n=444–504±SE)	0.2932 ± 0.1145	0.5550 ± 0.0246	0.2355 ± 0.0983	0.5608 ± 0.0115
	All data (n=4476)	0.5464	0.6570	0.4618	0.6727
80-20	3-weekly groups (n=1452–1512±SE)	0.3500 ± 0.1701	0.5639 ± 0.0235	0.4641 ± 0.0351	0.6953 ± 0.0123
	Weekly (n=444–504±SE)	0.5028 ± 0.0797	0.6162 ± 0.0193	0.5004 ± 0.0281	0.6437 ± 0.0127

Fig. 3: Average F_1* scores of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on 20% of input datasets. Models were developed on combined data from all trial sites from all weeks measured combined (weeks 1–9), from 3-week groupings and from combined data for each week from Zadoks (Z) 13, 21.

*F_1 - the harmonic mean of precision and recall

Fig. 4: Average F_1* scores of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on 80% of input datasets. Models were developed on combined data from all trial sites from all weeks measured combined (weeks 1–9), from 3-week groupings and from combined data for each week from Zadoks (Z) 13, 21.

*F_1 - the harmonic mean of precision and recall

C. Binary classification: feature number

The number of features in the models developed with data from the nine weeks combined ('All') with the highest classification accuracies were evaluated. In the 20% training, 80% testing dataset, PCA analysis returned fewer features in models than wavelet analysis in 67% of test cases. PCA outperformed wavelet analysis in regard to number of features using artificial neural network classifiers, decision tree classifiers, extreme random forest classifiers and k-nearest neighbor classifiers. In the 80% training, 20% testing data set, PCA analysis returned fewer features than wavelet analysis in 83% of test cases. PCA outperformed wavelet analysis in all modeling techniques tested with the exception of the extreme random forest classifier, where the preprocessing techniques performed equally (Table 4).

In the 20% training, 80% testing dataset, PCA analysis returned fewer features in models than wavelet analysis in 83% of test cases when information of each of the three-weekly groups with the highest classification accuracies were considered. PCA outperformed wavelet analysis in regard to number of features in all machine learning approaches tested with the exception of extreme random forest classification in the three-week grouping. In the 80% training, 20% testing data set, PCA analysis returned fewer features than wavelet analysis in 50% of test cases. PCA outperformed wavelet analysis in all modeling techniques except artificial neural network classification and decision tree classification, and logistic regression classification where both preprocessing techniques returned fewer features in an equal number of test cases (Table 4).

In the 20% training, 80% testing dataset, PCA analysis returned fewer features in models than wavelet analysis in 83% of test cases when information for each of the individual weeks with the highest classification accuracies were evaluated. PCA outperformed wavelet analysis in regard to number of features in all machine learning approaches tested except decision tree classification, where both models returned the same number of features in individual weeks. In the 80% training, 20% testing data set, PCA analysis returned fewer features than wavelet analysis in 83% of test cases, again. PCA outperformed wavelet analysis in all modeling techniques except extreme random forest classification, where both approaches returned the same number of features (Table 4).

Table 4: Number of features chosen, out of an available 10, optimised for accuracy of CR classification detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on 20 or 80% of input datasets. Models were developed on data combined across all 5 experiments. Data is reported as 'All' data (weeks 1–9); 3-week groupings (weeks 1-3, 4-6, 7-9); and 'Weekly' (each individual week) from Z 13, 21.

Data split	Temporal grouping	MLP		SVC		Logistic Regression		Decision Tree		Extreme Random Forest		K-Nearest	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
20-80	All data (n=4476)	8	6	1	10	1	3	10	9	10	5	10	6
	3-weekly groups (n=1452–1512±SE)	6 ± 2	4 ± 2	4 ± 3	1 ± 0	4 ± 3	1 ± 0	9 ± 0	5 ± 3	6 ± 1	7 ± 2	6 ± 2	1 ± 0
	Weekly (n=444–504±SE)	9 ± 1	6 ± 1	3 ± 1	1 ± 0	5 ± 1	1 ± 0	7 ± 1	7 ± 1	7 ± 1	6 ± 1	6 ± 1	1 ± 0
80-20	All data (n=4476)	7	2	2	1	10	1	7	2	8	8	5	1
	3-weekly groups (n=1452–1512±SE)	4 ± 2	6 ± 2	4 ±3	1 ± 0	1 ± 0	1 ± 0	5 ± 1	6 ± 1	7 ± 2	5 ± 1	3 ± 0	1 ± 0
	Weekly (n=444–504±SE)	8 ± 1	6 ± 1	3 ± 1	1 ± 0	5 ± 1	1 ± 0	6 ± 1	5 ± 1	5 ± 1	5 ± 1	5 ± 1	1 ± 0

D. Temporal effects

Analysis of variance (ANOVA) detected no significant temporal effects between datasets that were grouped into the 3-weekly categories in either the 20-80% or 80-20% data splits. However, significant temporal effects were observed for both data splits across weekly models, $p = 0.028199$ for the 20-80% split and $p = 0.000268$ for the 80-20% split ($\alpha = 0.05$) (Table 5).

Table 5: ANOVA analysis of variation between weeks for accuracy of CR classification detection (+ or -) of optimised machine learning models, trained on 20 or 80% of input datasets. Models were developed on combined data from all experiments (weeks 1–9). 'Groups' are denoted as from identification of Zadoks (Z) 13, 21.

Data split	Growth Stage	Groups	Count	Sum	Average	Variance
		1	14	7.6048	0.5432	0.00068
		2	14	7.3838	0.527414	0.001126
	Tillering	3	14	7.5006	0.535757	0.000824
		4	14	7.3882	0.527729	0.000768
20-80		5	14	7.806	0.557571	0.002078
		6	14	7.5025	0.535893	0.000871
	Stem Extension	7	14	7.1963	0.514021	0.000385
		8	14	7.3676	0.526257	0.00054
	Boot	9	14	7.5683	0.540593	0.001336
α		0.05				
P-value		0.028199				
		1	14	9.088	0.649143	0.006906
		2	14	7.8591	0.561364	0.004961
	Tillering	3	14	7.0426	0.503043	0.015527
		4	14	6.988	0.499143	0.028833
80-20		5	14	8.5935	0.613821	0.004414
		6	14	8.5512	0.6108	0.001821
	Stem Extension	7	14	8.2763	0.591164	0.005718
		8	14	8.1162	0.579729	0.003282
	Boot	9	14	8.0282	0.573443	0.004801
α		0.05				
P-value		0.000268				

IV. DISCUSSION

A. Crown rot detection

Currently CR detection and assessment methods rely on trained assessors and involve destructive sampling. These methods are time consuming and expensive, particularly to process a large number of samples to characterise infection levels in an entire field or paddock, or for germplasm assessment by researchers and breeding companies. Rapid, non-invasive, high-throughput CR detection is required to reduce costs and improve germplasm delivery timeframes. Near-infrared spectroscopy has the potential

to detect and quantify crop disease before visible symptoms appear, becoming particularly useful in situations where crops exhibit few symptoms, or where symptoms are difficult to perceive, for example, soil and stubble-borne diseases (Kumar et al. 2002). Results from Part A show potential for automated classification of CR using both DWT and PCA approaches. In Part A the DWT-based classification system outperformed the PCA-based system when training on a 20-80% split, while both systems performed best in 50% of test cases on the 80-20% split. In both training splits the PCA-based approach utilised fewer features in top performing models.

The results of our experiments, presented in this paper, illustrate potential for the use of optimised machine learning and near-infrared sensing technologies to detect CR across trial sites/paddocks (Tables 1–3). Results indicate that PCA has more potential in this regard when used in combination with tree-based machine learning algorithms (e.g. decision tree classifiers, extreme random forest classifiers), consistent with the results presented in Part A of this paper series (Chapter 3) and in the literature (Nasution et al. 2018). Wavelet-based models show particular potential when trained on smaller portions of datasets, indicating value in robust model development.

The models were trained on weekly, tri-weekly and all-inclusive groups across 5 experiments. The use of these generalised models, allows for models to be developed specifically for CR detection across environments, which can be evolved into a multispectral sensing system for direct deployment to growers and breeders, without the need to train new models at each site. The fewer features involved in the top performing tri-weekly models, relative to other data groupings, would allow for less expensive multispectral sensing solutions to be developed, as each additional feature corresponds to a new filter or sensor/camera. Further, tri-weekly models have the advantage of flexibility with sensing times, and therefore disease detection may be less constrained to a particular timepoint and will fit in readily with other breeding program or grower operations.

Accuracy and F_1 scores were reduced compared to training on individual data sets (Part A) when generalised models were developed. This drop is consistent across all data splits and groupings. Consistent with Part A, PCA-based approaches outperformed DWT-based approaches in number of features across both data splits. However, when models were generalised PCA-based approaches outperformed DWT-based approaches in both the 20-80% and 80-20% train-test splits. Further refinement of data

groupings needs be investigated to determine whether other temporal combinations are beneficial in model development. Additional analysis of temporal effects along with spatial effects will be evaluated in Part C of this study.

B. Temporal effects

Near-infrared hyperspectral and reflectance-based crop disease approaches are not currently sufficiently robust to handle site variability with success. This has been discussed in Anderegg et al. (2019), described in detail in previous studies (Devadas et al. 2015; Zheng et al. 2019) and noted in Part A of this research. Results indicate that temporal features must be investigated further to obtain the highest accuracy across environments and cropping sites. Results show that there are significant differences in accuracies temporally across weekly groups, but not across tri-weekly groups.

C. Effects of training splits

Reducing the train-test split from 80-20% to 20-80% results in less of a reduction in accuracy and F_1 scores, in DWT-based models than in PCA-based models (Tables 1, 3). PCA-based models generally perform with greater accuracy on higher percentages of training data. PCA also returned higher accuracies overall for models generated from multi-site data. Additionally, PCA-based models performed better with tree-based classifiers (decision tree, extreme random forests), indicating the potential for use of PCA with tree-based classification, consistent with recent literature (Nasution et al. 2018). Further, as using a higher train-test split may overfit the models by training on features specific to a particular dataset and not necessarily to a particular disease (i.e. CR), other training splits should be investigated as put forward in Part A. Finally, a smaller train-test split would decrease processing time.

D. Feature selection

In our results the PCA-based models generally utilised fewer overall features in best performing models as opposed to DWT-based models (Table 4). This is consistent to findings in Part A and should be a consideration for developing disease specific multispectral sensors with less features, where lower cost sensor development is required.

V. CONCLUSION

In Part A of this study, we tested the ability of near-infrared spectroscopy in conjunction with various machine learning methods and preprocessing techniques to detect CR (*Fusarium pseudograminearum*) in wheat. Results from Part A were further explored in this paper to determine the capacity to which these models could be generalised, both temporally and across environments. The successful generalisation of accurate CR models would allow for the development of a sensing system usable across breeding programs and trial sites while minimising required calibration of models. This generalised system would decrease sensor adoption costs while increasing system viability. Results show potential in combined dataset model development around CR detection.

Generalised models developed using PCA-based approaches outperformed DWT approaches overall in the tri-weekly grouping in the 20-80% train-test split, both in accuracy and in F_1 scores, while both processing approaches performed best across half of the test cases in the 'All' and 'Weekly' groups. PCA-based, generalised models outperformed DWT approaches across both the 'All' and tri-weekly groupings in the 80-20% train-test split, with both approaches performing best on 50% of cases in the 'Weekly' grouping. Further, PCA-based approaches consistently provided models with fewer features than wavelet-based approaches and performed better in conjunction with tree-based models, also consistent with Part A. Fewer features are an important consideration in the development of new sensing solutions as the cost of a sensor increases with each additional input band. The highest accuracies and F_1 scores for generalised CR models were seen using a PCA-based approach on tri-weekly groupings with an 80-20% train-test split. These results were obtained using an extreme random forest classifier with an average accuracy of 69.78 ± 0.72% and a F_1 of 0.6953 ± 0.0123. Significant temporal patterning effects were seen between individual weekly model accuracies, which will be investigated further in Part C of this paper series.

Further experiments, particularly around CR quantification and model development should and will be examined. These examinations will increase knowledge and understanding of CR pathology and model specificity for phenotyping systems. Further work should be undertaken to examine image-based system modelling, potentially increasing model deployment opportunities.

REFERENCES

Alahmad, S, Simpfendorfer, S, Bentley, A & Hickey, L 2018, 'Crown rot of wheat in Australia: Fusarium pseudograminearum taxonomy, population biology and disease management', *Australasian Plant Pathology*, vol. 47, no. 3, pp. 285-99.

Anderegg, J, Hund, A, Karisto, P & Mikaberidze, A 2019, 'In-field detection and quantification of Septoria tritici blotch in diverse wheat germplasm using spectral-temporal features', *BioRxiv*, p. 664011.

Bruce, LM & Li, J 2001, 'Wavelets for computationally efficient hyperspectral derivative analysis', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1540-6.

Bruce, LM, Koger, CH & Li, J 2002, 'Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331-8.

Daubechies, I 1988, 'Orthonormal bases of compactly supported wavelets', *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909-96.

Delalieux, S, Van Aardt, J, Keulemans, W, Schrevens, E & Coppin, P 2007, 'Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications', *European Journal of Agronomy*, vol. 27, no. 1, pp. 130-43.

Devadas, R, Lamb, D, Backhouse, D & Simpfendorfer, S 2015, 'Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat', *Precision agriculture*, vol. 16, no. 5, pp. 477-91.

Jacquemoud, S & Baret, F 1990, 'PROSPECT: A model of leaf optical properties spectra', *Remote Sensing of Environment*, vol. 34, no. 2, pp. 75-91.

Jacquemoud, S & Ustin, SL 2001, 'Leaf optical properties: A state of the art', in *8th International Symposium of Physical Measurements & Signatures in Remote Sensing*, pp. 223-332.

Jacquemoud, S, Verhoef, W, Baret, F, Bacour, C, Zarco-Tejada, PJ, Asner, GP, François, C & Ustin, SL 2009, 'PROSPECT+ SAIL models: A review of use

for vegetation characterization', *Remote Sensing of Environment*, vol. 113, pp. S56-S66.

Jones, E, Oliphant, T & Peterson, P 2001, 'SciPy: Open Source Scientific Tools for Python'.

Kazan, K & Gardiner, DM 2018, 'Fusarium crown rot caused by Fusarium pseudograminearum in cereal crops: recent progress and future prospects', *Molecular plant pathology*, vol. 19, no. 7, pp. 1547-62.

Klein, T, Burgess, L & Ellison, F 1991, 'The incidence and spatial patterns of wheat plants infected by Fusarium graminearum Group 1 and the effect of crown rot on yield', *Australian Journal of Agricultural Research*, vol. 42, no. 3, pp. 399-407.

Kumar, L, Schmidt, K, Dury, S & Skidmore, A 2002, 'Imaging spectrometry and vegetation science', in *Imaging spectrometry*, Springer, pp. 111-55.

Labsphere, 'Reflectance characteristics of Spectralon panels', Reflectance Calibration Laboratory, Sutton, NH, USA. (1998).

Lee, GR, Gommers, R, Waselewski, F, Wohlfahrt, K & O'Leary, A 2019, 'PyWavelets: A Python package for wavelet analysis', *Journal of Open Source Software*, vol. 4, no. 36, p. 1237.

Nasution, M, Sitompul, O & Ramli, M 2018, 'PCA based feature reduction to improve the accuracy of decision tree c4. 5 classification', in *Journal of Physics: Conference Series*, IOP Publishing, p. 012058.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R & Dubourg, V 2011, 'Scikit-learn: Machine learning in Python', *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-30.

Python version 3.6.8, Python Software Foundation 2019.

Savitzky, A & Golay, MJ 1964, 'Smoothing and differentiation of data by simplified least squares procedures', *Analytical chemistry*, vol. 36, no. 8, pp. 1627-39.

Zheng, Q, Huang, W, Cui, X, Dong, Y, Shi, Y, Ma, H & Liu, L 2019, 'Identification of Wheat Yellow Rust Using Optimal Three-Band Spectral Indices in Different Growth Stages', *Sensors*, vol. 19, no. 1, p. 35.

CHAPTER 5

MACHINE LEARNING FOR THE DETECTION OF CROWN ROT, *FUSARIUM PSEUDOGRAMINEARUM*, PART C: SENSING LOCATION AND TIMING

In this study, an extension of Part A and B, significant differences in accuracy between sensing dates (temporal) and sensing location (spatial) are evaluated in the scope of detection of *Fusarium pseudograminearum* induced crown rot, using the discrete wavelet transform (DWT) and six machine learning methods, developed from combined data across multiple trial sites. ANOVA F-values were used to select the optimal wavelet features, with data trained on an 80-20% train-test split. Significant temporal and spatial differences were observed using ANOVA and the Tukey HSD, post-hoc test.

Machine learning for the detection of crown rot, *Fusarium pseudograminearum*, Part C: sensing location and timing

Jacob Humpal[*], Cheryl McCarthy[*], Cassandra Percy[**], J. Alex Thomasson[***]

[*] Centre for Agricultural Engineering, University of Southern Queensland, Toowoomba, Queensland, Australia

[**] Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia

[***] Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, Mississippi, United States

J. Humpal, C. McCarthy (*)

Centre for Agricultural Engineering, University of Southern Queensland,

Toowoomba, Qld 4350, Australia

*e-mail: cheryl.mccarthy@usq.edu.au

Ph: +61 7 4631 2297

Abstract- In this paper, temporal and spatial sensing considerations are evaluated in the context of *Fusarium pseudograminearum* induced crown rot (CR) detection, using a commercially available near-infrared contact sensor. Machine learning models were developed from features derived from a discrete wavelet transform (DWT) for the detection of CR in five genotypes of known susceptibility (standards). Models were evaluated across nine weeks from three weeks after identification of Zadoks (Z) 13, 21 in three glasshouse and two field trials, with signatures recorded from the youngest leaf of the newest emerged tiller, centre-most leaf, flag leaf of the newest emerged tiller and grain head, when present. Significant differences in time of detection (temporal), from Z 13, 21 and sensing location (spatial) are reported.

Keywords*- crop disease, temporal, machine-learning, sensing, wavelet*

I. Introduction

Wheat CR is a significant cereal disease in Australia caused by one of three *Fusarium* species, *Fusarium culmorum*, *Fusarium graminearum* or *Fusarium pseudograminearum* (Parry et al. 1994). With the increase in conservation management agriculture in recent years, CR prevalence has grown, as infectious material is retained in the field between seasons (Simpfendorfer et al. 2019). Potential annual yield loss from CR has been estimated to have increased from 1998 to 2008 from 3.0 to 22.2%, across the Northern, 0.4 to 10.5%, across the Southern and 0.1 to 1.5% across the Western regions of Australia. However, reports in more recent years have indicated that the distribution of CR has increased again, particularly across the Western region, with incidence of paddocks with high levels of infected stubble across the grain belt during 2014–2017 at 31% in the Northern region, 21% in the Western region and 15% in the Southern region (Simpfendorfer et al. 2019).

Traditional CR assessment relies on disease correlation to noticeable stem browning and whiteheads, and therefore incidence and severity of disease is often determined well after infection. This limitation of traditional CR assessment makes it a suitable target for machine vision applications. However, spatial and temporal considerations need to be measured when developing such systems to provide a methodology for consistent, accurate disease assessment and to begin to better understand the underlying biological mechanisms that determine host tolerance and resistance.

Several recent publications have indicated that temporal patterning is important in disease detection and quantification, in addition to being advantageous in advanced and high-throughput (HTP) phenotyping applications (Franke & Menz 2007; Busemeyer et al. 2013; Bauriegel & Herppich 2014; Anderegg et al. 2019). However, there is no set best time to detect crop disease. While Bauriegel and Herppich (2014) reported that 10–14 days after infection was ideal for detecting downey mildew in lettuce cultivars, Franke and Menz (2007) reported detection ability of powdery mildew and leaf rust in winter wheat increased as the season progressed. These differences are likely due to differences in disease cycles, host response and specific individual sensor limitations. A novel, promising approach is the use of spectral-temporal features as independent inputs into disease resistance modeling applications, as changes over time are often unique and can be used to develop more specific disease signatures than those based on a single timepoint. This approach has been successfully

validated on 330 wheat genotypes under *Septoria tritici* blotch infection (Anderegg et al. 2019). Temporal effects have been considered in this study to determine if individual or groups of timepoints, from one to nine weeks after wheat inoculation with *F. pseudograminearum* (Z 13, 21), are significantly more effective at determining CR infection.

While the importance of temporal considerations when developing crop disease detection and quantification methodologies has been documented, there has been limited research focused on sensing location, or where the sensor is directed, to increase detection capability. Studies have included scaling drone and other aerial-based platforms and sensors to the resolutions and bands detected by satellites, but little research has been conducted on placement of the sensors within the crop, as most systems look directly downward onto the canopy (Clark 2017). This study also aims to determine the optimal sensing location or combination of locations for a contact near-infrared sensor for the analysis and discrimination of CR infection specifically for use in breeding programs.

In Part A of this study, machine learning models based on near-infrared spectroscopy were developed for CR detection in wheat and successfully exhibited detection capability (Chapter 3). In Part B these models were generalised to determine the impact of combining disease data, both temporally and across sites (Chapter 4). While detection F_1 scores and accuracies dropped in the generalised models, CR detection was again successful.

II. MATERIALS AND METHODS

A. Trial site, plant cultivation and experimental design

This study is part of a set of objectives derived from data from five CR trials. As such, the methodologies are briefly laid out in the following sections and are fully outlined in Part A of this paper series (Chapter 3). Three glasshouse trials and two field trials were conducted as part of this research. The three glasshouse trials were conducted in 2018 and 2019 at the Leslie Research Centre in Rockville, QLD and the University of Southern Queensland's Centre for Crop Health (USQ-CCH) in Darling Heights, QLD as described previously (Chapter 3, 4). Briefly, in the first glasshouse trial in 2018 four genotypes with known resistances to CR, known in the breeding community as standards, were examined in inoculated and non-inoculated pots with twelve

replicates, totaling 96 plants. In the final two glasshouse trials, five standards were observed over six replicates of inoculated and non-inoculated pots, for 60 total plants. Each plant was scanned throughout the season and was removed from the pot after flowering for manual scoring at the USQ-CCH research facilities. All trials were controlled at 20–25° C and plants were watered once per week to field capacity. Infection was achieved by applying *F. pseudograminearum* colonised wheat grain inoculum, at the two-leaf stage, to individual plant coleoptiles as outlined in (Percy et al. 2012). At maturity plants were harvested and removed from the glasshouse for disease severity assessment.

Two field trials were conducted in 2018 and 2019 at the Tosari Research Station, QLD (-27.859964, 151.452766). These trials consisted of three replicates of the same standards as the glasshouse trials, with plus and minus inoculated plots paired and arranged in a randomised strip plot design. Six plants from each plot were scanned throughout the season and were removed from the ground at flowering for manual scoring at the USQ-CCH research facilities. *F. pseudograminearum* colonised millet inoculum was applied above the seed in furrow at planting to inoculated plots in field trials. All glasshouse and field trials were statistically designed in collaboration with biostatisticians at Queensland Department of Agriculture and Fisheries (QDAF).

B. Near-infrared measurements

Near-infrared hyperspectral scans were collected by a handheld contact sensor, the DLP® NIRscan™ Nano (Texas Instruments, USA). Scans were taken from the centre of the youngest leaf of the newest emerged tiller, the flag leaf of the newest emerged tiller, the centre-most leaf and the grain head, when present. Grain heads began to appear across a few early maturing plants at weeks 7–9, only in glasshouse trials. Leaf measurements were collected once a week throughout the growing season from Z 13, 21. No data was collected in week 8 in the second glasshouse trial due to technical issues during collection. All other measurement dates were situated within 8 days of each other. Measurements were calibrated with 10% grey, 60% grey and 99% white reference Spectralon® panels, measured at the beginning of sampling. Measurements were corrected as described:

$$Corrected\ Reflectance = Calibration\ Coeffient \left(\frac{Target\ Radiance}{Reference} \right)$$

(Labsphere 1998)

C. Data grouping

Data was grouped into individual weekly datasets and a combined dataset, comprising all measurements across all trials for evaluation of any temporal patterning. For spatial patterning, data was grouped based on location of measurement. Groups were tiller, centre, flag and head. A combined group was also created for the spatial data, consisting of combined data across all locations, both including and excluding head data.

D. Data analysis

All data analysis was performed using the Python computing environment (Python version 3.6.8, Python Software Foundation 2019). Data analysis was primarily limited to the SciPy ecosystem (Jones et al. 2001) and the Scikit-learn library (Pedregosa et al. 2011). Any additional libraries and packages have been cited where relevant.

Spectral signatures were analysed and any outliers (z-score > 3) were removed. The resulting signatures were transformed using the discrete wavelet transform (DWT), using a Daubechies 2 wavelet from the PyWavelets library (Lee et al. 2019). This wavelet is asymmetric and biorthogonal (Daubechies 1988).

The DWT approach (1) uses a scaled operator to return transformed data features indicative of significant signatures at both local and global scales:

$$\omega_s f(\lambda) = f(\lambda) * \psi_s(\lambda) = f(\lambda) * \frac{1}{s}\psi\left(\frac{\lambda}{s}\right) \tag{1}$$

(Bruce & Li 2001)

where $f(\lambda)$ is the input signal, s is the scale and $\psi_s(\lambda)$ is the scaled wavelet function. Resulting transformed spectra were standardised and scaled in the Python SciPy library (Jones et al. 2001).

Any resulting features with bandwidths less than 25 nm were discarded to reduce model noise. Feature selection was based upon the SelectKBest function in the feature_selection package of the Scikit-learn library. The top k, 1–10 features were selected using the k highest ANOVA F-values (Pedregosa et al. 2011).

Six machine learning techniques were evaluated for effectiveness in classifying pots or plots into diseased (inoculated) or healthy (non-inoculated) plants including logistic regression classification, k-nearest neighbors, decision trees, extreme random forests, support vector machines and a multilayer perceptron artificial neural network. The multilayer perceptron was determined to be the most effective method of discriminating diseased from healthy plants (Chapter 3, 4). Multilayer perceptron, artificial neural networks (MLP, ANN) are classification algorithms that evaluate training data without specific programing to discover unique characteristics, otherwise overlooked from the input training data set. A MLP consists of a minimum of three layers of nodes, an input, a hidden layer and an output layer. MLPs are a class of feedforward ANN, which use backpropagation to distinguish non-linear data patterns (Wasserman & Schwartz 1988).

D.1 Evaluation and validation

Data was collected and evaluated for nine weeks after inoculation for each of the trials using the wavelet-machine learning approach. Each dataset was split into training and test sets using an 80-20% split using Scikit-learn's train_test_split function (Pedregosa et al. 2011). These data feature sets were trained to determine an optimal combination of hyperparameters using the GridSearchCV function from Scikit-learn's model_selection package (Pedregosa et al. 2011). This function performs a comprehensive search over specified input parameters using k-fold cross-validation to determine optimal model parameters. The resulting optimised models were tested on random, independent validation sets and scored for accuracy and F_1 (F-score). The equations used for accuracy, precision, recall and F_1 are reported below (2,3,4,5):

$$accuracy\left(y, \hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \cdot 1\left(\hat{y}_i = y_i\right) \qquad (2)$$

Where $\hat{y}_i$ is the predicted value of the ith sample and y_i is the corresponding true value

$$Precision = \frac{tp}{tp + fp} \qquad (3)$$

Where tp is number of true positive values and fp is number of false positive values

$$Recall = \frac{tp}{tp + fn} \qquad (4)$$

Where tp is number of true positive values and fn is number of false negative values

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \qquad (5)$$

D.2 Temporal features

Resulting accuracies were established for each week and the best models for each approach were used to determine temporal differences between the nine weeks evaluated, from one to nine weeks of data collection from Z 13, 21. An ANOVA was performed using the f_oneway function in the stats analysis package of the SciPy ecosystem, version 0.19.1 (Jones et al. 2001). To determine significant differences between weeks a Tukey HSD (Honest Significant Difference) test was performed with a significance threshold of α = 0.05 using the pairwise_tukeyhsd function from the statsmodels.stats.multicomp package, version 0.11.1 (Seabold & Perktold 2010).

D.3 Spatial features

The resulting accuracies determined for each of the six machine learning methods were also used to discriminate spatial differences between the four sensing locations tested, the newest tiller, centre-most leaf, flag leaf and grain head, as well as, combined data from all locations. The combined data group was split to both include and exclude head data. Both the single top model accuracies for each location and groups of the top twenty accuracies were evaluated to determine significant differences between locations. An ANOVA determined significant differences between the sensing locations, both when including grain head information and when it was excluded. To determine which locations performed significantly better than the others a Tukey HSD test was performed with a significance threshold of α = 0.05.

D.4 Temporal and spatial interaction

Finally, accuracies determined for each of the machine learning techniques were used to evaluate the interaction between sensing locations and timepoint. Interactions were evaluated at the individual trial level and on combined trial data. The top twenty accuracies were used to evaluate interaction variances between the temporal and spatial data using a two-factor ANOVA, significant interactions between sensing location and timepoint were observed. The two-factor ANOVA was performed using the anova_lm function in the statsmodels.stats.anova package, version 0.11.1 (Seabold & Perktold 2010). To determine which combinations of weeks and sensing location

were significant a Tukey HSD test was performed ($\alpha = 0.05$) using the pairwise_tukeyhsd function included in the statsmodels.stats.multicomp package, version 0.11.1 (Seabold & Perktold 2010).

III. RESULTS

A. Temporal effects

Timepoints which were significantly different from one another ($\alpha = 0.05$) were evaluated across trials and combined data. When evaluating the initial glasshouse trial, 2 and 4 weeks from Z 13, 21 were identified as temporally significant. In the second glasshouse trial, week 2 was again significant with the addition of week 1 and week 5, while the final glasshouse trial showed significant differences in weeks 1 and 7 of data collection (Table 1).

Both field trials had significant temporal effects at week 5 of data collection, with the second trial also showing significant differences in week 1. When looking at combined data, weeks 2, 4, 5 and 9 post-identification of Z 13, 21 were determined to be significantly advantageous in discriminating CR infection in an equal number of cases (Table 1).

Overall, timepoints of significance appeared across most trials in weeks 1–2 weeks after identification of Z 13, 21, at 4–5 weeks and again after 9 weeks post-identification of Z 13, 21 (Table 1). These timepoints correspond to important events in the wheat lifecycle, specifically early-tillering (1–2 weeks), late-tillering (4–5 weeks) and the boot stage (9 weeks).

Table 1: Weeks of data collection that were significantly different for each glasshouse and field experiment and for all data combined as determined by ANOVA and Tukey HSD (α=0.05). Week comparisons that did not exhibit significant differences were omitted from this table.

Multiple Comparison of Means, Tukey HSD, α = 0.05

Data grouping	Group 1	Group 2	Mean difference	p-adj	Lower	Upper
Glasshouse 1	Week 2	Week 4	-15.345	0.001	-24.9789	-5.7111
	Week 2	Week 9	-11.4383	0.0097	-21.0722	-1.8045
	Week 4	Week 5	11.6033	0.0082	1.9695	21.2372
Glasshouse 2	Week 1	Week 3	-28.8883	0.001	-41.4957	-16.281
	Week 1	Week 4	-37.9067	0.001	-50.514	-25.2993
	Week 1	Week 5	-19.1467	0.001	-31.654	-6.4393
	Week 1	Week 6	-31.9017	0.001	-44.509	-19.2943
	Week 1	Week 7	-38.9267	0.001	-51.534	-26.3193
	Week 1	Week 9	-34.1383	0.001	-46.7457	-21.531
	Week 2	Week 3	-20.3183	0.001	-32.9257	-7.711
	Week 2	Week 4	-29.3367	0.001	-41.944	-16.7293
	Week 2	Week 6	-23.3317	0.001	-35.939	-10.7243
	Week 2	Week 7	-30.3567	0.001	-42.964	-17.7493
	Week 2	Week 9	-25.5683	0.001	-38.1757	-12.961
	Week 5	Week 4	18.86	0.001	6.2527	31.4673
	Week 5	Week 6	-12.855	0.0429	-25.4623	-0.2477
	Week 5	Week 7	-19.88	0.001	-32.4873	-7.2727
	Week 5	Week 9	-15.0917	0.0096	-27.699	-2.4843
Glasshouse 3	Week 1	Week 2	18.82	0.0303	1.0649	36.5751
	Week 1	Week 7	27.6167	0.001	9.8615	45.3718
	Week 1	Week 9	20.04	0.0166	2.2849	37.7951
	Week 7	Week 5	20.51	0.013	2.7549	38.2651
	Week 7	Week 8	-24.0717	0.0019	-41.8268	-6.3165
Field 1	Week 1	Week 5	14.2233	0.0025	3.5176	24.929
	Week 3	Week 8	-11.4383	0.0282	-22.144	-0.7326
	Week 5	Week 7	-12.305	0.0137	-23.0107	-1.5993
	Week 5	Week 8	-15.0267	0.0012	-25.7324	-4.321
	Week 5	Week 9	-11.8017	0.021	-22.5074	-1.096
Field 2	Week 1	Week 5	15.4783	0.001	7.561	23.3957
	Week 1	Week 6	9.525	0.0083	1.6077	17.644
	Week 1	Week 7	8.5633	0.0251	0.646	16.4807
	Week 1	Week 8	10.715	0.0019	2.7977	18.6323
	Week 5	Week 2	9.7267	0.0066	1.8093	17.644
	Week 5	Week 3	10.4433	0.0027	2.526	18.3607
	Week 5	Week 4	9.02	0.0151	1.1027	16.9373
	Week 5	Week 9	-10.7183	0.0019	-18.6357	2.801
Combined (All Experiments)	Week 2	Week 4	-10.5307	0.0061	-19.2583	-1.8031
	Week 2	Week 8	-11.3439	0.002	-20.0715	-2.6163
	Week 5	Week 4	8.983	0.0383	0.2554	17.7106
	Week 5	Week 8	-9.7963	0.0152	-18.5239	-1.0686

B. Spatial effects

Sensing was significantly (α = 0.05) more accurate when evaluating the top twenty models developed on data from tillers, centre leaves and flag leaves, as opposed to individual locations. No significant differences were observed when only evaluating

the top models for each group, in either the head-retained or head-removed datasets. When head data was retained and the top 20 models were evaluated, combined data significantly outperformed tiller and flag data, while head data, alone, also outperformed tiller data (Table 2).

Table 2: Significant differences in sensing location (leaf or head of the center-most tiller, the newest emerged tiller or the youngest flag leaf) across 3 glasshouse and 2 field trials for individual sensing locations and combined data for either the single top performing accuracy for each model or the top twenty accuracies for each model as determined by ANOVA and Tukey HSD. Significance is denoted as 'S' (significant) or 'NS' (not significant).

Multiple Comparison of Means, Tukey HSD, α = 0.05

Data grouping	Group 1	Group 2	Mean difference	p-adj	Lower	Upper	Significance
			Combined Data				
Highest Accuracy - Head Data Removed	Centre	Flag	0.0064	0.9	-0.0943	0.17	NS
	Centre	Combined	0.0038	0.9	-0.0969	0.1044	NS
	Centre	Tiller	0.0013	0.9	-0.0993	0.102	NS
	Flag	Combined	-0.0026	0.9	-0.1033	0.098	NS
	Flag	Tiller	-0.005	0.9	-0.1057	0.0956	NS
	Combined	Tiller	-0.0024	0.9	-0.1031	0.0982	NS
Highest Accuracy - Head Data Retained	Combined	Centre	0.0075	0.9	-0.1285	0.1435	NS
	Combined	Flag	-0.0005	0.9	-0.1365	0.1355	NS
	Combined	Head	-0.0006	0.9	-0.1366	0.1354	NS
	Combined	Tiller	0.0008	0.9	-0.1352	0.1367	NS
	Centre	Flag	-0.008	0.9	-0.144	0.128	NS
	Centre	Head	-0.0081	0.9	-0.1441	0.1279	NS
	Centre	Tiller	-0.0067	0.9	-0.1427	0.1293	NS
	Flag	Head	-0.0001	0.9	-0.1361	0.1359	NS
	Flag	Tiller	0.0013	0.9	-0.1347	0.1372	NS
	Head	Tiller	0.0013	0.9	0.1373	0.13763	NS
Top 20 Accuracies - Head Data Removed	Centre	Flag	-0.0099	0.6776	-0.0333	0.0136	NS
	Centre	Combined	0.0366	0.001	0.0132	0.06	S
	Centre	Tiller	-0.0163	0.2791	-0.0397	0.0071	NS
	Flag	Combined	0.0464	0.001	0.023	0.0699	S
	Flag	Tiller	-0.0064	0.8888	-0.0299	0.017	NS
	Combined	Tiller	-0.0529	0.001	-0.0763	-0.0295	S
Top 20 Accuracies - Head Data Retained	Combined	Centre	-0.0326	0.1376	-0.0709	0.0057	NS
	Combined	Flag	-0.0421	0.0231	-0.0804	-0.0037	S
	Combined	Head	-0.0151	0.7932	-0.0534	0.0232	NS
	Combined	Tiller	-0.062	0.001	-0.1003	-0.0237	S
	Centre	Flag	-0.0095	0.9	-0.478	0.0289	NS
	Centre	Head	0.0175	0.6986	-0.0208	0.0558	NS
	Centre	Tiller	-0.0294	0.2222	-0.0677	0.0089	NS
	Flag	Head	0.0269	0.3072	-0.0114	0.0652	NS
	Flag	Tiller	-0.0199	0.5985	-0.0583	0.0184	NS
	Head	Tiller	-0.0469	0.0076	-0.0852	-0.0086	S

C. Temporal and spatial interaction

Significant variation ($\alpha = 0.05$) was seen in the interaction between timepoint (week) and sensing location (Table 3). Significant interactions were observed across all individual trials and combined data, with temporal and spatial interactions accounting for the majority of variance in four of the six data groupings (Eta^2, ω^2). This interaction was most important in analysis of combined data across all trials. Mean accuracies for each significant interaction as determined by Tukey's HSD test ($\alpha = 0.05$) ranged from 60% to 78% (Fig. 1). The interaction with the highest mean accuracy occurred at week five using combined scan data from the centre-most leaf, the newest emerged tiller leaf and the flag leaf.

Table 3: Significant differences of sensing location (leaf or head of the center most tiller, the newest emerged tiller or the youngest flag leaf), timepoint ('Week') and interaction ('Location:Week') between timepoint and sensing location across 3 glasshouse and 2 field trials for individual trials and combined data for the top twenty accuracies for each model as determined by a two-factor ANOVA. Significance is denoted as 'S' (significant) or 'NS' (not significant).

Two-Factor ANOVA, $\alpha = 0.05$

Data Grouping	Variable	Sum of Squares	df	F	PR (>F)	Eta^2	ω^2	Significance
Glasshouse 1	Location	1.156	5	55.494	5.25E-51	0.074	0.072	S
	Week	3.547	8	106.472	1.95E-128	0.226	0.224	S
	Location:Week	6.813	40	40.893	3.08E-181	0.434	0.423	S
	Residual	4.194	1007					
Glasshouse 2	Location	0.002	5	0.032	0.968	4.33E-05	-0.001	NS
	Week	28.711	7	343.110	4.60E-143	0.647	0.644	S
	Location:Week	6.154	35	14.709	1.16E-57	0.139	0.129	S
	Residual	9.540	798					
Glasshouse 3	Location	0.190	5	7.158	0.001	0.014	0.012	S
	Week	1.816	8	42.760	8.35E-40	0.130	0.127	S
	Location:Week	6.945	40	32.702	9.49E-143	0.496	0.481	S
	Residual	5.044	950					
Field 1	Location	0.625	4	47.244	6.60E-36	0.074	0.073	S
	Week	3.952	8	149.402	1.10E-156	0.470	0.466	S
	Location:Week	1.012	32	9.569	9.88E-39	0.120	0.108	S
	Residual	2.827	855					
Field 2	Location	0.781	4	66.971	2.57E-49	0.143	0.140	S
	Week	0.749	8	32.117	2.75E-44	0.137	0.133	S
	Location:Week	1.451	32	15.552	6.53E-65	0.265	0.248	S
	Residual	2.493	855					
Combined (All Experiments)	Location	2.759	5	16.565	3.02E-16	0.016	0.015	S
	Week	5.240	8	19.663	1.95E-29	0.030	0.029	S
	Location:Week	10.051	40	7.542	2.56E-40	0.058	0.050	S
	Residual	154.806	4647					
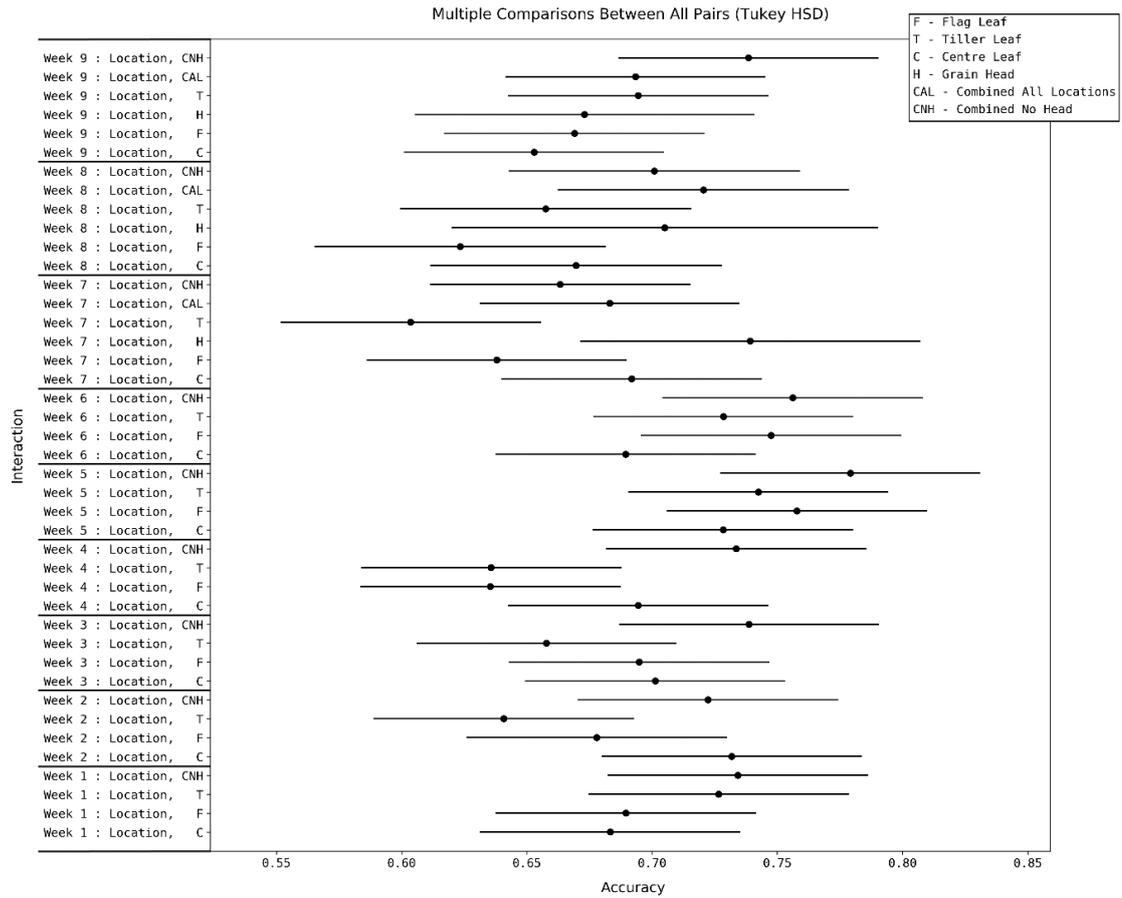
Fig. 1: Group means (●) and confidence intervals (–) of combined data of interactions between location and timepoint determined by Tukey HSD to be significant.

IV. DISCUSSION

A. Temporal effects

This paper describes significant ($\alpha = 0.05$) temporal effects as early as one to two weeks from identification of Z 13, 21 in wheat inoculated with *F. pseudograminearum* (Table 1). This suggests not only potential for early disease detection, but that early disease detection may be more accurate than CR detection later in the growing season. The temporal patterning observed in these experiments indicates a decrease in disease detection accuracy after weeks six to seven of measurement with accuracies increasing again around the eight to nine-week mark. Further work will develop these specific temporal patterns into features for model refinement and further model development which has been shown to increase model accuracy over single timepoint models (Anderegg et al. 2019).

Model accuracy fluctuations may correspond to pathogen toxin production and/or plant disease host responses. *F. pseudograminearum* is known to produce deoxynivalenol, 3-acetyldeoxynivalenol, and zearalenone, which may be detectible by near-infrared sensing (Leslie & Summerell 2006). Previous work has shown deoxynivalenol may contribute to shifts in the near-infrared absorption between healthy and *Fusarium* damaged wheat kernels (Peiris et al. 2009). Additionally, it has been hypothesised that when under attack by certain CR pathogens (*F. culmorum*), wheat synthesises an anti-fungal compound, which may be an additional target for near-infrared spectral sensing (Skov et al. 2004). These shifts may be more apparent at certain timepoints and as such may be responsible for the temporal patterning seen here. Additional biochemical research should be carried out to determine wavebands and assess signatures and quantities of isolated toxins and biological defense chemicals produced from green tissue for comparison to discovered temporal patterns. Further research is required to determine the relationship between these temporal differences in NIR signatures and the pathogen biology and disease cycle, in addition to furthering knowledge of cereal CR resistance and tolerance mechanisms.

B. Spatial effects

Significant ($\alpha = 0.05$) spatial effects were observed between sensing locations when comparing accuracies from the top twenty models in each group (tiller, centre, flag, head, combined). Tiller scans were taken from the centre of the youngest leaf of the newest emerged tiller; flag scans from the flag leaf of the newest emerged tiller; centre scans from the centre-most leaf. When evaluating differences involving these groups for each machine learning technique, using a combination of measurements throughout the canopy profile a combination of tiller, centre and flag measurements returned significantly higher accuracies than any of those measurements alone. This suggests that a combination of measurements, throughout the canopy should be used when discriminating CR before heads begin to emerge. When head data was included in the evaluation of the top twenty model groups, combined data again proved to perform better in most instances.

Interestingly, head data alone was shown to significantly outperform all data collected on leaves (Table 2). This could be due to detection of decreased grain fill or quality in the wheat kernels, increased moisture stress, or molecular changes caused by the *Fusarium* infection, as discussed above in *Temporal effects*. While this suggests

potential for the use of head data in disease detection and discrimination, if disease discrimination is desired early in the season, such as by a breeding program for use in quickly selecting germplasm to advance, models built on other measurements should be considered. Further, the results of this study indicate significant temporal effects early in the season using young tissues, consistent with findings reported in Percy et al. (2012), whereby the disease is shown to be progressive in its movement through each plant part. However, results of this study suggest that using a combination of scans produces significantly better results than using scans of a single tissue. These results indicate that the best potential combination of temporal and spatial parameters for near-infrared CR discrimination is measurement at late-tillering (week 5) using a combination of leaf scans throughout the canopy. Further, these results indicate that when transferring this system to a camera-based approach, images should be obtained including as much leaf information as possible, versus targeting a specific leaf.

Results of the data groupings and analysis conducted in this paper are consistent with Part A of the paper series, where increased accuracies and F_1 scores were achieved at 1–2 weeks, 5–6 weeks and 9 weeks of data collection (Chapter 3). These timepoints, corresponding to the wheat growth stages of early-tillering, late-tillering and booting, respectively, are shown here to be significant in the successful detection of CR, in conjunction with scans throughout the canopy, with an emphasis on young tissue scans and early identification.

C. Temporal and spatial interaction

Significant interaction ($\alpha = 0.05$) was observed between sensing location and timepoint when the top twenty accuracies were evaluated across machine learning methods (Table 3). Tukey HSD ($\alpha = 0.05$) determined that in early weeks, until week 6, significant interaction occurred between the weeks and all location measures. After week 6, the most significant interactions consistently involved grain head or combined sensing data (Fig. 1). This suggests that scan location increases in importance later in the season. These findings indicate that for optimal CR detection, the interaction between timing and location of sensing measurements is important to consider. This further suggests, in conjunction with the points discussed above, that as the disease is progressive, optimal detection time and sensing location move like a wave with this progression. As such, using a combination of sensing locations, such as observed from

a camera system, and data from early-tillering, late-tillering and booting significantly impacts and increases CR detection capability.

Further work should evaluate the accuracies of models targeting specific tillers throughout the growing season to determine how disease detection is impacted by tiller age and size. Additionally, this evaluation could be used to model disease progression through the host crop to improve our understanding of these pathogen-host interactions. Stem measurements, at the site of lesions, should also be included to determine suitability for rapid disease detection and quantification. However, while stem measurements may be suitable for post-harvest quantification, they are less likely to be adaptable to either real-time or in-season remote sensing systems as they are rapidly hidden by the crop canopy closing and would likely require a system that travels up every other row to image stem browning, as opposed to an overhead system, like those commonly integrated into unmanned aerial vehicle (UAV) platforms.

V. CONCLUSION

In this study, temporal and spatial features were evaluated for significance in *F. pseudograminearum* induced CR detection in bread wheat using near-infrared hyperspectral spectroscopy, pre-processed with the discrete wavelet transform (DWT) and discriminated with machine learning models. Results of this further research indicate significant temporal and spatial effects to consider when developing CR detection frameworks. Significant temporal patterning was observed at one to two weeks from Z 13, 21 in plants inoculated with *F. pseudograminearum*, dropping off and rising again around the four to five-week mark and again rising at the nine-week mark, consistent with observations reported in Part A of this study. This patterning should be investigated in future studies for use as an input feature to further refine disease models. Investigation of spatial features revealed that combinations of sensing locations throughout the canopy performed better than individual features. Further, the use of grain heads for sensing showed potential for disease discrimination, compared to tillers. However, combined features provide an advantage over grain head features when early detection is the goal, as with developing high-throughput breeding programs. Further work should investigate how these differences, both temporally and spatially, relate to pathogen toxicity, host response and disease lifecycle.

REFERENCES

Anderegg, J, Hund, A, Karisto, P & Mikaberidze, A 2019, 'In-field detection and quantification of Septoria tritici blotch in diverse wheat germplasm using spectral-temporal features', *BioRxiv*, p. 664011.

Bauriegel, E & Herppich, WB 2014, 'Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to Fusarium spec. infections on wheat', *Agriculture*, vol. 4, no. 1, pp. 32-57.

Bruce, LM & Li, J 2001, 'Wavelets for computationally efficient hyperspectral derivative analysis', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1540-6.

Busemeyer, L, Ruckelshausen, A, Möller, K, Melchinger, AE, Alheit, KV, Maurer, HP, Hahn, V, Weissmann, EA, Reif, JC & Würschum, T 2013, 'Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation', *Scientific reports*, vol. 3, p. 2442.

Clark, ML 2017, 'Comparison of simulated hyperspectral HyspIRI and multispectral Landsat 8 and Sentinel-2 imagery for multi-seasonal, regional land-cover mapping', *Remote Sensing of Environment*, vol. 200, pp. 311-25.

Daubechies, I 1988, 'Orthonormal bases of compactly supported wavelets', *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909-96.

Franke, J & Menz, G 2007, 'Multi-temporal wheat disease detection by multi-spectral remote sensing', *Precision agriculture*, vol. 8, no. 3, pp. 161-72.

Jones, E, Oliphant, T & Peterson, P 2001, 'SciPy: Open Source Scientific Tools for Python'.

Labsphere, 'Reflectance characteristics of Spectralon panels', Reflectance Calibration Laboratory, Sutton, NH, USA. (1998).

Lee, GR, Gommers, R, Waselewski, F, Wohlfahrt, K & O'Leary, A 2019, 'PyWavelets: A Python package for wavelet analysis', Journal of Open Source Software, vol. 4, no. 36, p. 1237.

Leslie, JF & Summerell, BA 2006, 'Species Descriptions', in *The Fusarium Laboratory Manual*, Wiley-Blackwell, ch 13, pp. 121-278.

Parry, D, Pettitt, T, Jenkinson, P & Lees, A 1994, 'The cereal Fusarium complex', *Ecology of plant pathogens*, pp. 301-20.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R & Dubourg, V 2011, 'Scikit-learn: Machine

learning in Python', *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-30.

Peiris, KH, Pumphrey, MO & Dowell, FE 2009, 'NIR absorbance characteristics of deoxynivalenol and of sound and Fusarium-damaged wheat kernels', *Journal of Near Infrared Spectroscopy*, vol. 17, no. 4, pp. 213-21.

Percy, C, Wildermuth, G & Sutherland, M 2012, 'Symptom development proceeds at different rates in susceptible and partially resistant cereal seedlings infected with Fusarium pseudograminearum', *Australasian Plant Pathology*, vol. 41, no. 6, pp. 621-31.

Python version 3.6.8, Python Software Foundation 2019.

Seabold, S & Perktold, J 2010, 'Statsmodels: Econometric and statistical modeling with python', in *Proceedings of the 9th Python in Science Conference*, Austin, TX, p. 61.

Simpfendorfer, S, McKay, A & Ophel-Keller, K 2019, 'New approaches to crop disease management in conservation agriculture. In (Eds J Pratley and J Kirkegaard) "Australian Agriculture in 2020: From Conservation to Automation" pp 173-188 (Agronomy Australia and Charles Sturt University: Wagga Wagga)', pp. 173-88.

Skov, J, Lemmens, M & Giese, H 2004, 'Role of a Fusarium culmorum ABC transporter (FcABC1) during infection of wheat and barley', *Physiological and molecular plant pathology*, vol. 64, no. 5, pp. 245-54.

Wasserman, PD & Schwartz, T 1988, 'Neural networks. II. What are they and why is everybody so interested in them now?', *IEEE Expert*, vol. 3, no. 1, pp. 10-5.

CHAPTER 6

DEVELOPMENT OF MULTILAYER PERCEPTRON MODELS FOR HIGH-THROUGHPUT PHENOTYPING OF CROWN ROT, *FUSARIUM PSEUDOGRAMINEARUM*

In this study, multilayer perceptron artificial neural networks (MLP, ANN), preprocessed for feature reduction with either a discrete wavelet transform (DWT) or principal component analysis (PCA) were used to develop models to phenotype *Fusarium pseudograminearum* induced crown rot of wheat by quantifying disease using six developed scales. The scales correlate to percentage stem browning and were tested across glasshouse and field environments, with models developed from individual trials and combined site data. Models were shown to perform best in field environments. PCA and DWT approaches performed with similar accuracy in developed models.

Humpal J., Percy C., McCarthy C., & Thomasson AJ. (2020d). Development of multilayer perceptron models for high-throughput phenotyping of crown rot, *Fusarium pseudograminearum*. This chapter was prepared according to the instructions to authors given by **Computers and Electronics in Agriculture**.

Development of multilayer perceptron models for high-throughput phenotyping of crown rot, *Fusarium pseudograminearum*

Jacob Humpal[*], Cassandra Percy[**], Cheryl McCarthy[*], J. Alex Thomasson[***]

[*] Centre for Agricultural Engineering, University of Southern Queensland, Toowoomba, Queensland, Australia

[**] Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia

[***] Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, Mississippi, United States

J. Humpal, C. Percy (*)

Centre for Agricultural Engineering, University of Southern Queensland,

Toowoomba, Qld 4350, Australia

*e-mail: cassy.percy@usq.edu.au

Ph: +61 7 4631 1241

Abstract- In this paper, multilayer perceptron artificial neural networks (MLP, ANN), preprocessed with a discrete wavelet transform (DWT) or principal component analysis (PCA) were used to build models to quantify *Fusarium pseudograminearum* induced crown rot (CR) in bread wheat. Six scales were developed and tested, correlating to categories of percentage stem browning, traditionally used to quantify resistance and tolerance of wheat to CR. The scales were tested across three glasshouse and two field trials, with models being developed from individual trial data, as well as, combined data from all experiments. Average accuracy ranged from 32.17% to 96.40% with models performing better with field data than in the glasshouse. Both PCA and DWT approaches performed similarly in developed models.

Keywords- *crop disease, crown rot, ANN, proximal sensing, DWT*

I. INTRODUCTION

Fusarium pseudograminearum is the predominant pathogen responsible for CR infection in Australia. Recent reports indicate that CR, traditionally a significant issue in winter cereals in the northern grains region of Australia, is becoming more widespread, with increasing significance across the Northern, Western and Southern regions (Simpfendorfer et al. 2019). As CR is a stubble-borne disease, the increasing adoption of conservation agriculture has allowed inoculum to build up in the soil, particularly in rotations with limited cycles out of host crops (Simpfendorfer et al. 2019).

There are currently no viable options for fungicide control of CR infection, although targeting application to the base of the crop using inter-row drop applicators has been reported to show some success in yield retention (Simpfendorfer et al. 2014). As the most successful management approaches involve rotation into a non-host crop or inter-row seeding, genetic solutions to CR are a priority to Australian grain growers.

Current breeding programs must grow plants through to maturity to screen for tolerance and resistance. Tolerance is the ability of germplasm to resist yield loss under disease pressure while resistance is the ability of germplasm to combat the disease directly (Forknall et al. 2019), reducing the spread, whether that be through the production of anti-fungal compounds or other host responses. A machine vision system that could determine CR severity early in the season would allow for rapid screening of resistance and would also be a step toward the development of a high-throughput phenotyping (HTP) platform for early resistance and tolerance screening based on machine vision technologies. Early detection would allow for only the most promising material to be moved forward through the program, freeing capital for investment in additional lines in each step of germplasm development.

Additional advantages to developing machine-based HTP methodologies include shifting from reliance on human interpretation of visual disease symptoms. Traditional CR assessment, in breeding programs, relies on removing the plant from the field, often off-site, to score percentage stem browning around the lower internodes. However, human assessment, particularly in plant disease rating, has been shown to contain significant variability, both between samples scored by an individual assessor and between multiple assessors (Nutter Jr et al. 1993; Newton & Hackett 1994; Nutter

1997). While this variability has been hypothesised to be due to imperfect scoring systems, it is likely to be specifically due to the variation between human colour interpretations in CR scoring systems, an issue evident in any interpretive, colour-based system. Further, traditional CR assessment is time consuming. A machine-vision system would allow trained assessors to reproduce scores more reliably and more quickly.

Previous research has reported the successful detection of CR using multiple machine learning methods and processing techniques on data obtained from a contact near-infrared sensor (900–1700 nm) (Chapter 3, 4, 5). Successful disease prediction models were developed and evaluated from both site-specific data and generalised data collected across multiple trial sites and timepoints (Chapter 3, 4). Additionally, significant differences in sensing location, as well as, temporal and locative-temporal differences in detection accuracy were determined (Chapter 5).

This study was conducted to expand upon these findings and examine the phenotyping capability of these approaches. The previous chapters used analysis to determine detection (positive or negative) capability of developed analyses while this chapter examines the capability for multi-class classification. Specifically, a multilayer perceptron (MLP) approach was developed, preprocessed with either DWT or PCA for dimensionality reduction and feature selection. The goal of this method was to successfully quantify CR using a machine-vision approach which could be beneficial to CR breeding programs.

II. MATERIALS AND METHODS

A. Trial site, plant cultivation and experimental design

Five experimental trials were conducted for this study, including three glasshouse trials and two field trials in collaboration with researchers from the University of Southern Queensland Centre for Crop Health (USQ-CCH), as previously described (Chapter 3). All glasshouse and field trials were statistically designed in collaboration with biostatisticians at Queensland Department of Agriculture and Fisheries (QDAF).

Glasshouse trials were conducted from 2018–2019 at the Leslie Research Centre, Rockville, QLD and at the USQ-CCH, Darling Heights, QLD. The initial glasshouse trial consisted of four standard genotypes, with known resistances to CR. These were examined in 96 inoculated and uninoculated pots for twelve replicates. In the 2nd and

3rd glasshouse trials, five standards were observed over six replicates (60 plants). All experiments were arranged in a randomised block design and grown in a glasshouse which maintained temperature at 20–25° C. Plants were watered to field capacity once per week to minimise external influences. Colonised wheat grain inoculum (Percy et al. 2012) was applied at the two-leaf stage, to individual plant coleoptiles of the inoculated treatments. Plants were harvested at maturity and rated for disease severity.

Two field trials were conducted in the winter of 2018 and 2019 at the Tosari Research Station, QLD (-27.859964, 151.452766). *F. pseudograminearum* colonised millet inoculum was applied to inoculated plots in furrow, above the seed at planting. These trials contained three replicates of the same standards as the glasshouse trials and were arranged in a paired strip plot, where genotypes were placed in a randomised block with paired inoculated and non-inoculated plots. Six plants from each plot were evaluated for nine weeks throughout the season, from Zadoks (Z) 13, 21 (Zadoks et al. 1974), and removed for manual scoring at USQ-CCH.

<div align="center">B. Near-infrared measurements</div>

The DLP® NIRscan™ Nano (Texas Instruments, USA) was used to take near-infrared hyperspectral measurements of inoculated and non-inoculated plants for nine weeks, from Z 13, 21. Measurements were taken from the centre of the leaf of the youngest fully emerged tiller, the newest flag leaf, the determined centre-most leaf and a grain head if present, for both inoculated and non-inoculated plants. Due to technical issues during data collection, no data was collected in week 8 of the second glasshouse trial at Leslie Research Centre. Every other measurement date was within 8 days of each other. Spectral scans were calibrated with 10% grey, 60% grey and 99% white reference Spectralon® panels, measured at the beginning of the sampling session. Measurements were corrected as follows:

$$Corrected\ Reflectance = Calibration\ Coefficient \left(\frac{Target\ Radiance}{Reference} \right)$$

(Labsphere 1998)

C. Data analysis

Data analysis was performed inside the Python computing environment, unless otherwise stated (Python version 3.6.8, Python Software Foundation 2019). Data analysis was primarily conducted using the SciPy ecosystem (Jones et al. 2001) and the Scikit-learn library (Pedregosa et al. 2011). Figures were developed using the seaborn data visualisation library (Waskom et al. 2014) on top of the matplotlib data visualisation library (Hunter 2007). Additional libraries and packages have been cited where appropriate.

C.1 Discrete wavelet transform

Raw spectral signatures from the DLP® NIRscan™ Nano (Nano) were analysed and outliers (z-score > 3) were removed. The resulting cleaned signatures were transformed using the discrete wavelet transform (DWT), with a Daubechies 2 wavelet from the PyWavelets library (Lee et al. 2019). The Daubechies wavelet family is asymmetric and biorthogonal (Daubechies 1988).

The DWT approach (1) uses a scaled operator to return transformed features or wavelets revealing significant patterns at both high-resolution (local) and low-resolution (global scales):

$$\omega_s f(\lambda) = f(\lambda) * \psi_s (\lambda) = f(\lambda) * \frac{1}{s} \psi \left(\frac{\lambda}{s}\right) \tag{1}$$

(Bruce & Li 2001)

where $f(\lambda)$ is the input signal, from the cleaned spectral signature, s is the scale variable and $\psi_s (\lambda)$ is the scaled wavelet function. Transformed spectral features were standardised and scaled (0–1) in the Python SciPy library (Jones et al. 2001).

Resulting features with bandwidths less than 25 nm were removed to reduce model noise caused by local environmental factors. Features were selected based on highest k, 1–10 ANOVA F-values using the SelectKBest function of the feature_selection package in the Scikit-learn library (Pedregosa et al. 2011).

C.2 Principal component analysis

Raw spectra of adjacent wavebands from the Nano were averaged to approximate the wider bandwidths produced in the DWT approach and outliers were removed (z-score > 3). The resulting bands were standardised and scaled (0–1) using the preprocessing package in the Python SciPy library, version 0.19.1 (Jones et al. 2001). Resulting spectra were smoothed using the Savitzky-Golay filter (Savitzky & Golay 1964). The window size was set at 5 bands and a second order polynomial was used by applying the savgol_filter in the Python library SciPy (Jones et al. 2001). PCA was applied to the resulting spectra and restricted to 1–10 principal components using the decomposition package from the Python SciPy library (Jones et al. 2001). Features of maximum importance were determined from the principal components. Resulting features with bandwidths less than 25 nm were removed and transformed features were used in model development.

C.3 Modelling

A multilayer perceptron, artificial neural network (MLP, ANN) technique was evaluated for effectiveness in quantifying plant disease based on six scoring systems, with differing numbers of data groupings from two to eleven groups, associated with traditional stem browning percentage (Fig. 1). ANNs consist of multiple nodes which act like the neurons in an animal's brain, allowing the models to find patterns in the data where driving mechanisms are not fully understood. MLPs are a class of forward feeding ANN, which uses backpropagation to classify data when linear patterns do not exist (Wasserman & Schwartz 1988). Analysis was performed across individual trials, in addition to a dataset containing data combined from all trials. Further, three additional groupings were developed around early-tillering (weeks 1–2), late-tillering (weeks 4–5) and booting (week 9), based on their impact in successful CR detection from previous research (Chapter 3, 4, 5).

Rating System	% Stem Browning																				
	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
1	ND	Group 1 (5)		Group 2 (14)		Group 3 (25)		Group 4 (36)		Group 5 (44)		Group 6 (54)		Group 7 (65)		Group 8 (77)		Group 9 (85)		Group 10 (95)	
2	ND	Group 1 (8)				Group 2 (29)				Group 3 (50)				Group 4 (72)				Group 5 (91)			
3	ND	Group 1 (9)					Group 2 (35)					Group 3 (60)					Group 4 (87)				
4	Group 1 (9)													Group 2 (69)							
5	ND	Group 1 (13)												Group 2 (69)							
6	ND	Group 1 (9)				Group 2 (49)															

Fig. 1. The grouping of the six rating systems developed for evaluation of CR quantification ability and the average stem browning score (%) of samples used to develop each grouping across all trials in parenthesis. Not detectable is denoted as ND.

C.4 Evaluation and validation

Each experimental dataset was split into train-test datasets using an 80-20% split and used as inputs into the MLP models. The splits were determined using the train_test_split function in Scikit-learn (Pedregosa et al. 2011). The transformed signatures were labelled according to the groupings in Figure 1 based on visual ratings by trained assessors. Traditional visual CR rating involves removing the plant from the trial, removing the leaf sheaths around the lower internodes and scoring based on percentage of colour variation of the stem. The resulting datasets were trained to determine the optimal combination of hyperparameters for accurate quantification with the GridSearchCV function from Scikit-learn's model_selection package (Pedregosa et al. 2011). This performs a complete search over specified input parameters using *k*-fold cross-validation to determine best model parameters for a given input dataset. The resulting models were tested and scored for accuracy using random independent validation sets. The equation used for accuracy is reported below (2):

$$accuracy\left(y, \hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \cdot 1\left(\hat{y}_i = y_i\right) \qquad (2)$$

Where $\hat{y}_i$ is the predicted value of the ith sample and y_i is the corresponding true value

III. RESULTS

A. Glasshouse trials

PCA based models returned average accuracies from glasshouse trials ranging from 32.17% when separating the plants into eleven groups (rating system 1), to 80.83% when separating into the three groups, not detectible (ND), less than 25% browning and greater than 25% browning (Fig. 2). DWT based models followed a similar pattern for glasshouse trials with fewer groupings generally performing better than rating systems with more groups. DWT model means ranged from 35.31% for eleven groups to 73.67% for two to three groups (Fig. 2). When using the PCA approach, glasshouse data returned the highest average accuracy, 80.83%, of the groupings for rating system 6 (Fig. 2).
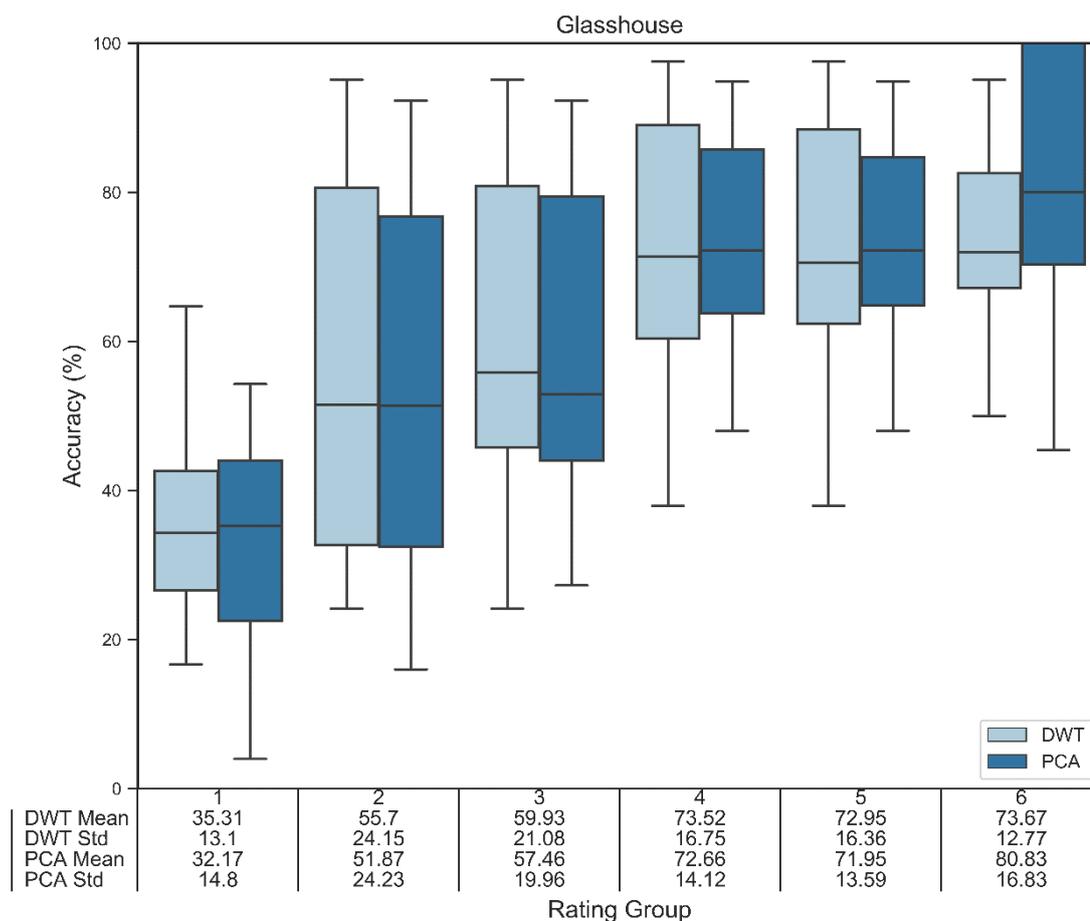
Fig 2. Distribution of accuracies and average mean and standard deviation (std) (table) across machine learning methods for each of the six rating systems (Fig. 1) for both PCA and DWT preprocessing techniques for glasshouse trials.

B. Field trials

Overall, field trial data performed better than all other data groupings across rating systems 1–4 with both PCA and DWT approaches (Fig. 3). Higher accuracies were achieved in field trials than in glasshouse trials (Table 1) with average accuracies of 54.95% to 94.46% for PCA based models and 55.16% to 96.40% for DWT based models. Further, rating systems with larger numbers of categories, rating systems 1–3, performed better under field conditions with mean accuracies from 54.95% to 67.86% (Fig. 3). Under glasshouse conditions rating systems 1–3 returned mean accuracies from 32.17% to 59.93% (Fig. 2). Field trials followed a pattern similar to that seen in the glasshouse trials with groupings with 2–3 categories outperforming other rating systems (Fig. 3).

	1	2	3	4	5	6
DWT Mean	55.16	63.12	66.98	96.4	75.03	67.11
DWT Std	4.47	4.52	7.33	1.4	13.1	7.14
PCA Mean	54.95	65.81	67.86	94.46	73.04	76.17
PCA Std	3.58	8.12	9.84	1.53	13.67	17.44
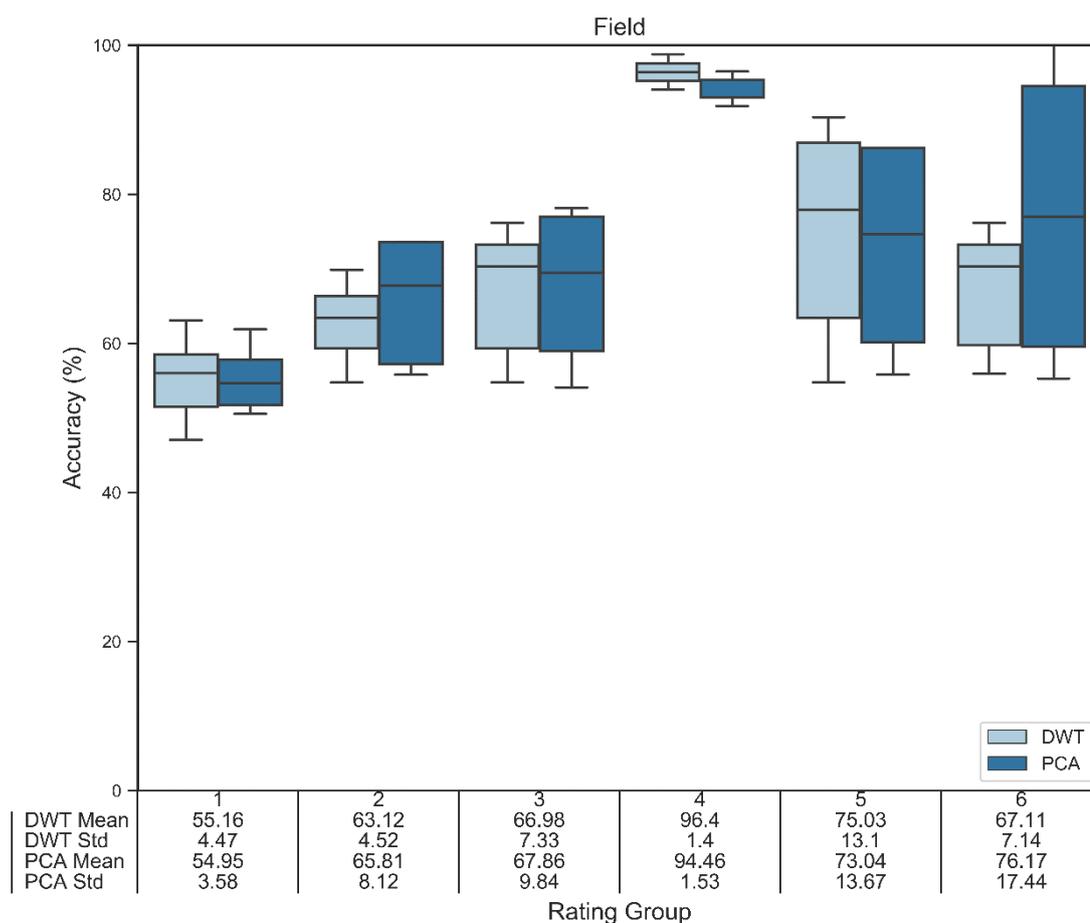
Fig 3. Distribution of accuracies and average mean and standard deviation (std) (table) across machine learning methods for each of the six rating systems (Fig. 1) for both PCA and DWT preprocessing techniques for field trials.

C. Combined Data

When data from all five trial sites was combined, average accuracies followed similar patterns to both glasshouse and field data with higher scores seen when classifying plants into fewer groups (2–3). PCA and DWT approaches performed similarly with means ranging from 41.28–81.38% and 43.25–82.68%, respectively. The highest average accuracy was achieved using rating system 4 (Fig. 4).

	1	2	3	4	5	6
DWT Mean	43.25	58.67	62.75	82.68	73.78	71.04
DWT Std	14.35	19.13	17.19	17.18	15.02	11.25
PCA Mean	41.28	57.45	61.62	81.38	72.39	78.97
PCA Std	16.18	20.5	17.3	15.34	13.48	17.04
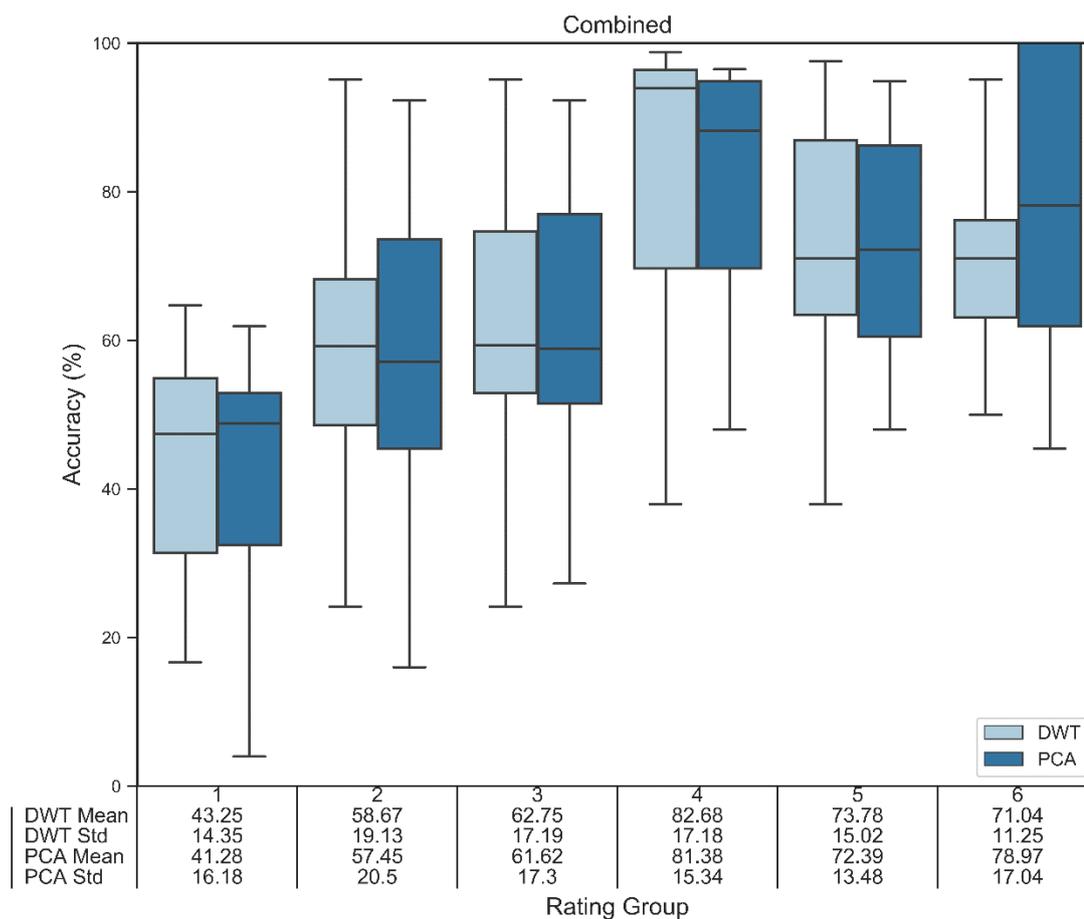
Fig 4. Distribution of accuracies and average mean and standard deviation (std) (table) across machine learning methods for each of the six rating systems (Fig. 1) for both PCA and DWT preprocessing techniques for models built on combined data from all trial sites.

D. Growth Stages

When early-tillering data was evaluated across both glasshouse and field trials, DWT means ranged from 44.53–83.21% while PCA means ranged from 40.66–82.48% (Fig. 5). Late-tillering data presented DWT means of 45.59–83.20% with PCA means from 41.01–82.37% (Fig. 6). Late-tillering performed better than all other data groupings using the DWT approach and rating system 5, with an average accuracy of 76.08% (Fig. 6). Booting data returned DWT means of 42.36–85.38% with PCA means ranging from 42.48–82.23% (Fig. 7). DWT-based approaches returned slightly higher maximum means across all growth stage groupings of 83.20–85.38% compared to 82.23–82.48% (Fig. 5–7). Rating systems containing fewer categories (systems 4–6) returned higher accuracies than systems with larger numbers of categories, consistent to what was seen in the glasshouse, field and combined data groupings. Grouping data

into late-tillering and booting growth stages improved accuracies over the combined data grouping in rating system 6 (Fig. 4, 6, 7).
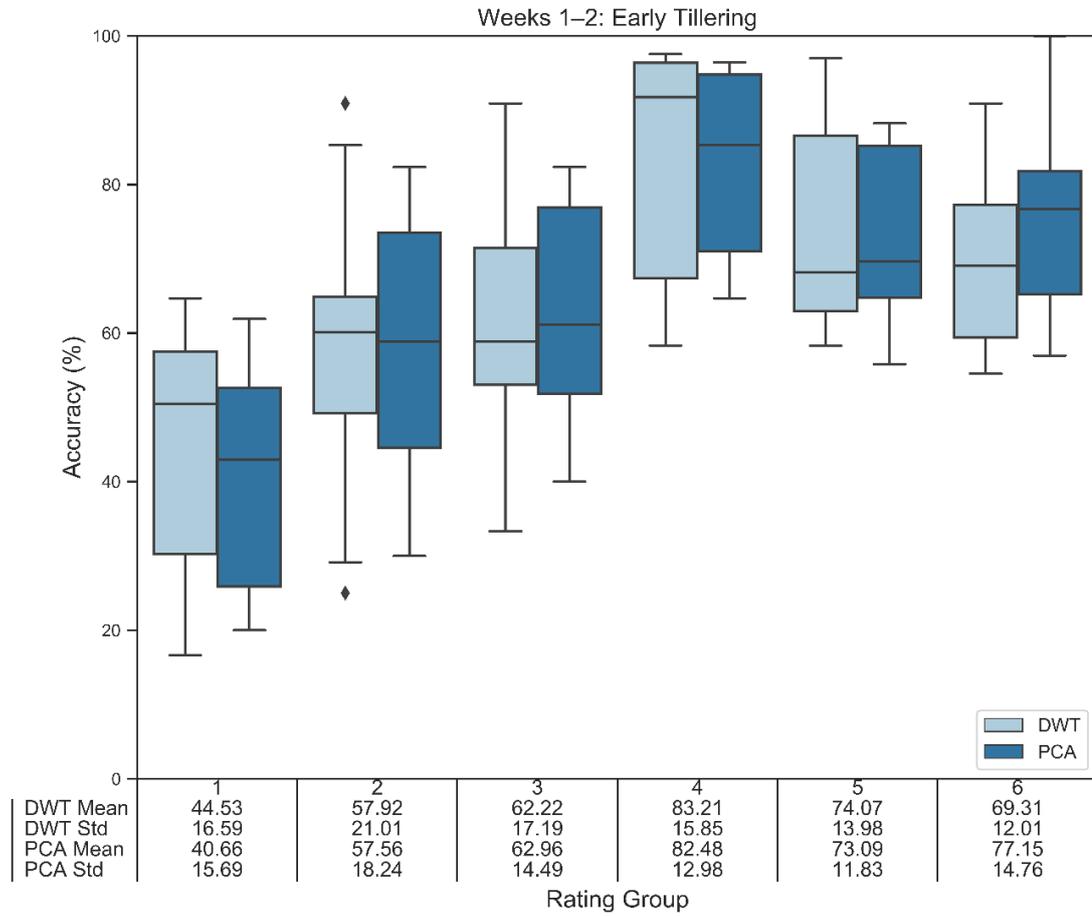
Fig 5. Distribution of accuracies and average mean and standard deviation (std) (table) across machine learning methods for early-tillering for each of the six rating systems (Fig. 1) for both PCA and DWT preprocessing techniques for models built on combined data from all trial sites.
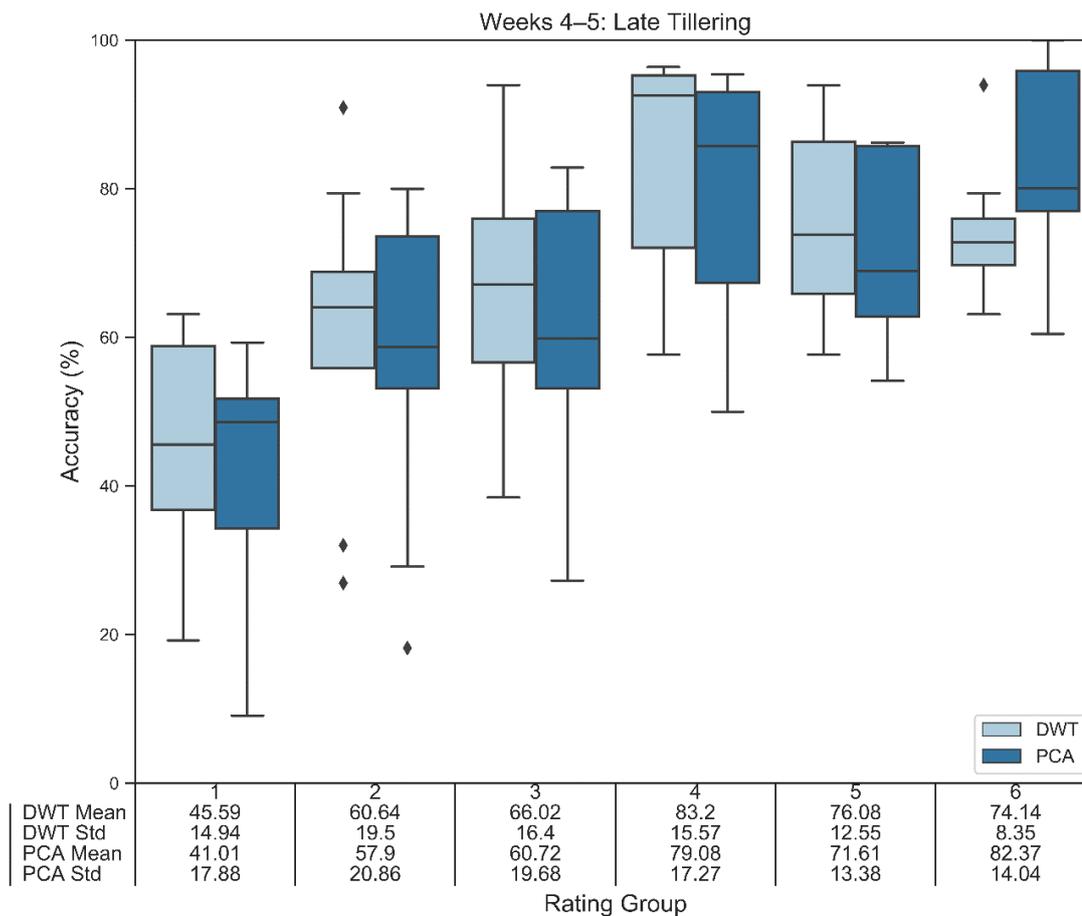
Fig 6. Distribution of accuracies and average mean and standard deviation (std) (table) across machine learning methods for late-tillering for each of the six rating systems (Fig. 1) for both PCA and DWT preprocessing techniques for models built on combined data from all trial sites.

	1	2	3	4	5	6
DWT Mean	42.36	60.9	64.71	85.38	75.61	73.96
DWT Std	16.12	23.92	19.84	15.86	17.83	14.38
PCA Mean	42.48	55.61	59.77	82.23	72.61	80.49
PCA Std	18.42	27.03	23.77	16.26	16.55	27.03
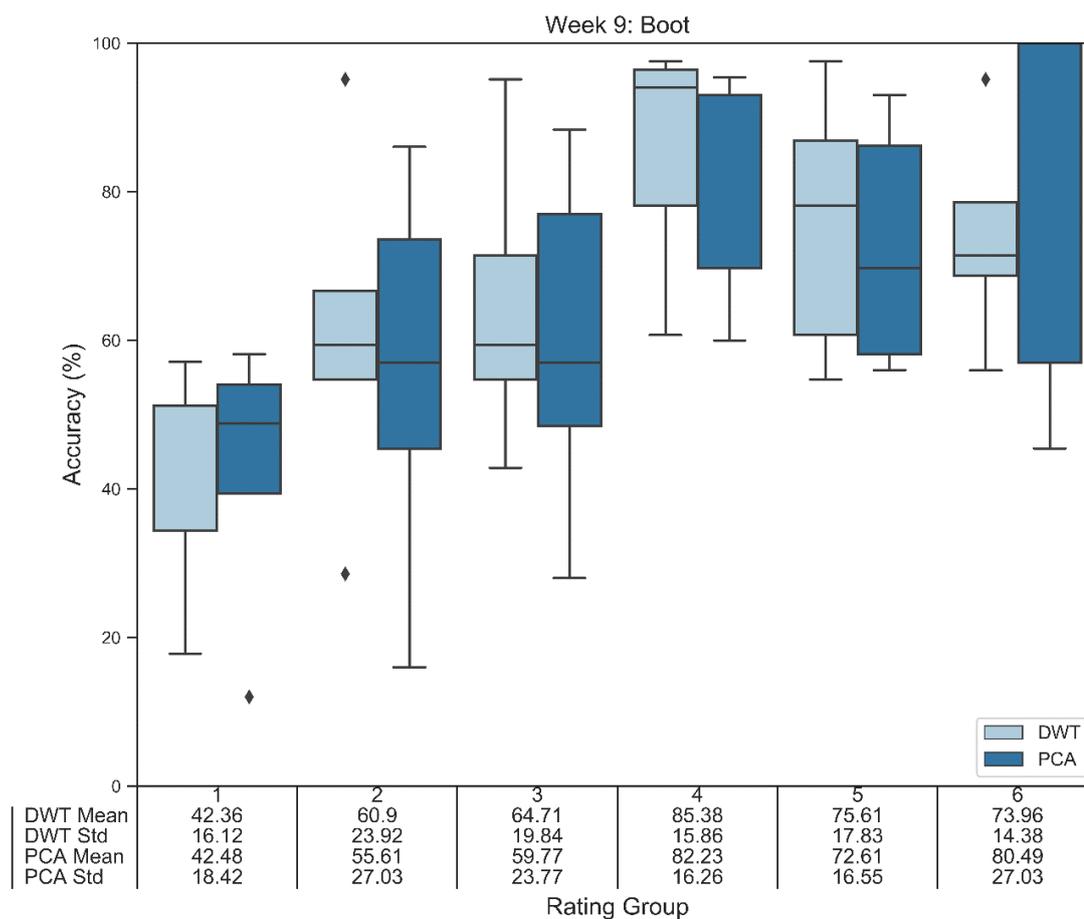
Fig 7. Distribution of accuracies and average mean and standard deviation (std) (table) across machine learning methods for booting for each of the six rating systems (Fig. 1) for both PCA and DWT preprocessing techniques for models built on combined data from all trial sites.

IV. DISCUSSION

The results of this research indicate potential for a CR resistance quantification framework to be developed from the methodologies outlined in this study. A resistance framework is the first step in further developing technologies and methodologies to provide tools for resistance and tolerance quantification to breeding companies and research requiring CR phenotyping. Results indicate that data from field conditions resulted in higher average accuracies in developing quantification models. Germplasm can express CR symptoms differently in different environments including between seasons and also glasshouse and field environments (Wildermuth & McNamara 1994). Furthermore, in this study plants were scored at flowering in the field as compared to maturity in the glasshouse. CR discrimination ability has previously been reported to be higher at flowering than at maturity, as infected phloem cells began to disintegrate

and may fade by maturity (Malligan 2009; Knight & Sutherland 2016). Further increases in phenotyping accuracy may be seen if assessor scores used for training are obtained at the same time as the NIR reading.

Results also indicate potential for the further development of a phenotyping system that can be used in controlled environments with average accuracies as high as 80.83% in the glasshouse in this study. The significantly higher number of lines that can be screened in a field environment should be considered. Further, field trials are required for tolerance screening while resistance trials can be conducted in both glasshouse and field environments (Forknall et al. 2019). High accuracy can be achieved in both the glasshouse and field. If early quantification is acceptable and the seed does not need to be retained, seedling assays in the glasshouse may be desirable as glasshouse trials have the advantages of temperature and moisture control to reduce variation while also allowing multiple experiments to be conducted in a single season (Purss 1969; Dodman & Wildermuth 1987; Wildermuth & McNamara 1994; Wallwork et al. 2004; Mitter et al. 2006). A quantification framework discriminating disease into three categories (not-detectible, low and high), with high accuracy, as shown here, would be beneficial in breeding pipelines by allowing early discrimination of lines with high severity of CR. These lines could be removed from the pipeline earlier than possible with traditional methods, saving resources for the development of more successful germplasm.

Further, the results of this study indicate the potential for the use of growth stage targeting to improve precision of CR discrimination by specifically targeting early-tillering, late-tillering and booting. This finding is consistent with previous findings in Chapter 5 which indicated significant differences existed between sensing timepoints for CR detection. This timepoint sensitivity has been previously documented in multiple cropping systems (Busemeyer et al. 2013; Bauriegel & Herppich 2014; Anderegg et al. 2019).

The success of this study shows the potential for the development of a high-throughput phenotyping (HTP) methodology for the quantification of CR resistance and tolerance in both glasshouse and field environments. Successful development of an HTP system could be transferred into an automated robotic or tractor-based system for the detection and quantification of CR, such as that previously developed for use in yellow rust detection in wheat (Moshou et al. 2004; Moshou et al. 2005; Moshou et al. 2011).

Further work will begin to evaluate the potential of camera-based systems to discriminate CR, using the approaches developed with contact-based sensors, potentially enabling equipment, robotic and aerial-based phenotyping frameworks.

V.   CONCLUSION

In this study, we evaluated the potential for PCA-MLP or DWT-MLP based quantification of *F. pseudograminearum* induced CR in bread wheat. Quantification was based upon near-infrared spectroscopy across six developed scoring systems consisting of differing numbers of scoring categories, with separation into two to eleven groups. Early detection of CR has been reported using this NIR point based sensor (Chapter 3, 4). However, successful quantification of CR in both field and glasshouse environments would enable the development of an HTP methodology for CR breeding in Australia and around the world.

Our results indicate that CR discrimination is possible across both field and glasshouse environments and in combined datasets, using a contact NIR sensor for data collection and processing with either PCA-MLP or DWT-MLP machine learning approaches. Further, our results indicate discrimination capability can be further improved in combined data by targeting timepoints corresponding to the specific growth stages of early-tillering, late-tillering and booting, consistent with findings in Chapter 5. The results indicate that while discrimination can be obtained using up to eleven separate levels of infection classification, the highest accuracy can be obtained by using two to three categories. Further, these results indicate that field trial data may be more successful in future model development, possibly due to the increased variation these environments produce.

Future work should investigate whether these methodologies can be applied to camera-based systems to allow for varied data-collection methodologies (e.g. robotic, UAV, human). Further, future work should also begin to validate these findings across a wider array of germplasm and cereals.

REFERENCES

Anderegg, J, Hund, A, Karisto, P & Mikaberidze, A 2019, 'In-field detection and quantification of Septoria tritici blotch in diverse wheat germplasm using spectral-temporal features', *BioRxiv*, p. 664011.

Bauriegel, E & Herppich, WB 2014, 'Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to Fusarium spec. infections on wheat', *Agriculture*, vol. 4, no. 1, pp. 32-57.

Bruce, LM & Li, J 2001, 'Wavelets for computationally efficient hyperspectral derivative analysis', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1540-6.

Busemeyer, L, Ruckelshausen, A, Möller, K, Melchinger, AE, Alheit, KV, Maurer, HP, Hahn, V, Weissmann, EA, Reif, JC & Würschum, T 2013, 'Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation', *Scientific reports*, vol. 3, p. 2442.

Daubechies, I 1988, 'Orthonormal bases of compactly supported wavelets', *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909-96.

Dodman, R & Wildermuth, G 1987, 'Inoculation methods for assessing resistance in wheat to crown rot caused by Fusarium graminearum Group 1', *Australian Journal of Agricultural Research*, vol. 38, no. 3, pp. 473-86.

Forknall, CR, Simpfendorfer, S & Kelly, AM 2019, 'Using yield response curves to measure variation in the tolerance and resistance of wheat cultivars to Fusarium crown rot', *Phytopathology*, vol. 109, no. 6, pp. 932-41.

Hunter, JD 2007, 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-5.

Jones, E, Oliphant, T & Peterson, P 2001, 'SciPy: Open Source Scientific Tools for Python'.

Knight, NL & Sutherland, MW 2016, 'Histopathological assessment of Fusarium pseudograminearum colonization of cereal culms during crown rot infections', *Plant disease*, vol. 100, no. 2, pp. 252-9.

Labsphere, 'Reflectance characteristics of Spectralon panels', Reflectance Calibration Laboratory, Sutton, NH, USA. (1998).

Lee, GR, Gommers, R, Waselewski, F, Wohlfahrt, K & O'Leary, A 2019, 'PyWavelets: A Python package for wavelet analysis', *Journal of Open Source Software*, vol. 4, no. 36, p. 1237.

Malligan, CD 2009, 'Crown rot (Fusarium pseudograminearum) symptom development and pathogen spread in wheat genotypes with varying disease resistance', University of Southern Queensland.

Mitter, V, Zhang, M, Liu, C, Ghosh, R, Ghosh, M & Chakraborty, S 2006, 'A high-throughput glasshouse bioassay to detect crown rot resistance in wheat germplasm', *Plant pathology*, vol. 55, no. 3, pp. 433-41.

Moshou, D, Bravo, C, West, J, Wahlen, S, McCartney, A & Ramon, H 2004, 'Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks', *Computers and electronics in agriculture*, vol. 44, no. 3, pp. 173-88.

Moshou, D, Bravo, C, Oberti, R, West, J, Bodria, L, McCartney, A & Ramon, H 2005, 'Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using Kohonen maps', *Real-Time Imaging*, vol. 11, no. 2, pp. 75-83.

Moshou, D, Bravo, C, Oberti, R, West, J, Ramon, H, Vougioukas, S & Bochtis, D 2011, 'Intelligent multi-sensor system for the detection and treatment of fungal diseases in arable crops', *Biosystems Engineering*, vol. 108, no. 4, pp. 311-21.

Newton, A & Hackett, C 1994, 'Subjective components of mildew assessment on spring barley', *European Journal of Plant Pathology*, vol. 100, no. 6, pp. 395-412.

Nutter, J 1997, 'Disease severity assessment training', *Exercises in plant disease epidemiology*, pp. 1-7.

Nutter Jr, F, Gleason, M, Jenco, J & Christians, N 1993, 'Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems', *Phytopathology*, vol. 83, no. 8, pp. 806-12.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R & Dubourg, V 2011, 'Scikit-learn: Machine learning in Python', *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-30.

Percy, C, Wildermuth, G & Sutherland, M 2012, 'Symptom development proceeds at different rates in susceptible and partially resistant cereal seedlings infected with Fusarium pseudograminearum', *Australasian Plant Pathology*, vol. 41, no. 6, pp. 621-31.

Purss, G 1969, 'Studies of varietal resistance to crown rot of wheat caused by Fusarium graminearum Schw', *Queensland Journal of Agriculture and Animal Science*, vol. 23, pp. 475-98.

Python version 3.6.8, Python Software Foundation 2019.

Savitzky, A & Golay, MJ 1964, 'Smoothing and differentiation of data by simplified least squares procedures', *Analytical chemistry*, vol. 36, no. 8, pp. 1627-39.

Simpfendorfer, S, McKay, A & Ophel-Keller, K 2019, 'New approaches to crop disease management in conservation agriculture. In (Eds J Pratley and J

Kirkegaard) "Australian Agriculture in 2020: From Conservation to Automation" pp 173-188 (Agronomy Australia and Charles Sturt University: Wagga Wagga)', pp. 173-88.

Wallwork, H, Butt, M, Cheong, J & Williams, K 2004, 'Resistance to crown rot in wheat identified through an improved method for screening adult plants', *Australasian Plant Pathology*, vol. 33, no. 1, pp. 1-7.

Waskom, M, Botvinnik, O, Hobson, P, Cole, JB, Halchenko, Y, Hoyer, S, Miles, A, Augspurger, T, Yarkoni, T, Megies, T, Coelho, LP, Wehner, D, cynddl, Ziegler, E, diego0020, Zaytsev, YV, Hoppe, T, Seabold, S, Cloud, P, Koskinen, M, Meyer, K, Qalieh, A & Allan, D 2014, seaborn: v0.5.0 (November 2014), 0.5.0, <https://doi.org/10.5281/zenodo.12710>.

Wasserman, PD & Schwartz, T 1988, 'Neural networks. II. What are they and why is everybody so interested in them now?', *IEEE Expert*, vol. 3, no. 1, pp. 10-5.

Wildermuth, G & McNamara, R 1994, 'Testing wheat seedlings for resistance to crown rot caused by Fusarium graminearum Group 1', *Plant disease* (USA).

Zadoks, JC, Chang, TT & Konzak, CF 1974, 'A decimal code for the growth stages of cereals', *Weed Research*, vol. 14, no. 6, pp. 415-21.

CHAPTER 7

CNNs FOR CROWN ROT DISCRIMINATION IN WHEAT, PART A: DETECTION

In this study, one single input and two multi-input convolutional neural networks (CNN) were developed for the discrimination of *Fusarium pseudograminearum* induced crown rot of wheat. Five wavebands of importance for crown rot detection were identified in the near-infrared (NIR) range from analysis in chapters 3 and 4. The wavebands were each selected to correlate to a narrow-bandpass filter with 50 nm widths for use with a full spectrum continuous NIR camera system (900–1700 nm). CNN models were developed incorporating single, combinations of two and all five wavebands. The resulting systems were tested in a glasshouse trial in 2019 across four weeks of early crown rot infection, from three weeks post-inoculation (Zadoks (Z) 13, 21) (Zadoks et al. 1974). Models were shown to achieve average accuracies across 10 random independent validation splits of 53–100%, with a dual-band model centered at 950 and 1350 nm performing best on average across data groupings. Resulting models have been successfully ported to a development board for real-time detection applications.

Humpal J., McCarthy C., Percy C., & Thomasson AJ. (2020e). CNNs for crown rot discrimination in wheat, Part A: detection. This chapter was prepared according to the instructions to authors given by **Computers and Electronics in Agriculture**.

CNNs for crown rot discrimination in wheat, Part A: detection

Jacob Humpal[*], Cheryl McCarthy[*], Cassandra Percy[**], J. Alex Thomasson[***]

[*] Centre for Agricultural Engineering, University of Southern Queensland, Toowoomba, Queensland, Australia

[**] Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia

[***] Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, Mississippi, United States

J. Humpal, C. McCarthy (*)

Centre for Agricultural Engineering, University of Southern Queensland,

Toowoomba, Qld 4350, Australia

*e-mail: cheryl.mccarthy@usq.edu.au

Ph: +61 7 4631 2297

Abstract- In this paper, Part A of a pair of papers around neural network development for crown rot (CR) discrimination, one single input and two multi-input convolutional neural networks (CNNs) were developed for the discrimination of CR in wheat induced by the *Fusarium pseudograminearum* pathogen. Five wavebands of importance for CR detection were identified from previous research in the near-infrared (NIR) range. Each waveband was selected to correlate to a narrow-bandpass filter with 50 nm widths for use with a full spectrum NIR camera system (900–1700 nm). Wavebands were used as inputs into the developed CNNs consisting of single, two input and five input models and were trained and tested on four weeks of data from a glasshouse trial in 2019, from three weeks post-inoculation (Zadoks (Z) 13, 21). This is early in the CR lifecycle before symptoms become apparent. Models were shown to achieve average accuracies across 10 independent random validation splits of 53–100%, with a dual band model (950, 1350 nm) performing best, with an average accuracy across datasets of 87%. Further, models were successfully deployed on a development board, enabling real-time detection applications.

Keywords*- crop disease, fusarium pseudograminearum, CNN, remote sensing, NIR*

I. INTRODUCTION

CR is a cereal stubble-borne disease that is estimated to cost the Australian wheat sector $79 million per year in lost profits (Murray & Brennan 2009). Recent reports have indicated that these losses have increased along with the distribution of CR, across the Australian grain belt from reports from 2014–2017 (Simpfendorfer et al. 2019). Incidence of paddocks with high levels of infected stubble were reported at 31% in the Northern region, 21% in the Western region and 15% in the Southern region.

The fungal pathogen predominantly responsible for cereal CR in Australia is *Fusarium pseudograminearum.* CR infection begins when infected grass or cereal debris comes into contact with the roots, stem, sub-crown internode or crown of the host crop (Burgess et al. 1993). It is hypothesised that the recent increase in the adoption of conservation agriculture has allowed infectious material to build up in the soil, increasing CR inoculum levels throughout Australia (Simpfendorfer et al. 2019).

Current CR disease identification is time-consuming and labour intensive as it relies on sampling plants throughout the paddock. As CR is difficult to identify early in the season without sufficient sampling, emerging areas of disease may be overlooked. This can cause delays in production decision-making around input selection and timing (e.g. nitrogen and water). Delays in identification can potentially lead to expending resources in areas of the crop where productivity will already be low, while early identification could provide guidance around obtaining maximum potential yield under disease. Further, a machine-vision-based analysis tool could potentially allow growers to identify areas of disease which otherwise would be mistaken as damage from other factors such as frost. A camera-based machine-vision system would allow growers to identify in paddock CR infections rapidly and reliably.

Multispectral imaging takes readings of light in multiple spaced bands at positions across the electromagnetic spectrum. Unique reflectance signatures have been reported in different agricultural disease systems (Steddom et al. 2005; Franke & Menz 2007; Polder et al. 2014). However, the wide-spread use of broad spectral vegetation indices (SVIs) like NDVI when attempting to discriminate plant stressors may reduce output reliability as these indices are seldom stress specific.

Previous chapters have reported the successful detection of CR in both the field and glasshouse using a near-infrared contact sensor (900–1700 nm) and machine learning methodologies (Chapter 3, 4, 5). This study aims to build upon the previous research to develop a camera-based system for the detection of CR. A camera-based system would allow for more rapid identification of the disease by removing the requirement to come into contact with the plant tissue, potentially enabling airborne disease identification systems. Convolutional neural networks (CNNs) were developed to include image data from one, two or five feature wavebands identified as important for CR detection in previous chapters. A CNN is a class of deep neural networks, widely used for image analysis and classification, which learns patterns internally (LeCun et al. 1998).

Disease specific indices which allow early detection of CR can help growers make in-crop management decisions to maximise profitability in the presence of CR infection both by optimising input timing and restricting input application. A machine vision-based system for the discrimination of early-season CR would also be a step forward in the development of a high-throughput phenotyping (HTP) methodology for rapid germplasm screening. Current breeding programs grow plants through harvest in order to screen for tolerance and resistance. Early detection capability would allow for the most effective germplasm to be moved through the program, reducing breeding costs and retaining additional capital for line development.

II. MATERIALS AND METHODS

A. Experimental design and plant cultivation

An experimental glasshouse trial was conducted in 2019 at the University of Southern Queensland's Centre for Crop Health (CCH), Darling Heights, QLD. Six replicates of plus and minus inoculated pots of five standard genotypes, with known resistances to CR were grown in a glasshouse with temperature controlled at 20–25° C and were watered once per week to field capacity. *F. pseudograminearum* colonised wheat grain inoculum (Percy et al. 2012) was applied to individual plant coleoptiles of the inoculated treatments at the two-leaf stage. Plants were harvested and rated for disease severity at physiological maturity. The glasshouse trial was statistically designed in cooperation with biostatisticians at Queensland Department of Agriculture and Fisheries (QDAF).

B. Near-infrared camera measurements

The FLIR Tau™ SWIR camera core (900–1700 nm) was used to record near-infrared video on an attached mini-DVR (digital video recorder), of inoculated and non-inoculated plants once a week for four weeks, from three weeks post-inoculation. Videos were recorded for the purpose of still image acquisition. Separate videos were recorded using five narrow-bandpass filters, centered at 950, 1100, 1300, 1350 and 1600 nm with 50 nm bandwidths. These bands were selected based upon past research (Chapter 3, 4). Each filter was installed on the camera one at a time and imaging was completed for sixty plants, before the filter was changed and imaging repeated for all plants for each bandpass filter. Imaging took less than one minute for each bandpass filter and capturing all plants with all filters took less than one hour. Variations in leaf position seen in the images are due to slight variations in time of image acquisition and due to the relative movement of the camera to the plant while filters were changed and while individual plants were moved and removed from view of the camera. Measurements were taken from a height of one metre above the pot for both inoculated and non-inoculated pots. Pots were placed in a tray with 2 inch calibrated 2, 5, 10, 20, 40, 60 and 80% grey and a 99% white reference Spectralon® panels, visible to the camera system (Fig. 1).
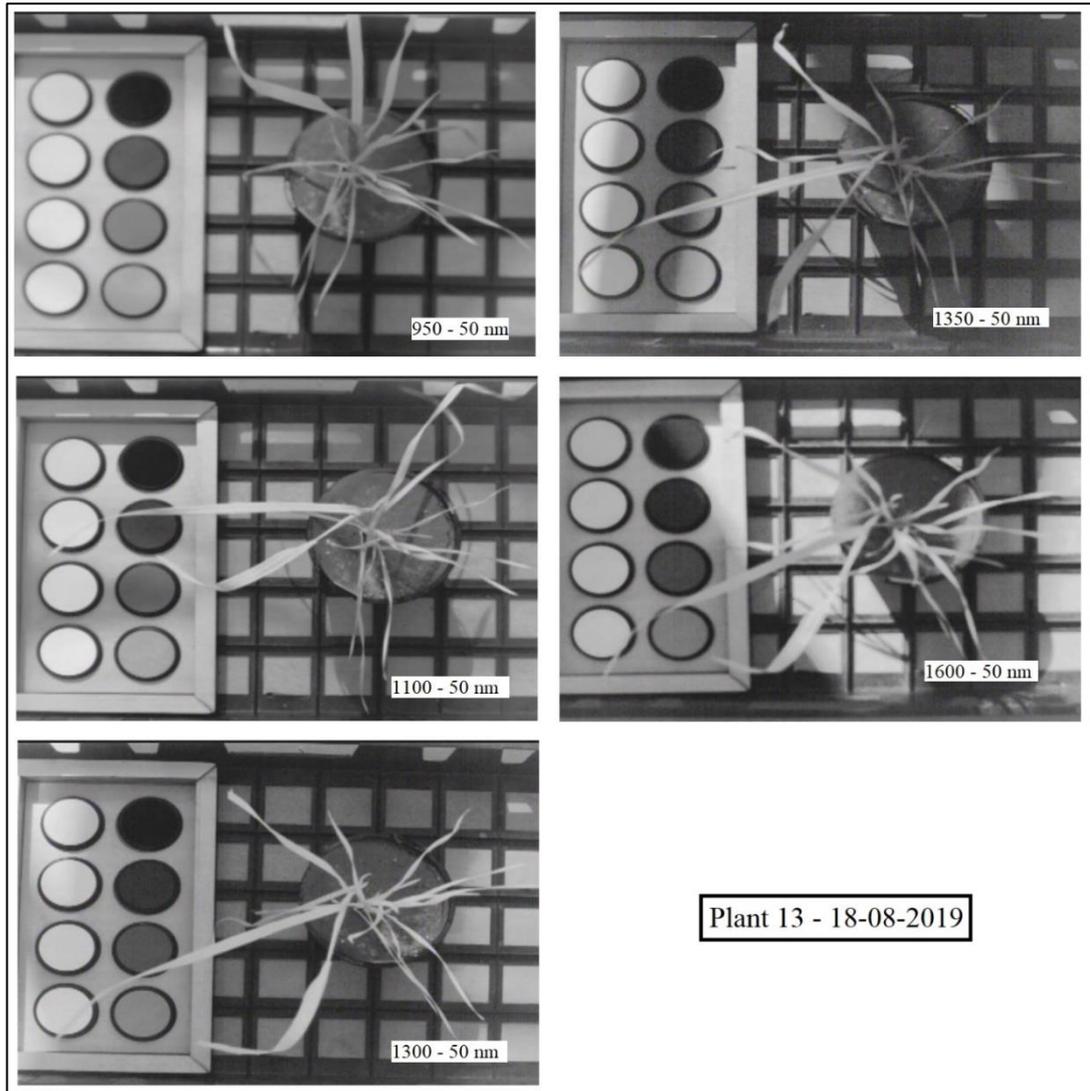
Fig. 1: *F. pseudograminearum* inoculated wheat plant (Plant 13) and reference panels, still images from video feed for each of the five narrow-bandpass filters, taken in the USQ-CCH glasshouse, Darling Heights QLD on 18-08-2019. Variations in leaf position are due to slight variations in time of image acquisition and due to the relative movement of the camera to the plant while filters were changed and while individual plants were moved and removed from view of the camera.

C. Data analysis

Data analysis was performed inside the Python computing environment (Python version 3.6.8, Python Software Foundation 2019). Data analysis was conducted using open-source libraries including Keras (Chollet 2015) a neural network library API, on top of TensorFlow, a machine learning platform (TensorFlow version 2.0.0-alpha0; Abadi et al. 2016). Additionally, the SciPy ecosystem (Jones et al. 2001) and the Scikit-learn library (Pedregosa et al. 2011) were used in model development.

C.1 CNN Development

Near-infrared videos were collected weekly for each plant and each of the five narrow-bandpass filters for four weeks, from three weeks post-inoculation. Individual video frames were removed for each plant from each of the five wavelength videos. The resulting sixty image datasets were labeled for use as sets of one, two or five inputs into a convolutional neural network (CNN). Models were developed and tested on data from individual weeks (60 images/input) and combined data across the four weeks (240 images/input). Input images were reshaped to 100x100 pixels, converted to grayscale and scaled to 0–1. Each dataset was split into training-test and validation datasets using a 75-25% split and applied as inputs into the CNN model. The CNN model was developed using the Keras library on top of TensorFlow (Chollet 2015; Abadi et al. 2016). The resulting models used a stochastic gradient descent (SGD) optimiser and consisted of a concatenation layer in the two and five input models to combine data from individual bands, a flatten layer to convert data to a simple vector output, a dense layer of 128 nodes, a dense layer of 64 nodes, a 40% dropout layer to reduce overfitting and a final sigmoid dense layer for binary output (Fig. 2). The resulting CNN was validated and scored for accuracy based upon the average accuracy of 10 independent random validation splits per dataset. The equation used for accuracy is reported below (1):

$$accuracy\left(y, \hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \cdot 1\left(\hat{y}_i = y_i\right) \qquad (1)$$

Where $\hat{y}_i$ is the predicted value of the ith sample and y_i is the corresponding true value
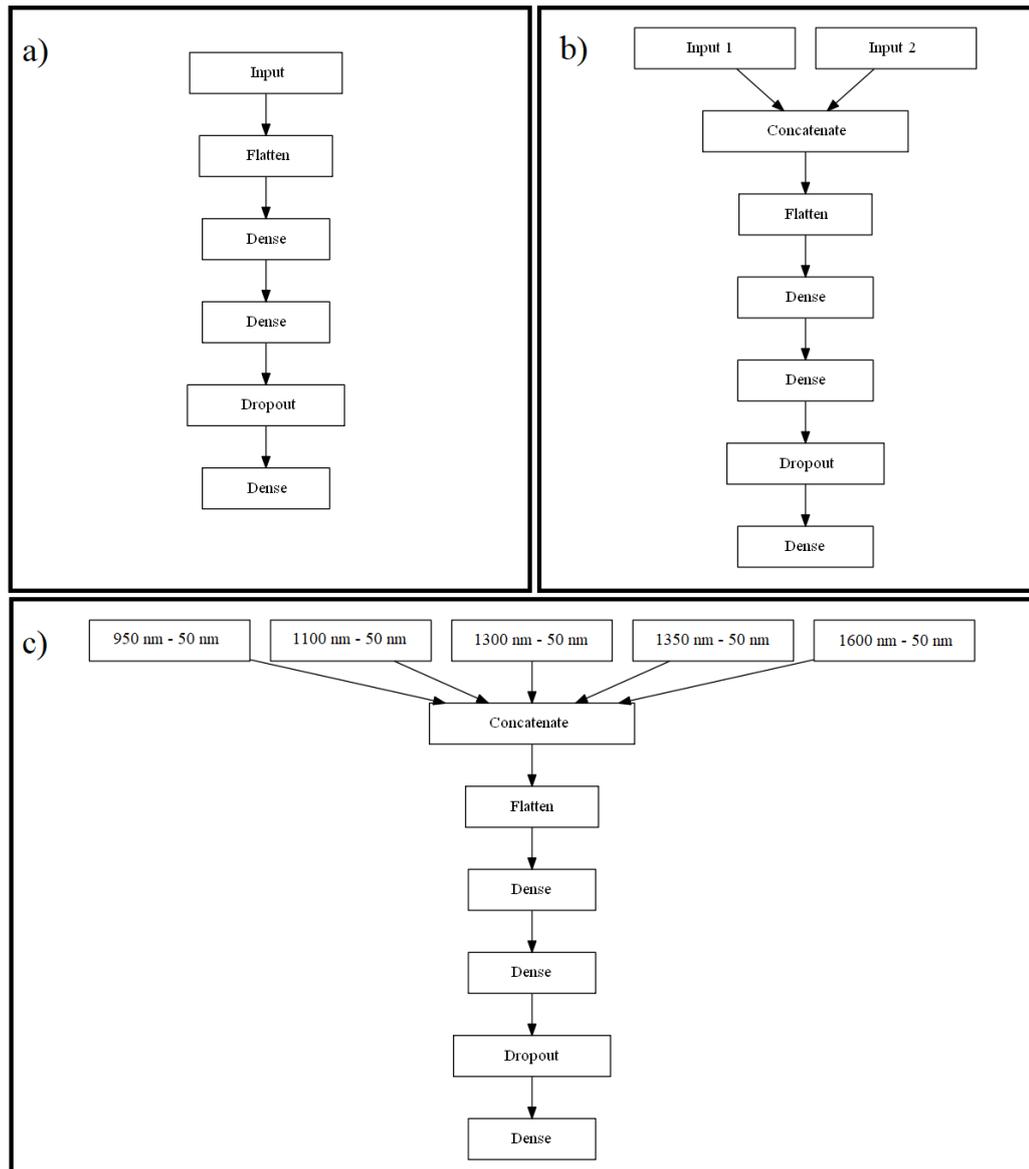
Fig. 2: Developed CNN architectures, consisting of 1 (a), 2 (b) or 5 (c) input bands.

C.2 Real-time mobile detection

The model framework was successfully deployed on a Jetson Nano™ Developer Kit (NVIDIA Corporation, USA) as the first step towards a proof-of-concept real-time robotic detection system. The Jetson Nano™ has maximum dimensions of 105 x 85 x 40 mm and contains a NVIDIA Maxwell GPU, quad-core ARM Cortex-A57 processor and 4GB LPDDR4 memory. The board allows direct camera connection through two MIPI-CSI camera connectors, as well as, four USB 3.0 ports. Further, data can be visualised in real-time using an HDMI or DisplayPort output or saved onboard for later retrieval. Models were successfully developed, trained and validated on the board. A speed performance test was conducted between a high-end laptop and the Jetson

Nano™, to compare the time it took to process the images and score them per plant for the five-input model. The laptop contained an Intel® Core™ i7-8550U CPU @ 1.80GHz 1.99 GHz and 16.0 GB installed RAM. The Jetson Nano™ processed and scored one plant every ~0.1693 s, while the laptop scored one plant every ~0.0776 s.

III.   RESULTS

A. Discrimination of disease: weekly models

Weekly datasets were collected 3, 4, 5 and 6 weeks after inoculation (weeks 1–4 after Zadoks 13, 21). Average accuracies across models for each weekly dataset ranged from 65 to 90%, with five weeks after inoculation returning the highest accuracies (week 3). This accuracy ranged from 74 to 100% (Fig. 3).   Three weeks post-inoculation (week 1) returned the lowest accuracies, with a range of 53–78%. Standard deviation decreased while overall accuracy increased from weeks 2 to 3, with accuracy slightly decreasing again in week four.

B. Discrimination of disease: combined models

When data from all four weeks was combined, overall average accuracies of 72% were observed with a standard deviation of 8% (Fig. 3). Combined data returned the largest average accuracy range (61–87%), the lowest overall average accuracy and highest standard deviation behind week 1.
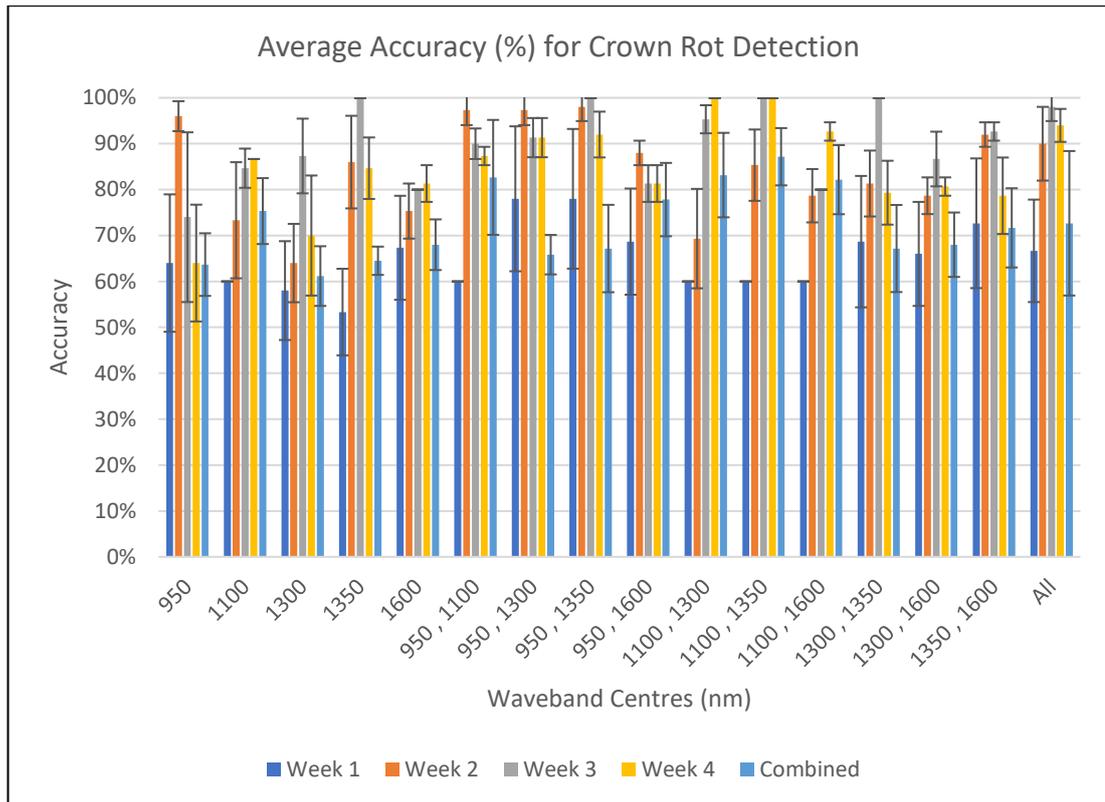
Fig. 3: Average mean accuracies and standard deviations for each dataset (weekly and combined) for each set of wavebands used in model development. The 'Combined' grouping includes all data from weeks 1, 2, 3 and 4. The 'All' grouping includes all five wavebands (950, 1100, 1300, 1350 and 1600 nm). Weeks are reported as weeks from Zadoks (Z) 13, 21 (Zadoks et al. 1974).

C. Discrimination of disease: single waveband models

Single waveband models returned average accuracies of 68–78%, dependent on the waveband used. The lowest average accuracy was reported using a model trained using the 1300 nm waveband data. The highest average accuracy was achieved using a waveband centered at 1350 nm. The 1350 nm single band model returned an accuracy of 100% on the week 3 dataset. In the single band models the highest accuracies were again observed in week 3, with an average of 85% across models and the lowest accuracies were again observed in week 1, with an average of 61% across models (Fig. 3).

D. Discrimination of disease: dual waveband models

Models developed using two wavebands (dual waveband models) returned average accuracies of 76–87%, dependent on waveband combination. The lowest average accuracy was reported using a combination of bands centered at 1300 and 1600 nm. The highest average accuracy was achieved using a model developed from wavebands centered at 950 and 1350 nm. The 950-1350 nm dual band model returned an accuracy of 98% on the week 2 and 100% on the week 3 datasets. The highest accuracies in the dual waveband models were again observed in week 3, with an average of 92% across models and the lowest accuracies were observed in week 1, with an average of 67% across models (Fig. 3).

E. Discrimination of disease: five waveband model

A model developed using combined data from all five wavebands returned an average accuracy of 84% across datasets. The lowest average accuracy of 67% was reported in week 1, while the highest average accuracy of 98% was reported in week 3 (Fig. 3).

IV. DISCUSSION

In this study multi-input CNNs were developed and tested for the ability to successfully discriminate *Fusarium pseudograminearum* induced CR of wheat from single, dual or sets of five near-infrared image wavebands, in a glasshouse, for four weeks from Z 13, 21. A combination of two 50 nm wavebands, with centres at 950 and 1350 nm exhibited strong ability to discriminate between *F. pseudograminearum* inoculated and non-inoculated wheat plants, with an average accuracy across datasets of 87%. The 1350 nm centered waveband appeared both as the top performing single waveband model and in the top performing dual waveband model, as such, variations in this waveband's intensity are likely to be important in the further development of a specific spectral marker of CR infection. However, multi-waveband models are preferable to single waveband models as single waveband models must rely on relationships between the waveband and references, requiring reference material in each image. A multi-waveband model allows internal relationships between bands to discriminate crop disease (Mahlein et al. 2010; Zheng et al. 2019). The results of this study indicate successful development of a proof-of-concept camera-based CR discrimination framework, which can rapidly discriminate CR under glasshouse conditions. Further, resulting detection accuracies were higher than previously

reported using a contact near-infrared sensor (Chapter 3, 4) from which the initial wavebands were chosen. Previous research has indicated that combining scan locations, throughout the canopy, can increase detection capability using a contact sensor (Chapter 5). As the camera-based system is able to simultaneously image multiple leaves on both the tillers and main stem, this could potentially contribute to the higher accuracies seen with this system. Finally, results indicate the successful development of models which can be transferred to a mobile format with potential for real-time disease detection. A camera-based detection methodology is a step towards developing real-time solutions for grower disease detection and mapping in the paddock in addition to phenotyping technologies to provide tools for breeding programs.

Our results indicate that week three from Z 13, 21, or five weeks post-inoculation is the best time to use this system for the detection of CR with an average of 90% accuracy over 10 random independent validation splits across all developed models (Fig. 3). However, accuracies greater than 80% are seen in weeks two through four (84–90%) indicating that early detection is possible within this window around early tillering. Further, models validated on combined data returned average accuracies of 72% across all models. However, combined models also exhibited larger standard deviation than models developed on weeks 2–4 which may be due to differences in disease progression appearing in the near-infrared range or simply to the larger dataset included.

These results indicate that reliable detection can be accomplished from four to six weeks after inoculation, with five weeks being the ideal target. The importance of such spectral-temporal features has been reported in previous research, both in CR and other crop disease systems (Busemeyer et al. 2013; Anderegg et al. 2019).

The discrimination success shown in this study indicates potential for the development of real-time CR detection technologies utilising near-infrared cameras, which could be developed into a multispectral sensing array. Further work should validate these results on a larger, diverse set of germplasm and the potential of this system to discriminate CR in the field should be evaluated. Further work should also focus on determining whether accuracy and sensitivity can be improved by changing the number of filters required for successful detection. CR camera-based detection technology would facilitate the early detection of disease, enabling growers to make informed decisions

regarding CR control strategies and enabling researchers to track infection cycles, increasing our understanding of this disease and its underlying pathogens. Further, the quantification potential of this technology should be evaluated, both in glasshouse and field environments. A successful quantification model using this approach would enable rapid high-throughput phenotyping (HTP) of CR, enabling breeding companies to deliver germplasm to the market more efficiently. This success also potentially enables the development of robotic and aerial-based, real-time phenotyping systems for soil and stubble-borne crop diseases, such as those previously developed for foliar yellow rust detection in wheat (Moshou et al. 2011).

V. CONCLUSION

In this study, we evaluated the potential for single and multi-input CNN classifiers based on single, dual or five near-infrared bands obtained using a filtered NIR camera to successfully discriminate *F. pseudograminearum* induced CR in wheat in a glasshouse environment. The results of this study indicate that CR discrimination is possible in the glasshouse both using weekly and multi-weekly (combined) datasets, using a multi-input CNN. These results indicate that while discrimination was successful across all databases, images for discrimination should be obtained between two and four weeks from Z 13, 21, with three weeks being the ideal timepoint.

Part B of this pair of papers will determine whether quantification models can be developed with success, using this approach. Future work should investigate whether this approach can be applied to field/natural systems and evaluate the potential to integrate this system into a platform for real-time detection, for deployment on robotic or UAV-based platforms. This research also requires validation of this model across a larger set of germplasm.

ACKNOWLEDGEMENT

REFERENCES

Abadi, M, Agarwal, A, Barham, P, Brevdo, E, Chen, Z, Citro, C, Corrado, GS, Davis, A, Dean, J & Devin, M 2016, 'Tensorflow: Large-scale machine learning on heterogeneous distributed systems', *arXiv preprint* arXiv:1603.04467.

Anderegg, J, Hund, A, Karisto, P & Mikaberidze, A 2019, 'In-field detection and quantification of Septoria tritici blotch in diverse wheat germplasm using spectral-temporal features', *BioRxiv*, p. 664011.

Burgess, L, Backhouse, D, Summerell, B, Pattison, A, Klein, T, Esdaile, R & Ticehurst, G 1993, 'Long-term effects of stubble management on the incidence of infection of wheat by Fusarium graminearum Schw. Group 1', *Australian Journal of Experimental Agriculture*, vol. 33, no. 4, pp. 451-6.

Busemeyer, L, Ruckelshausen, A, Möller, K, Melchinger, AE, Alheit, KV, Maurer, HP, Hahn, V, Weissmann, EA, Reif, JC & Würschum, T 2013, 'Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation', *Scientific reports*, vol. 3, p. 2442.

Chollet, F 2015, Keras, https://<keras.io>.

Franke, J & Menz, G 2007, 'Multi-temporal wheat disease detection by multi-spectral remote sensing', *Precision agriculture*, vol. 8, no. 3, pp. 161-72.

Jones, E, Oliphant, T & Peterson, P 2001, 'SciPy: Open Source Scientific Tools for Python'.

LeCun, Y, Bottou, L, Bengio, Y & Haffner, P 1998, 'Gradient-based learning applied to document recognition', Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-324.

Mahlein, A-K, Steiner, U, Dehne, H-W & Oerke, E-C 2010, 'Spectral signatures of sugar beet leaves for the detection and differentiation of diseases', *Precision agriculture*, vol. 11, no. 4, pp. 413-31.

Moshou, D, Bravo, C, Oberti, R, West, J, Ramon, H, Vougioukas, S & Bochtis, D 2011, 'Intelligent multi-sensor system for the detection and treatment of fungal diseases in arable crops', *Biosystems Engineering*, vol. 108, no. 4, pp. 311-21.

Murray, GM & Brennan, JP 2009, 'Estimating disease losses to the Australian wheat industry', *Australasian Plant Pathology*, vol. 38, no. 6, pp. 558-70.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R & Dubourg, V 2011, 'Scikit-learn: Machine learning in Python', *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-30.

Percy, C, Wildermuth, G & Sutherland, M 2012, 'Symptom development proceeds at different rates in susceptible and partially resistant cereal seedlings infected with Fusarium pseudograminearum', *Australasian Plant Pathology*, vol. 41, no. 6, pp. 621-31.

Polder, G, van der Heijden, GW, van Doorn, J & Baltissen, TA 2014, 'Automatic detection of tulip breaking virus (TBV) in tulip fields using machine vision', *Biosystems Engineering*, vol. 117, pp. 35-42.

Python version 3.6.8, Python Software Foundation 2019.

Simpfendorfer, S, McKay, A & Ophel-Keller, K 2019, 'New approaches to crop disease management in conservation agriculture. In (Eds J Pratley and J Kirkegaard) "Australian Agriculture in 2020: From Conservation to Automation" pp 173-188 (Agronomy Australia and Charles Sturt University: Wagga Wagga)', pp. 173-88.

Steddom, K, Bredehoeft, M, Khan, M & Rush, C 2005, 'Comparison of visual and multispectral radiometric disease evaluations of Cercospora leaf spot of sugar beet', *Plant Disease*, vol. 89, no. 2, pp. 153-8.

Zadoks, JC, Chang, TT & Konzak, CF 1974, 'A decimal code for the growth stages of cereals', *Weed Research*, vol. 14, no. 6, pp. 415-21.

Zheng, Q, Huang, W, Cui, X, Dong, Y, Shi, Y, Ma, H & Liu, L 2019, 'Identification of Wheat Yellow Rust Using Optimal Three-Band Spectral Indices in Different Growth Stages', *Sensors*, vol. 19, no. 1, p. 35.

CHAPTER 8

CNNs FOR CROWN ROT DISCRIMINATION IN WHEAT, PART B: QUANTIFICATION

In this study, an extension of Chapter 7, three convolutional neural networks (CNNs) were developed for the quantification of *Fusarium pseudograminearum* induced crown rot of bread wheat (*Triticum aestivum*). Five wavebands of importance for crown rot detection were identified in the near-infrared (NIR) range from analysis in chapters 3 and 4 and utilised for crown rot discrimination from image datasets in Chapter 7. Each waveband correlates to a narrow-bandpass filter with 50 nm widths for use with a continuous NIR camera (900–1700 nm) system. Single waveband, dual waveband and five waveband quantification models were built upon three scales developed in Chapter 6. Quantification models sorted plants into either two or three classes dependent on the scale used. The quantification models were developed and tested on glasshouse trial data from four weeks of early crown rot infection in 2019. Models achieved average accuracies across 10 independent random validation splits of 21–73%, dependent upon rating system, dataset and wavebands used. A binary rating system performed best with combined data returning significantly higher accuracies across all models than individual weeks.

CNNs for crown rot discrimination in wheat, Part B: quantification
Jacob Humpal[*], Cassandra Percy[**], Cheryl McCarthy[*], J. Alex Thomasson[***]

[*] Centre for Agricultural Engineering, University of Southern Queensland, Toowoomba, Queensland, Australia

[**] Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia

[***] Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, Mississippi, United States

J. Humpal, C. Percy (*)

Centre for Agricultural Engineering, University of Southern Queensland,

Toowoomba, Qld 4350, Australia

*e-mail: cassy.percy@usq.edu.au

Ph: +61 7 4631 1241

Abstract- In this paper, Part B of a pair of papers on neural network development for crown rot (CR) discrimination, a single and two multi-input convolutional neural networks (CNNs) were developed for the quantification of CR in bread wheat (*Triticum aestivum*) induced by *Fusarium pseudograminearum* using three developed disease scales based on percentage stem browning. Scales included two or three output classes determined from previous research. Near-infrared (900–1700 nm) narrow bandpass filters correlated to five wavebands of importance for CR detection were identified from previous research for use in developing a phenotyping model. All filters had 50 nm bandwidths. Resulting filtered images were used as inputs into the developed CNNs and trained and validated on four weeks of glasshouse trial data in 2019. Developed models were shown to achieve average accuracies across 10 independent random validation splits of 21–73%, with combined data from all four weeks performing best, specifically when discriminating between less than 50% stem browning and greater than 50% stem browning.

Keywords- *phenotyping, Fusarium pseudograminearum, CNN, remote sensing, NIR*

I. INTRODUCTION

Wheat CR, caused by the fungus *Fusarium pseudograminearum* is a cereal stubble-borne disease that can be responsible for a greater than 50% yield reduction under conducive conditions (Klein et al. 1991). This disease is estimated to cost the Australian wheat sector $79 million per year with recent reports indicating a further increase in losses and an increased distribution of the disease (Murray & Brennan 2009; Simpfendorfer et al. 2019).

CR is generally identified by the presence of honey brown coloured lesions along the stem and the appearance of dead grain heads, known as whiteheads after flowering (Klein et al. 1991; Malligan 2009; Knight et al. 2017; Knight et al. 2020). Current identification methods cause CR assessment to be time consuming and labour intensive, as plants are usually not scored until after flowering and rating often involves the removal and storage of material, off-site for manual scoring of stem browning to determine resistance and tolerance. In traditional CR phenotyping a trained assessor removes the leaf sheaths from the lower internodes of the plant and scores disease severity based on the percentage of honey brown lesioning on the stem. Potential exists for improvements in phenotyping technologies to increase breeding capacity and reduce associated costs. Specifically, the development of an image-based high-throughput phenotyping system (HTP) for CR quantification early in plant development would enable breeders to deliver germplasm with improved resistance and tolerance to growers in less time. Such a system could be designed to work with handheld, robotic or unmanned aerial vehicle (UAV) platforms.

A novel approach for CR discrimination and quantification, the use of near-infrared (NIR) reflectance in conjunction with machine learning methods has recently shown potential for disease discrimination and quantification using both contact and image-based NIR sensors (Chapter 3, 4, 5, 6, 7). The NIR spectrum consists of electromagnetic wavelengths from 750–2500 nm, including wavelengths invisible to the human eye. In the literature, reflectance within the NIR spectrum has been reported to correlate to specific plant molecules, such as, starches and proteins, as well as, water content (Kumar et al. 2002). These correlations suggest that wavelengths within the NIR spectrum could be used for the detection of molecular changes to plant structure, indicative of stress. Further, these changes can be detected throughout the plant, not only where visible symptoms are presented, indicating potential for use in disease

detection early in the disease cycle, before significant visible symptoms appear (Kumar et al. 2002).

II. MATERIALS AND METHODS

A. Experimental design and plant cultivation

An experimental glasshouse trial was conducted in 2019 at the University of Southern Queensland's Centre for Crop Health (CCH) following the methodology outlined in Chapter 7. The trial included six replicates of plus and minus inoculated pots of five standard bread wheat genotypes, with a total of 60 plants. Plants were arranged in a randomised block design in a glasshouse with temperature maintained at 20–25° C. All plants were watered to field capacity once per week. *F. pseudograminearum* colonised wheat grain inoculum (Percy et al. 2012) was applied to the coleoptiles of the inoculated treatments at the two-leaf stage (14 days after planting). Wheat was harvested at maturity for disease severity assessment. Statistical design and analysis of disease severity data was conducted in collaboration with biostatisticians at Queensland Department of Agriculture and Fisheries (QDAF).

B. Near-infrared camera measurements

Video was recorded on a mini-DVR (digital video recorder) using five narrow-bandpass filters attached to the FLIR Tau™ SWIR camera core (900–1700 nm). Video was recorded each week for four weeks from Zadoks (Z) 13, 21 (Zadoks et al. 1974) to enable extraction of still image datasets. Bandpass filters were centered at 950, 1100, 1300, 1350 and 1600 nm with 50 nm bandwidths. These bands were selected based upon previous research in disease discrimination in Chapter 6. All measurements were taken from a tripod at a height of 1 metre above the plant for both inoculated and non-inoculated pots. Pots were placed in a tray containing a set of 2-inch calibrated reflectance Spectralon® panels. Panel reflectance was 2, 5, 10, 20, 40, 60 and 80% grey and 99% white.

C. Data analysis

All analysis was performed inside the Python computing environment (Python version 3.6.8, Python Software Foundation 2019), primarily in the Keras functional API (Chollet 2015) on top of TensorFlow (TensorFlow version 2.0.0-alpha0; Abadi et al. 2016). The SciPy ecosystem (Jones et al. 2001) and the Scikit-learn library (Pedregosa et al. 2011) were used for preprocessing.

C.1 Scoring systems

Three different scoring systems, with two to three classes, assigned from traditional stem browning percentages were evaluated (Fig. 1). These scoring systems were developed in Chapter 6 for quantification analysis of contact sensor data. While six scoring systems were developed as part of that study, the three evaluated here produced the highest accuracies in previous research of 67–96% dependent on rating system and data grouping (i.e. glasshouse, field, combined).

Rating System	% Stem Browning																				
	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
4	Group 1 (9)											Group 2 (69)									
5	ND	Group 1 (13)										Group 2 (69)									
6	ND	Group 1 (9)					Group 2 (49)														

Fig. 1: The three rating systems used for near-infrared image classification of CR and the average stem browning score (%) of samples used to develop each grouping. Not detectable is denoted as ND. Adapted from: (Chapter 6).

C.2 CNN Development

Using the Tau SWIR, videos were collected weekly for each plant using each of the five narrow-bandpass filters for four weeks, from three weeks post-inoculation at Zadoks (Z) 13, 21. Frames were removed from each video and labeled to develop a CR image classification dataset consisting of near-infrared images at the five wavelengths. The resulting weekly image datasets comprised 60 images and were labeled to correspond to a key comprising measurements for the three scales utilised. Three convolutional neural networks (CNNs) were developed, trained and validated on these datasets, comprising data from individual weeks (60 images/input), as well as, a combined dataset incorporating images from all of the four weeks (240 images/input) (Fig. 2).

Each image within these datasets was reshaped to 100x100 pixels, converted to grayscale and scaled to 0–1. A 75–25% train-test split was selected and applied to the resulting transformed datasets. The CNN models were then developed using the Keras functional API with TensorFlow (Chollet 2015; Abadi et al. 2016). These models used the Adam algorithm for stochastic gradient-based optimisation (Kingma & Ba 2014) as an optimiser and consisted of one, two or five-input models, one for each input waveband, a concatenation layer in the dual and five waveband models, a flatten layer, two dense layers, a dropout layer and a final dense layer for output (Fig. 2). The resulting CNNs were tested for accuracy based upon the average of 10 independent random validation splits per dataset. The equation used for accuracy is reported below (1):

$$accuracy\left(y, \hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \cdot 1\left(\hat{y}_i = y_i\right) \qquad (1)$$

Where $\hat{y}_i$ is the predicted value of the ith sample and y_i is the corresponding true value
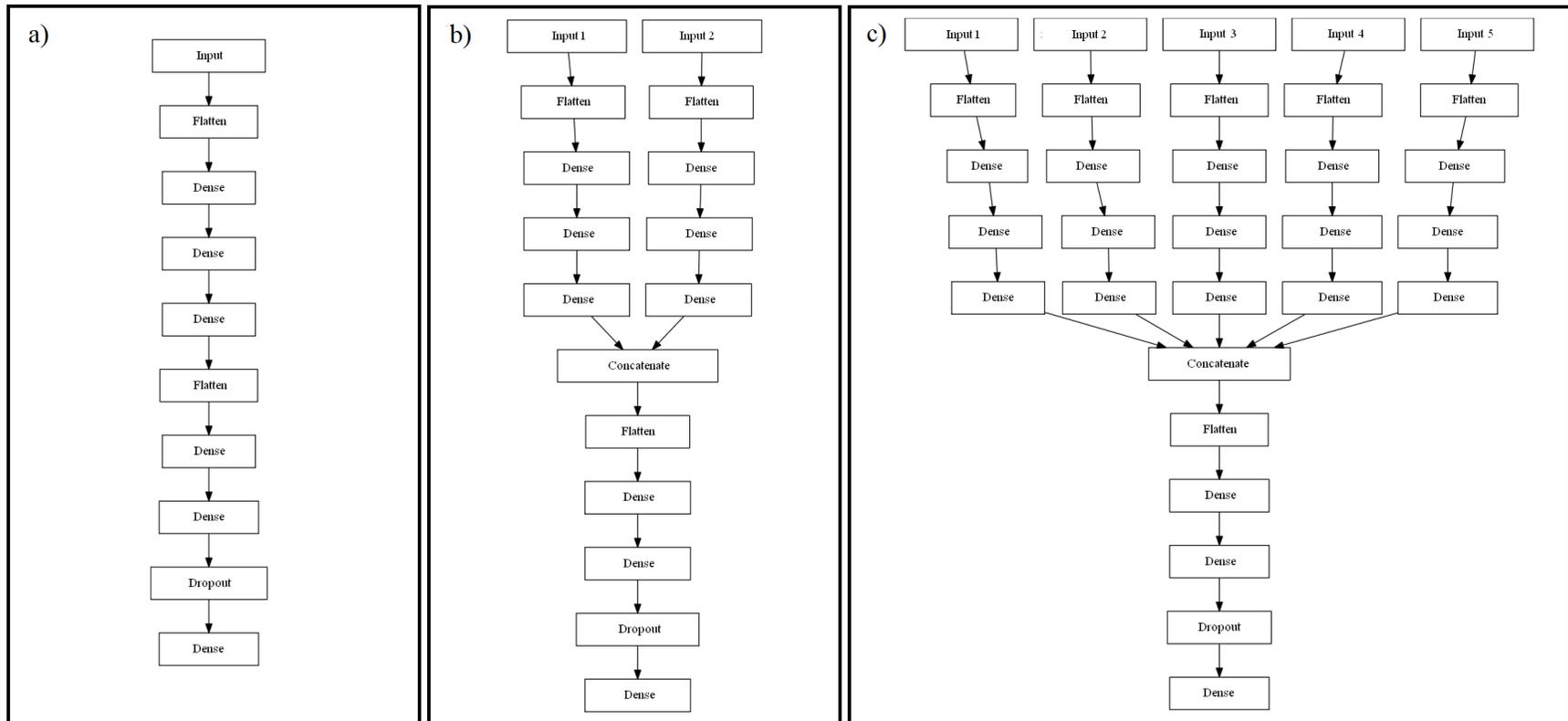
Fig. 2: Developed CNN architectures, consisting of 1 (a), 2 (b) or 5 (c) input networks.

III. RESULTS

A. Discrimination of disease: rating system 4

Rating system 4 divided the images into one of two classes, consisting of 0–50% stem browning and > 50% stem browning (Fig. 1). Individual average accuracies across the weekly datasets for rating system 4 ranged from 51% to 53%. Differences in accuracies between weeks and models were not statistically significant. However, accuracy across models was significantly higher with an average of 73% when trained on the combined dataset. Standard deviation was low with a range of 0–3% (Fig. 3). No single model and dataset combination could be identified as best using rating system 4, from the results of this trial.
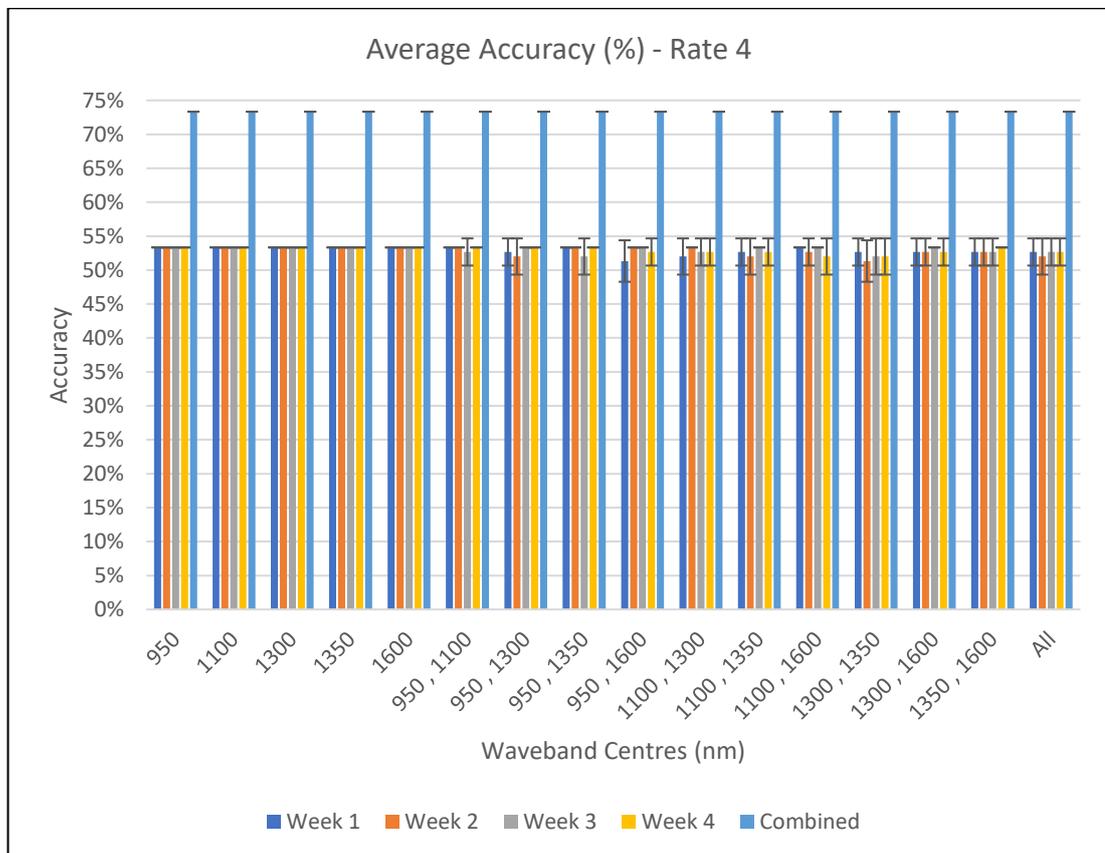
Fig. 3: Average mean accuracies and standard deviations for each dataset (weekly and combined) for each set of wavebands used in model development using rating system 4. The 'Combined' grouping includes all data from weeks 1– 4. The 'All' grouping includes data from all five wavebands (950, 1100, 1300, 1350 and 1600 nm). Weeks are reported as weeks from Zadoks (Z) 13, 21 (Zadoks et al. 1974).

B. Discrimination of disease: rating system 5

Rating system 5 divided the images into three classes, consisting of not detectable (ND), 0–50% stem browning and > 50% stem browning (Fig. 1). Average accuracies across the weekly datasets for rating system 5 ranged from 18% in week 1, using a model incorporating wavebands centered at 950 and 1600 nm to 40%, in week 2 using a single waveband model centered at 950 nm and week 3 using a single waveband model centered at 1300 nm (Fig. 4). Differences in average accuracies between weeks across models were not statistically significant. Accuracy across models was significantly highest with an average of 48% when trained on the combined dataset. Average standard deviation varied across weeks from 0 to 14% (Fig. 4). The best model across datasets using rating system 5, from the results of this trial, was identified as a single waveband model centered at 950 nm.
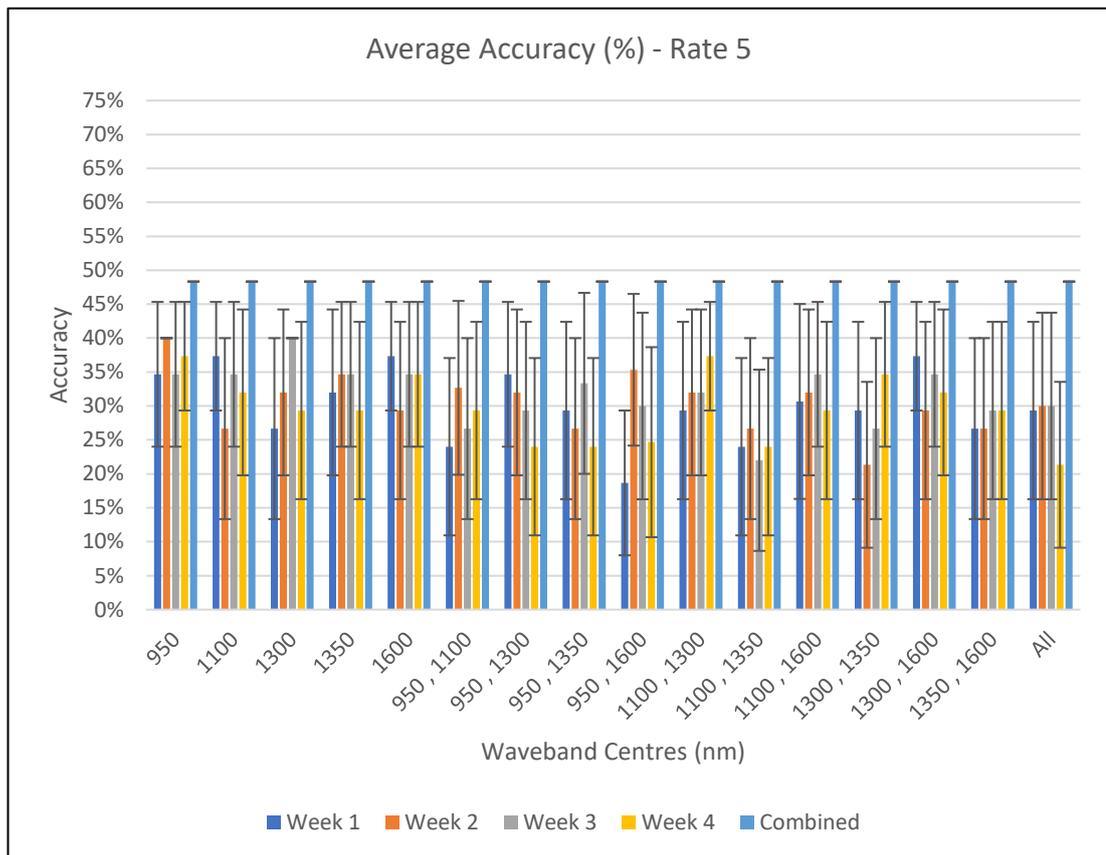
Fig. 4: Average mean accuracies and standard deviations for each dataset (weekly and combined) for each set of wavebands used in model development using rating system 5. The 'Combined' grouping includes all data from weeks 1– 4. The 'All' grouping includes data from all five wavebands (950, 1100, 1300, 1350 and 1600 nm). Weeks are reported as weeks from Zadoks (Z) 13, 21 (Zadoks et al. 1974).

C. Discrimination of disease: rating system 6

Rating system 6 also divided the images into three classes, consisting of not detectable (ND), 0–25% stem browning and > 25% stem browning (Fig. 1). Average accuracies for rating system 6 varied across the weekly datasets from 35–49% (Fig. 5). Evaluation of average accuracies across all developed models on each dataset showed no significant difference between weeks but indicated that developing models on combined data across weeks was significantly better on average, with an average of 47% across models. Standard deviation varied across datasets and models from 0 to 20% (Fig. 5). The best model developed using rating system 6 returned an average accuracy of 47% across datasets and was identified as a dual waveband model with wavebands centered at 950 and 1100 nm (Fig. 5).
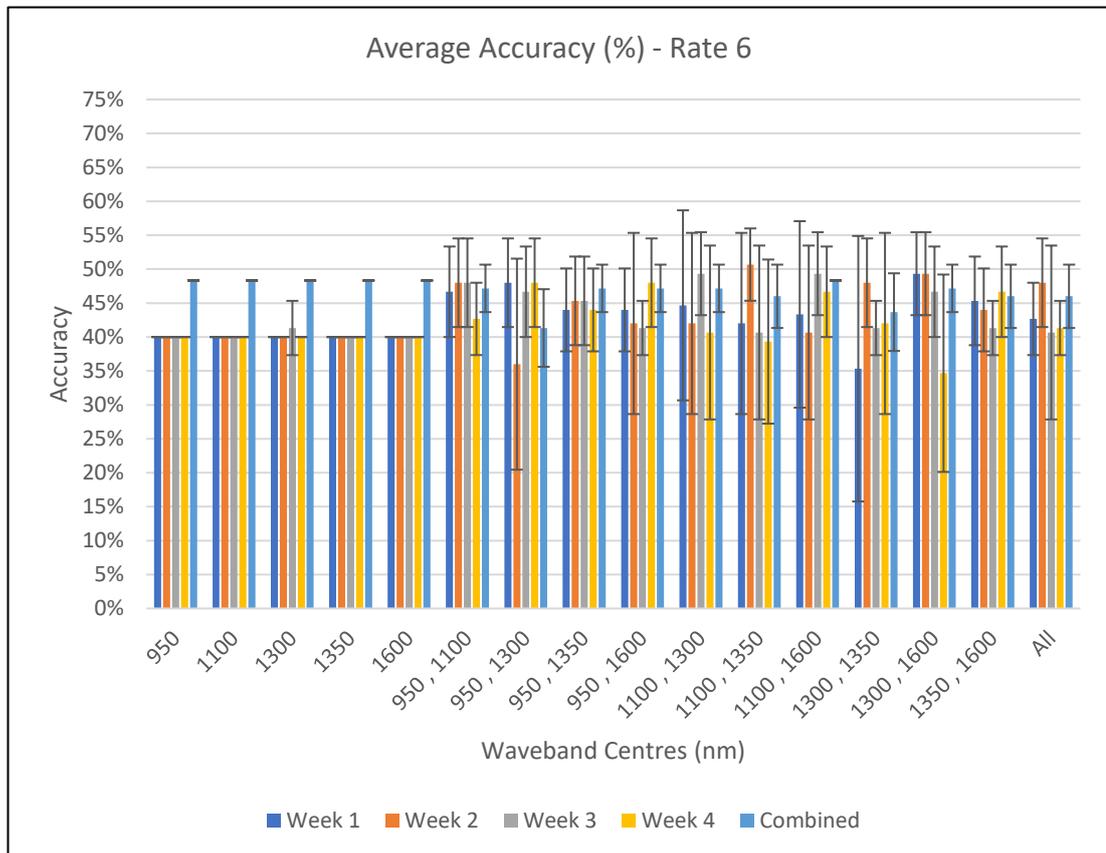
Fig. 5: Average mean accuracies and standard deviations for each dataset (weekly and combined) for each set of wavebands used in model development using rating system 6. The 'Combined' grouping includes all data from weeks 1– 4. The 'All' grouping includes data from all five wavebands (950, 1100, 1300, 1350 and 1600 nm). Weeks are reported as weeks from Zadoks (Z) 13, 21 (Zadoks et al. 1974).

IV. DISCUSSION

In Part A of this study, *Fusarium pseudograminearum* induced CR of wheat detection models were successfully developed using single, dual and five-input CNNs on a camera-based near-infrared sensor (Chapter 7). In this study single and multi-input CNNs were developed and tested for use in an HTP system to quantify CR into two or three disease severity classes. Data from five near-infrared image datasets, obtained from plus and minus inoculated bread wheat was analysed from three to six weeks post-inoculation, starting at Z 13, 21. The results of this study suggest potential for quantification of CR using a near-infrared camera system and a CNN to rapidly quantify CR under glasshouse conditions. Single waveband data and combinations of two and five 50 nm wavebands, with centres on 950, 1100, 1300, 1350 and 1600 nm exhibited varying ability to discriminate CR, depending on the model and dataset examined. In both rating system 5 and 6, the top performing models incorporated data from the 950 nm waveband, indicating some importance, relative to the other wavebands examined, in CR quantification using a camera-based system. However, average accuracies above 50% were only achieved using a binary quantification model on a combined dataset, with no single model determined to be best in this instance. This finding is counter to the findings in the development of a CNN for discrimination (+,-) of CR in Chapter 7, where average accuracies ranged from 53–100% and models performed better on individual weeks, than on combined data. Further, when developing crop disease detection models, multi-waveband models are often preferable to single waveband models as single waveband models must rely on relationships with external references. Multi-waveband models can incorporate internal relationships between bands to quantify crop disease (Mahlein et al. 2010; Zheng et al. 2019).

A stagnation of accuracies is seen in the combined, 'Rate 4' data in Fig. 3 suggesting that more training data are needed to increase accuracies. This is supported by the flattening of training accuracies (below) at epoch 4 in Figure 6, (a), while training loss continues to decrease (b), suggesting the stagnation of accuracies is due to a lack of new training data and not from overfitting the model to the available data.
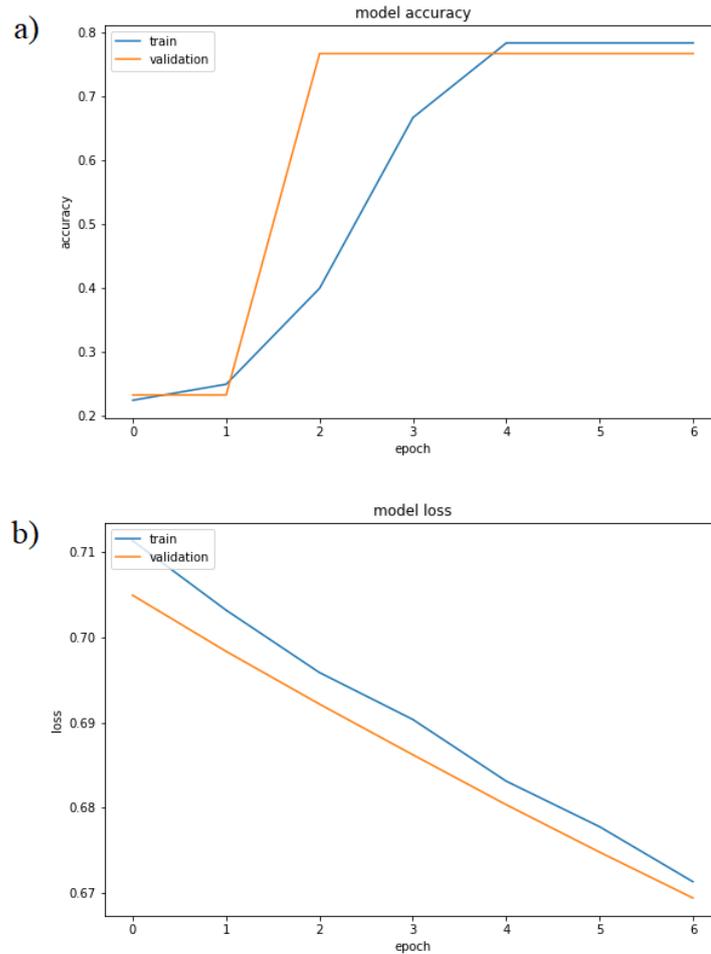
Fig. 6: Model accuracy (a) and loss (b) for training and validation datasets across epochs for a five-waveband CNN.

Our results indicate that the 'Rate 4', binary quantification system is the best system to use for the quantification of CR in bread wheat and the ideal time is on combined data for four weeks from Z 13, 21, with an average accuracy of 73% over 10 independent random validation splits (Fig. 3). However, further work should build upon the current dataset to increase available training data. The number of input images in this dataset is relatively small for a neural network multi-classifier, as compared to commonly used open-source image classification datasets like Imagenet (Deng et al. 2009) and Google's Open Images Dataset (Kuznetsova et al. 2018) with approximately 14 million and 9 million images respectively. Expanding the current CR image dataset could not only potentially increase accuracies but allow for training of quantification models with more than three classes.

The literature reports successful quantification of visible fungal infections of fruit and other commercial crops (i.e. chili, cotton, sugarcane) using machine learning and

specifically neural networks (Pujari et al. 2013a; Pujari et al. 2013b). However, the quantification success of the binary combined model in this study indicates potential for the future development of real-time phenotyping systems, using near-infrared cameras or multispectral sensing systems for both CR and other disease systems with difficult to observe symptoms. A system similar to that developed for the detection of septoria leaf blotch and yellow rust in wheat (Moshou et al. 2011) could be developed for the quantification of CR in wheat, improving germplasm delivery systems. Further work should validate these results on a larger, diverse set of training images and the potential of a system of this nature to phenotype CR in the field and in real-time should be evaluated. A successful CNN quantification model enables rapid high-throughput phenotyping (HTP) of CR. An HTP CR system would enable breeding companies to deliver germplasm to the market more rapidly and effectively.

Further indications from model analysis, discussed in this study, propose that a larger CR disease dataset should be obtained to both increase model accuracies and build successful quantification models with greater than two output classes.

Future work should focus on developing a larger, more diverse CR image dataset for model development, both with glasshouse and field images. Further work should also investigate whether models can be developed and successfully applied in field environments. Additional work should evaluate the potential to integrate this system into an HTP platform for use in CR breeding programs.

V. CONCLUSION

In this study, we built single, dual and five-waveband multi-input CNN classifiers for the quantification of *F. pseudograminearum* induced CR of wheat and evaluated the potential of these classifiers to phenotype CR resistance based on single and combinations of five near-infrared bands. These bands were obtained using a filtered NIR camera to simulate a multispectral system.

Potential for detection was evaluated across five image databases, built from images obtained in a glasshouse trial. These databases include sets of images for four individual weeks and a combined dataset including images from all four weeks.

The results of this study indicate that using CNNs to quantify CR in the glasshouse is possible. This is the first study, to the authors' knowledge to show potential for the successful quantification of CR using an image-based system. However, accuracies

greater than 50% were only obtained using a binary classification system on combined data from four weeks.

REFERENCES

Abadi, M, Agarwal, A, Barham, P, Brevdo, E, Chen, Z, Citro, C, Corrado, GS, Davis, A, Dean, J & Devin, M 2016, 'Tensorflow: Large-scale machine learning on heterogeneous distributed systems', *arXiv preprint* arXiv:1603.04467.

Chollet, F 2015, Keras, <https://keras.io>.

Deng, J, Dong, W, Socher, R, Li, L-J, Li, K & Fei-Fei, L 2009, 'Imagenet: A large-scale hierarchical image database', in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, pp. 248-55.

Jones, E, Oliphant, T & Peterson, P 2001, 'SciPy: Open Source Scientific Tools for Python'.

Kingma, DP & Ba, J 2014, 'Adam: A method for stochastic optimization', arXiv *preprint arXiv*:1412.6980.

Klein, T, Burgess, L & Ellison, F 1991, 'The incidence and spatial patterns of wheat plants infected by Fusarium graminearum Group 1 and the effect of crown rot on yield', *Australian Journal of Agricultural Research*, vol. 42, no. 3, pp. 399-407.

Knight, NL, Macdonald, B & Sutherland, MW 2017, 'Colonization of durum wheat (Triticum turgidum L. var. durum) culms exhibiting premature senescence (dead heads) associated with Fusarium pseudograminearum crown rot', *Plant Disease*, vol. 101, no. 10, pp. 1788-94.

Knight, NL, Macdonald, B, Percy, C & Sutherland, M 2020, 'Disease responses of hexaploid spring wheat (Triticum aestivum) culms exhibiting premature senescence (dead heads) associated with Fusarium pseudograminearum crown rot', *BioRxiv*.

Kumar, L, Schmidt, K, Dury, S & Skidmore, A 2002, 'Imaging spectrometry and vegetation science', in *Imaging spectrometry*, Springer, pp. 111-55.

Kuznetsova, A, Rom, H, Alldrin, N, Uijlings, J, Krasin, I, Pont-Tuset, J, Kamali, S, Popov, S, Malloci, M & Duerig, T 2018, 'The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale', *arXiv preprint* arXiv:1811.00982.

Mahlein, A-K, Steiner, U, Dehne, H-W & Oerke, E-C 2010, 'Spectral signatures of sugar beet leaves for the detection and differentiation of diseases', *Precision agriculture*, vol. 11, no. 4, pp. 413-31.

Malligan, CD 2009, 'Crown rot (Fusarium pseudograminearum) symptom development and pathogen spread in wheat genotypes with varying disease resistance', University of Southern Queensland.

Moshou, D, Bravo, C, Oberti, R, West, J, Ramon, H, Vougioukas, S & Bochtis, D 2011, 'Intelligent multi-sensor system for the detection and treatment of fungal diseases in arable crops', *Biosystems Engineering*, vol. 108, no. 4, pp. 311-21.

Murray, GM & Brennan, JP 2009, 'Estimating disease losses to the Australian wheat industry', *Australasian Plant Pathology*, vol. 38, no. 6, pp. 558-70.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R & Dubourg, V 2011, 'Scikit-learn: Machine learning in Python', *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-30.

Percy, C, Wildermuth, G & Sutherland, M 2012, 'Symptom development proceeds at different rates in susceptible and partially resistant cereal seedlings infected with Fusarium pseudograminearum', *Australasian Plant Pathology*, vol. 41, no. 6, pp. 621-31.

Pujari, JD, Yakkundimath, R & Byadgi, AS 2013a, 'Automatic fungal disease detection based on wavelet feature extraction and PCA analysis in commercial crops', *International Journal of Image, Graphics and Signal Processing*, vol. 6, no. 1, pp. 24-31.

Pujari, JD, Yakkundimath, RS & Byadgi, AS 2013b, 'Statistical methods for quantitatively detecting fungal disease from fruits' images', *International Journal of Intelligent Systems and Applications in Engineering*, vol. 1, no. 4, pp. 60-7.

Python version 3.6.8, Python Software Foundation 2019.

Simpfendorfer, S, McKay, A & Ophel-Keller, K 2019, 'New approaches to crop disease management in conservation agriculture. In (Eds J Pratley and J Kirkegaard) "Australian Agriculture in 2020: From Conservation to Automation" pp 173-188 (Agronomy Australia and Charles Sturt University: Wagga Wagga)', pp. 173-88.

Zadoks, JC, Chang, TT & Konzak, CF 1974, 'A decimal code for the growth stages of cereals', *Weed Research*, vol. 14, no. 6, pp. 415-21.

Zheng, Q, Huang, W, Cui, X, Dong, Y, Shi, Y, Ma, H & Liu, L 2019, 'Identification of Wheat Yellow Rust Using Optimal Three-Band Spectral Indices in Different Growth Stages', *Sensors*, vol. 19, no. 1, p. 35.

CHAPTER 9

CONCLUSIONS AND FUTURE RECOMENDATIONS

1 KEY FINDINGS

1.1 OBJECTIVE 1 - investigate available wavelet-based techniques for discriminating crop disease in the literature.

Discrete wavelet transform (DWT) techniques for signal processing in agriculture were investigated in a thorough literature review (Chapter 2). DWT-based analysis techniques have been shown to outperform traditional data analysis in hyperspectral data manipulation (Bruce & Li 2001; Bruce et al. 2002; Koger et al. 2003; Ge & Thomasson 2006; Ge et al. 2007).

Specifically of interest for this study, DWT was found to have been used to successfully detect fungal disease in fruit trees (Kempeneers et al. 2005). It was reported that features extracted from the DWT approach performed better than features extracted from original hyperspectral signatures. DWT was also used as a preprocessing step in a data fusion system for the detection of soybean rust disease (West et al. 2007). This system was derived from multiple DWTs at different scales. The DWT-based system outperformed an LDA-based system with more consistent accuracy range of 89–90% versus the 60–90% obtained by LDA. Recent work has indicated that wavelet-based spectral features can be used to successfully discriminate yellow rust and powdery mildew in winter wheat (Shi et al. 2017; Shi et al. 2018). Additionally, wavelets have been used to successfully extract fungal disease from multiple crop images for use in neural network classification, with accuracies of 86.48% (Pujari et al. 2013a).

It was determined that DWT used for hyperspectral data dimensionality reduction has been repeatedly shown to outperform traditional analysis in the literature. DWT presents advantages in original curve representation, feature number, computational efficiency and spectral waveband selection (Bruce & Li 2001; Ge & Thomasson 2006; Ge et al. 2007; Shi et al. 2017; Shi et al. 2018).

1.2 OBJECTIVE 2 - determine whether a discrete wavelet transform-based model can outperform traditional analysis techniques in the discrimination of crown rot of wheat in the near-infrared spectrum (900–1700 nm).

To determine whether a discrete wavelet transform-based model can outperform traditional analysis techniques in the discrimination of CR of wheat in the near-infrared spectrum (900–1700 nm) data was evaluated using a contact near-infrared sensor in three glasshouse and two field trials. Principal component analysis (PCA) and the discrete wavelet transform (DWT) were compared for ability to increase machine learning model disease prediction accuracy for the discrimination of CR, when used as a dimensionality reduction technique, across six machine learning methods (logistic regression classification, k-nearest neighbors, decision trees, extreme random forests, support vector machines, artificial neural networks). Additionally, two train-test data splits were evaluated, 20-80% and 80-20%. Accuracy was determined using four (initial glasshouse trial) or five standard bread wheat genotypes with varying susceptibility to CR. Discrimination ability was compared for models developed on individual weekly datasets (Chapter 3) and on combined datasets across weeks and trials (Chapter 4).

In this study, wavelet-based models performed better than PCA models when using a smaller train-test split. However, PCA models demonstrated more ability to detect CR under the larger, 80-20%, data split. Wavelet-based models also had higher minimum accuracies and a smaller, more consistent accuracy range than PCA models. These results indicate that wavelet-based models should be considered for agriculture applications where models must be more generalised to account for naturally occurring signature variances, influenced by environment. The discrimination difficulty caused by environmental and genetic diversity, and compounding of both abiotic and biotic stresses has been addressed in detail in the literature (Jacquemoud & Ustin 2001; Jacquemoud et al. 2009; Zhang et al. 2012; Devadas et al. 2015; Anderegg et al. 2019; Zheng et al. 2019). In addition to reducing spectral noise, smaller train-test splits may increase model development speed. Using train-test splits that are overly large could potentially lead to overfitting by including environmentally specific signatures, not only general signatures associated with CR infection.

Results of this study indicate that PCA has greater potential in CR discrimination (+,-) when used in combination with tree-based machine learning algorithms (e.g. decision

tree classifiers, extreme random forest classifiers). This finding is consistent with recent reports in the literature that PCA-based feature reduction can improve the accuracy of decision tree classification (Nasution et al. 2018). PCA-based models performed better than wavelet-based models when trained on multi-site data, returning higher overall accuracies. Additionally, our results indicate that PCA-based models used less features than wavelet models, overall, when obtaining top accuracies. This may be a consideration when developing multispectral sensors from models derived from hyperspectral signatures, as each additional feature equates to another sensor in the system. Fewer sensors equate to lower cost and weight. This is particularly important when developing sensing systems for use on UAV platforms such as those previously developed for the detection of lemon myrtle rust (*Austropuccinia psidii*) (Heim et al. 2018; 2019).

1.3 OBJECTIVE 3 - develop models to identify whether bandwidths corresponding to available commercial near-infrared filter sizes can discriminate and phenotype CR of wheat.

Models developed in Chapter 3 and 4 using a contact near-infrared sensor to discriminate CR removed all derived bandwidths less than 25 nm, resulting in all feature bands being greater than 27 nm. These models returned acceptable accuracies, dependent on grouping, of up to 79.17% for models developed from weekly data and 69.78% for generalised models.

Further, when limiting feature choice to bands larger than 25 nm for developing a phenotyping methodology using the contact NIR sensor (Chapter 6), average accuracy was as high as 96% when dividing the plants into two groupings using a DWT-based approach. In this case, field data outperformed models developed on glasshouse data.

Finally, single-band, dual-band and five-band multi-input CNNs were developed for the analysis of near-infrared image data with 50 nm bandwidths, to both detect and quantify CR (Chapter 7–8). The top performing detection model returned maximum average accuracies of 100% across 10 independent random validation splits at four and five weeks post-inoculation and consisted of two wavebands centered at 950 and 1350 nm. The top performing quantification models returned average accuracies of 73% across 10 independent random validation splits on a combined dataset incorporating data across a four week window.

These studies indicate that successful models for both the detection and phenotyping of CR in bread wheat can be developed using large bandwidths. Large bandwidths can easily be incorporated into novel multispectral sensing systems.

Further, preprocessing with the Savitzky-Golay filter and PCA or DWT to minimise noise in the signatures of large spectral bands was shown to be successful. This noise reduction is necessary as the narrow bands recorded by hyperspectral sensors are highly sensitive (Shafri & Yusof 2009; Rasti et al. 2018). Further, discrimination of CR using a contact NIR sensor or individual bands from a NIR camera has not been reported before these studies.

1.4 OBJECTIVE 4 - use commercially available filters on an imaging sensor to determine if the developed models can be utilised in an imaging system.

In the study outlined in Chapter 7, we developed a multi-input CNN classifier and evaluated the potential of this system to discriminate CR based on five near-infrared bands, deemed important in Chapters 3 and 4, obtained using a filtered NIR camera in a glasshouse environment. Detection ability was evaluated across five databases, including images from each of four individual weeks, from three weeks post-inoculation with *Fusarium pseudograminearum*, and a dataset using combined data from all weeks. Discrimination of CR was successful using this system and had not been reported before this study. Further, three scoring systems developed in Chapter 6 were evaluated for potential integration into a quantification CNN for CR phenotyping, using the five near-infrared bands (Chapter 8). Discrimination and quantification of CR using individual bands from a NIR imaging system, with the potential for real-time detection, indicates further potential for the development of a new CR specific multispectral sensor for deployment on multiple sensing platforms (e.g. robotic, UAV, handheld, equipment-based). An intelligent multi-sensor system has been developed for the detection of visible fungal disease on wheat (Moshou et al. 2011). However, no system yet exists for the detection or quantification of CR. The development of such a system could be expanded to include other stubble or soil-borne diseases with few visible symptoms.

Successful discrimination was achieved in the glasshouse using individual (weekly) and combined datasets. Results indicate that images used for discrimination should be taken between four and six weeks post-inoculation, with five weeks being the optimal

timepoint. This is consistent with reports of the importance of temporal patterning in plant disease detection and quantification (Franke & Menz 2007; Busemeyer et al. 2013; Bauriegel & Herppich 2014; Anderegg et al. 2019). Successful quantification was also obtained using the combined dataset and a binary classification system (Chapter 8), although further improvements in accuracy are likely with expansion of the CR image training database. This study indicates that CR discrimination and quantification is possible using commercially available filters on a NIR imaging sensor, with real-time detection capability.

2 ADDITIONAL FINDINGS

Six machine learning methods (logistic regression classification, k-nearest neighbors, decision trees, extreme random forests, support vector machines, artificial neural networks) were evaluated for ability to discriminate CR in wheat, in conjunction with PCA or DWT dimensionality reduction. Artificial neural networks, specifically multilayer perceptrons, were determined to be the most accurate machine learning method for the discrimination of CR, particularly when used in conjunction with DWT (Chapters 3, 4). This finding informed the decision to use a multilayer perceptron for the development of CR quantification models in Chapter 6 and the development of CNNs for image discrimination and quantification in Chapters 7 and 8. However, tree-based classifiers were found to be effective for use in conjunction with PCA-derived features, in agreement with the literature (Nasution et al. 2018).

Overall, DWT-based models performed better than PCA models when trained on smaller data splits (20-80%), while PCA-based models performed better when using larger data splits (80-20%). These findings may be important in informing preprocessing decisions in agriculture. If less training data are available, a DWT-based approach may be more desirable than a PCA-based approach, while also decreasing training time.

Results from ANOVA and post-hoc analysis indicate significant ($\alpha = 0.05$) temporal effects as early as one to two weeks from Z 13, 21, suggesting potential for early disease detection (Chapter 5). The resulting temporal patterning indicates timepoints of significance at early-tillering, late-tillering and the boot stage.

Further to the temporal effects observed, significant ($\alpha = 0.05$) spatial effects were seen when accuracies from the top twenty models in each group (tiller, centre, flag,

head, combined) were compared (Chapter 5). The observation of these spatial effects is significant as most current crop sensing systems look directly downward onto the canopy (Clark 2017) without taking into account specific areas of interest before data acquisition. Results indicate that a combination of measurements throughout the canopy profile (tiller, centre, flag) return significantly higher accuracies than any individual measurement. This finding indicates that a combination of measurements should be used when discriminating or phenotyping CR, specifically before flowering. The inclusion of head data into the models showed that combined data, again, outperformed individual scans in a majority of instances.

Finally, an ANOVA determined significant ($\alpha = 0.05$) interaction between timepoint and location of sensor scan across all data groupings (individual trials and combined data). The interaction between temporal and spatial data accounted for the majority of the variance observed across four of six data groupings (Eta^2, ω^2). The use of combined sensing data at week 5 from identification of Z 13, 21 was determined to be the optimal combination of timepoint and sensing location for successful CR detection.

3 FUTURE RECOMMENDATIONS

This study is the only study to our knowledge to evaluate near-infrared changes in CR reflectance throughout the growing season. Further, it is the first study to use wavelet-based, near-infrared signatures, in combination with machine learning techniques, to discriminate and phenotype CR of wheat. Future work should begin to investigate how temporal and spatial differences relate to pathogen toxicity, host response and disease progression.

Additional work should also investigate whether imaging-based CR detection can successfully be applied to field or natural systems and whether the quantification model developed in Chapter 8 can be applied to field systems with success as natural systems introduce increased spectral variation and canopy spectral properties are still in the process of being fully understood (Jacquemoud & Ustin 2001; Jacquemoud et al. 2009; Anderegg et al. 2019). Work should also evaluate the potential to integrate this image-based methodology into a platform for real-time detection of CR, available for deployment on robotic or drone-based platforms. Models developed in this research should also begin to be validated on a larger set of diverse germplasm.

Future work should also investigate the potential for a system to be built using this methodology for the detection and quantification of other crop diseases in both wheat and other cropping systems. The ability of and level to which these models can detect the differences between stresses, both biotic and abiotic should be evaluated. Finally, the ability of this methodology to be used to develop models for the discrimination of disease complexes and interactions should be assessed.

The overall success of the studies presented in this thesis shows the potential for the development of these models and frameworks into a high-throughput phenotyping (HTP) methodology for the quantification of CR resistance and tolerance, for the use in breeding programs both in Australia and globally. A machine-based HTP methodology would shift the breeding programs from reliance on human interpretation of visible disease symptoms, increasing rating accuracy and repeatability. Early disease detection and quantification would allow for only germplasm with the most potential for yield retention to be moved through the breeding program. This would free additional capital and resources throughout the program for investment into further lines.

4 CONCLUSIONS

The studies presented in this thesis resulted in several novel findings including the successful, automatic, non-destructive detection of CR that could previously only be detected manually; the successful implementation of wavelet analysis to reduce the volume of sensor data required for CR detection; the discovery of spatial and temporal effects impacting disease detection capability; the successful development and implementation of a CR phenotyping methodology using data from a contact NIR sensor; the development of real-time capable CNNs for the detection and quantification of CR using multiple NIR images as inputs. These findings are the first steps in developing an HTP system to provide plant breeders with new tools for quantifying CR resistance and tolerance, increasing the cost-effectiveness of such programs and ultimately delivering genetic solutions to growers more quickly. This research may also lead to the development of an independent multispectral sensor for the discrimination and/or quantification of CR, capable of large-scale coverage on an aerial platform, increasing opportunities for growers, researchers and breeders to understand and combat CR.

REFERENCES

Abadi, M, Agarwal, A, Barham, P, Brevdo, E, Chen, Z, Citro, C, Corrado, GS, Davis, A, Dean, J & Devin, M 2016, 'Tensorflow: Large-scale machine learning on heterogeneous distributed systems', *arXiv preprint arXiv:1603.04467*.

Adam, E, Deng, H, Odindi, J, Abdel-Rahman, EM & Mutanga, O 2017, 'Detecting the early stage of phaeosphaeria leaf spot infestations in maize crop using in situ hyperspectral data and guided regularized random forest algorithm', *Journal of Spectroscopy*, vol. 2017.

Ahmed, F, Al-Mamun, HA, Bari, AH, Hossain, E & Kwan, P 2012, 'Classification of crops and weeds from digital images: A support vector machine approach', *Crop Protection*, vol. 40, pp. 98-104.

Alahmad, S, Simpfendorfer, S, Bentley, A & Hickey, L 2018, 'Crown rot of wheat in Australia: Fusarium pseudograminearum taxonomy, population biology and disease management', *Australasian Plant Pathology*, vol. 47, no. 3, pp. 285-99.

Anderegg, J, Hund, A, Karisto, P & Mikaberidze, A 2019, 'In-field detection and quantification of Septoria tritici blotch in diverse wheat germplasm using spectral-temporal features', *BioRxiv*, p. 664011.

Arribas, JI, Sánchez-Ferrero, GV, Ruiz-Ruiz, G & Gómez-Gil, J 2011, 'Leaf classification in sunflower crops by computer vision and neural networks', *Computers and electronics in agriculture*, vol. 78, no. 1, pp. 9-18.

Backhouse, D 2006, 'Forecasting the risk of crown rot between successive wheat crops', *Australian Journal of Experimental Agriculture*, vol. 46, no. 11, pp. 1499-506.

Backhouse, D 2014, 'Modelling the behaviour of crown rot in wheat caused by Fusarium pseudograminearum', *Australasian Plant Pathology*, vol. 43, no. 1, pp. 15-23.

Backhouse, D & Burgess, L 2002, 'Climatic analysis of the distribution of Fusarium graminearum, F. pseudograminearum and F. culmorum on cereals in Australia', *Australasian Plant Pathology*, vol. 31, no. 4, pp. 321-7.

Bauriegel, E & Herppich, WB 2014, 'Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to Fusarium spec. infections on wheat', *Agriculture*, vol. 4, no. 1, pp. 32-57.

Behmann, J, Mahlein, A-K, Rumpf, T, Römer, C & Plümer, L 2015, 'A review of advanced machine learning methods for the detection of biotic stress in precision crop protection', *Precision agriculture*, vol. 16, no. 3, pp. 239-60.

REFERENCES

Bergsträsser, S, Fanourakis, D, Schmittgen, S, Cendrero-Mateo, MP, Jansen, M, Scharr, H & Rascher, U 2015, 'HyperART: non-invasive quantification of leaf traits using hyperspectral absorption-reflectance-transmittance imaging', *Plant methods*, vol. 11, no. 1, p. 1.

Bock, C, Parker, P, Cook, A & Gottwald, T 2008, 'Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves', *Plant Disease*, vol. 92, no. 4, pp. 530-41.

Bogue, R 2017, 'Sensors key to advances in precision agriculture', *Sensor Review*.

Bradley, AP 2003, 'Shift-invariance in the discrete wavelet transform', *Proceedings of VIIth Digital Image Computing: Techniques and Applications.* Sydney.

Bravo, C, Moshou, D, West, J, McCartney, A & Ramon, H 2003, 'Early disease detection in wheat fields using spectral reflectance', *Biosystems Engineering*, vol. 84, no. 2, pp. 137-45.

Breiman, L, Friedman, J, Olshen, R & Stone, C 1984, 'Classification and regression trees. Wadsworth Int', *Group*, vol. 37, no. 15, pp. 237-51.

Bruce, LM & Li, J 2001, 'Wavelets for computationally efficient hyperspectral derivative analysis', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1540-6.

Bruce, LM, Koger, CH & Li, J 2002, 'Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331-8.

Buitrago, MF, Groen, TA, Hecker, CA & Skidmore, AK 2016, 'Changes in thermal infrared spectra of plants caused by temperature and water stress', *ISPRS journal of photogrammetry and remote sensing*, vol. 111, pp. 22-31.

Burgess, L, Backhouse, D, Swan, L & Esdaile, R 1996, 'Control of Fusarium crown rot of wheat by late stubble burning and rotation with sorghum', *Australasian Plant Pathology*, vol. 25, no. 4, pp. 229-33.

Burgess, L, Backhouse, D, Summerell, B, Pattison, A, Klein, T, Esdaile, R & Ticehurst, G 1993, 'Long-term effects of stubble management on the incidence of infection of wheat by Fusarium graminearum Schw. Group 1', *Australian Journal of Experimental Agriculture*, vol. 33, no. 4, pp. 451-6.

Burks, T, Shearer, S, Heath, J & Donohue, K 2005, 'Evaluation of neural-network classifiers for weed species discrimination', *Biosystems Engineering*, vol. 91, no. 3, pp. 293-304.

Burrus, SC, Gopinath, RA & Guo, H 1998, 'Introduction to wavelets and wavelet transforms', *A Primer; Prentice Hall: Upper Saddle River, NJ, USA.*

Busemeyer, L, Ruckelshausen, A, Möller, K, Melchinger, AE, Alheit, KV, Maurer, HP, Hahn, V, Weissmann, EA, Reif, JC & Würschum, T 2013, 'Precision phenotyping of biomass accumulation in triticale reveals

temporal genetic patterns of regulation', *Scientific reports*, vol. 3, p. 2442.

Camargo, A & Smith, J 2009, 'Image pattern classification for the identification of disease causing agents in plants', *Computers and electronics in agriculture*, vol. 66, no. 2, pp. 121-5.

Campbell, PE, Middleton, E, Corp, L & Kim, M 2008, 'Contribution of chlorophyll fluorescence to the apparent vegetation reflectance', *Science of the total Environment*, vol. 404, no. 2-3, pp. 433-9.

Carter, GA, Freedman, A, Kebabian, P & Scott, H 2004, 'Use of a prototype instrument to detect short-term changes in solar-excited leaf fluorescence', *International Journal of Remote Sensing*, vol. 25, no. 9, pp. 1779-84.

Carter, GA, Jones, JH, Mitchell, RJ & Brewer, CH 1996, 'Detection of solar-excited chlorophyll a fluorescence and leaf photosynthetic capacity using a Fraunhofer line radiometer', *Remote Sensing of Environment*, vol. 55, no. 1, pp. 89-92.

Carter, GA & Knapp, AK 2001, 'Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration', *American journal of botany*, vol. 88, no. 4, pp. 677-84.

Chaerle, L, Hagenbeek, D, De Bruyne, E & Van Der Straeten, D 2007, 'Chlorophyll fluorescence imaging for disease-resistance screening of sugar beet', *Plant Cell, Tissue and Organ Culture*, vol. 91, no. 2, pp. 97-106.

Chaerle, L, Hagenbeek, D, De Bruyne, E, Valcke, R & Van Der Straeten, D 2004, 'Thermal and chlorophyll-fluorescence imaging distinguish plant-pathogen interactions at an early stage', *Plant and Cell Physiology*, vol. 45, no. 7, pp. 887-96.

Chakraborty, S, Obanor, F, Westecott, R & Abeywickrama, K 2010, 'Wheat crown rot pathogens Fusarium graminearum and F. pseudograminearum lack specialization', *Phytopathology*, vol. 100, no. 10, pp. 1057-65.

Chen, M, Glaz, B, Gilbert, RA, Daroub, SH, Barton, FE & Wan, Y 2002, 'Near-infrared reflectance spectroscopy analysis of phosphorus in sugarcane leaves', *Agronomy journal*, vol. 94, no. 6, pp. 1324-31.

Cheng, D, Zhang, S, Deng, Z, Zhu, Y & Zong, M 2014, *kNN Algorithm with Data-Driven k Value*.

Chollet, F 2015, *Keras*, <https://keras.io>.

Cintra, ME, Meira, CA, Monard, MC, Camargo, HA & Rodrigues, LH 2011, 'The use of fuzzy decision trees for coffee rust warning in Brazilian crops', in *2011 11th International Conference on Intelligent Systems Design and Applications*, IEEE, pp. 1347-52.

Clark, ML 2017, 'Comparison of simulated hyperspectral HyspIRI and multispectral Landsat 8 and Sentinel-2 imagery for multi-seasonal, regional land-cover mapping', *Remote Sensing of Environment*, vol. 200, pp. 311-25.

REFERENCES

Corp, L, Middleton, E, Daughtry, C & Campbell, P 2006, 'Solar induced fluorescence and reflectance sensing techniques for monitoring nitrogen utilization in corn', in *2006 IEEE International Symposium on Geoscience and Remote Sensing*, IEEE, pp. 2267-70.

Daubechies, I 1988, 'Orthonormal bases of compactly supported wavelets', *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909-96.

Delalieux, S, Van Aardt, J, Keulemans, W, Schrevens, E & Coppin, P 2007, 'Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications', *European Journal of Agronomy*, vol. 27, no. 1, pp. 130-43.

Deng, J, Dong, W, Socher, R, Li, L-J, Li, K & Fei-Fei, L 2009, 'Imagenet: A large-scale hierarchical image database', in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, pp. 248-55.

Devadas, R, Lamb, D, Backhouse, D & Simpfendorfer, S 2015, 'Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat', *Precision agriculture*, vol. 16, no. 5, pp. 477-91.

Dhau, I, Adam, E, Mutanga, O, Ayisi, K, Abdel-Rahman, EM, Odindi, J & Masocha, M 2018, 'Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot disease in maize crop', *Geocarto International*, vol. 33, no. 11, pp. 1223-36.

Dodman, R & Wildermuth, G 1987, 'Inoculation methods for assessing resistance in wheat to crown rot caused by Fusarium graminearum Group 1', *Australian Journal of Agricultural Research*, vol. 38, no. 3, pp. 473-86.

Elvanidi, A, Katsoulas, N, Augoustaki, D, Loulou, I & Kittas, C 2018, 'Crop reflectance measurements for nitrogen deficiency detection in a soilless tomato crop', *Biosystems Engineering*, vol. 176, pp. 1-11.

Evain, S, Ounis, A, Baret, F, Goulas, Y, Louis, J, Ducruet, J, Cerovic, Z & Moya, I 2002, 'Passive vegetation fluorosensing using atmospheric oxygen absorption bands', *Recent advances in quantitative remote sensing*, pp. 16-20.

Evans, ML, Hollaway, GJ, Dennis, JI, Correll, R & Wallwork, H 2010, 'Crop sequence as a tool for managing populations of Fusarium pseudograminearum and F. culmorum in south-eastern Australia', *Australasian Plant Pathology*, vol. 39, no. 4, pp. 376-82.

Filella, I & Peñuelas, J 1999, 'Altitudinal differences in UV absorbance, UV reflectance and related morphological traits of Quercus ilex and Rhododendron ferrugineum in the Mediterranean region', *Plant Ecology*, vol. 145, no. 1, pp. 157-65.

Fix, E 1951, *Discriminatory analysis: nonparametric discrimination, consistency properties*, USAF school of Aviation Medicine.

REFERENCES

Forknall, CR, Simpfendorfer, S & Kelly, AM 2019, 'Using yield response curves to measure variation in the tolerance and resistance of wheat cultivars to Fusarium crown rot', *Phytopathology*, vol. 109, no. 6, pp. 932-41.

Franke, J & Menz, G 2007, 'Multi-temporal wheat disease detection by multi-spectral remote sensing', *Precision agriculture*, vol. 8, no. 3, pp. 161-72.

Gamon, J, Penuelas, J & Field, C 1992, 'A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency', *Remote Sensing of Environment*, vol. 41, no. 1, pp. 35-44.

García-Gonzalo, E, Fernández-Muñiz, Z, Garcia Nieto, PJ, Sánchez, A & Menéndez, M 2016, 'Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers', *Materials*, vol. 9, p. 531.

Garcia-Ruiz, F, Sankaran, S, Maja, JM, Lee, WS, Rasmussen, J & Ehsani, R 2013, 'Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees', *Computers and electronics in agriculture*, vol. 91, pp. 106-15.

Ge, Y & Thomasson, J 2006, 'Wavelet incorporated spectral analysis for soil property determination', *Transactions of the ASABE*, vol. 49, no. 4, pp. 1193-201.

Ge, Y, Morgan, C, Thomasson, J & Waiser, T 2007, 'A new perspective to near-infrared reflectance spectroscopy: A wavelet approach', *Transactions of the ASABE*, vol. 50, no. 1, pp. 303-11.

Geurts, P, Ernst, D & Wehenkel, L 2006, 'Extremely randomized trees', *Machine learning*, vol. 63, no. 1, pp. 3-42.

Gitelson, AA, Merzlyak, MN & Chivkunova, OB 2001, 'Optical properties and nondestructive estimation of anthocyanin content in plant leaves¶', *Photochemistry and photobiology*, vol. 74, no. 1, pp. 38-45.

Gitelson, AA, Kaufman, YJ, Stark, R & Rundquist, D 2002, 'Novel algorithms for remote estimation of vegetation fraction', *Remote Sensing of Environment*, vol. 80, no. 1, pp. 76-87.

Goel, PK, Prasher, SO, Patel, RM, Landry, J-A, Bonnell, R & Viau, AA 2003, 'Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn', *Computers and electronics in agriculture*, vol. 39, no. 2, pp. 67-93.

Grieve, B, Hammersley, S, Mahlein, A-K, Oerke, E-C & Goldbach, H 2015, 'Localized multispectral crop imaging sensors: Engineering & validation of a cost effective plant stress and disease sensor', in *Sensors Applications Symposium (SAS), 2015 IEEE*, IEEE, pp. 1-6.

Grossmann, A, Kronland-Martinet, R & Morlet, J 1990, 'Reading and understanding continuous wavelet transforms', in *Wavelets*, Springer, pp. 2-20.

Haar, A 1909, *Zur theorie der orthogonalen funktionensysteme*, Georg-August-Universitat, Gottingen.

REFERENCES

Hagen, NA & Kudenov, MW 2013, 'Review of snapshot spectral imaging technologies', *Optical Engineering*, vol. 52, no. 9, p. 090901.

Hawkins, DM 2004, 'The problem of overfitting', *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 1-12.

Heim, R, Wright, I, Chang, HC, Carnegie, A, Pegg, G, Lancaster, E, Falster, D & Oldeland, J 2018, 'Detecting myrtle rust (Austropuccinia psidii) on lemon myrtle trees using spectral signatures and machine learning', *Plant pathology*, vol. 67, no. 5, pp. 1114-21.

Heim, RH, Wright, IJ, Scarth, P, Carnegie, AJ, Taylor, D & Oldeland, J 2019, 'Multispectral, aerial disease detection for myrtle rust (Austropuccinia psidii) on a lemon myrtle plantation', *Drones*, vol. 3, no. 1, p. 25.

Henderson, D, Williams, CJ & Miller, JS 2007, 'Forecasting late blight in potato crops of southern Idaho using logistic regression analysis', *Plant Disease*, vol. 91, no. 8, pp. 951-6.

Hofmann, M 2006, 'Support vector machines-kernels and the kernel trick', *Notes*, vol. 26, no. 3.

Hollaway, G, Evans, M, Wallwork, H, Dyson, C & McKay, A 2013, 'Yield loss in cereals, caused by Fusarium culmorum and F. pseudograminearum, is related to fungal DNA in soil prior to planting, rainfall, and cereal type', *Plant Disease*, vol. 97, no. 7, pp. 977-82.

Hollaway, GJ & Exell, GK 2010, 'Survey of wheat crops for white heads caused by crown rot in Victoria, 1997–2009', *Australasian Plant Pathology*, vol. 39, no. 4, pp. 363-7.

Holmes, MG & Keiller, DR 2002, 'Effects of pubescence and waxes on the reflectance of leaves in the ultraviolet and photosynthetic wavebands: a comparison of a range of species', *Plant, Cell & Environment*, vol. 25, no. 1, pp. 85-93.

Hou, M, Tian, F, Zhang, L, Li, S, Du, T, Huang, M & Yuan, Y 2019, 'Estimating crop transpiration of soybean under different irrigation treatments using thermal infrared remote sensing imagery', *Agronomy*, vol. 9, no. 1, p. 8.

Huang, W, Lamb, DW, Niu, Z, Zhang, Y, Liu, L & Wang, J 2007, 'Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging', *Precision agriculture*, vol. 8, no. 4-5, pp. 187-97.

Huete, A 1988, 'A soil-adjusted vegetation index (SAVI)', *Remote Sensing of Environment*, vol. 25, pp. 295-309.

Hunter, JD 2007, 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-5.

Jacquemoud, S & Baret, F 1990, 'PROSPECT: A model of leaf optical properties spectra', *Remote Sensing of Environment*, vol. 34, no. 2, pp. 75-91.

Jacquemoud, S & Ustin, SL 2001, 'Leaf optical properties: A state of the art', in *8th International Symposium of Physical Measurements & Signatures in Remote Sensing*, pp. 223-332.

REFERENCES

Jacquemoud, S, Verhoef, W, Baret, F, Bacour, C, Zarco-Tejada, PJ, Asner, GP, François, C & Ustin, SL 2009, 'PROSPECT+ SAIL models: A review of use for vegetation characterization', *Remote Sensing of Environment*, vol. 113, pp. S56-S66.

Jing, X & Bai, Z 2019, 'Remote Sensing Detection of Wheat Stripe Rust by Synergized Solar-Induced Chlorophyll Fluorescence and Differential Spectral Index', in *2019 6th International Conference on Information Science and Control Engineering (ICISCE)*, IEEE, pp. 1109-17.

Jolliffe, IT & Cadima, J 2016, 'Principal component analysis: a review and recent developments', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202.

Jones, E, Oliphant, T & Peterson, P 2001, 'SciPy: Open Source Scientific Tools for Python'.

Jones, HG, Serraj, R, Loveys, BR, Xiong, L, Wheaton, A & Price, AH 2009, 'Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field', *Functional Plant Biology*, vol. 36, no. 11, pp. 978-89.

Jones, HG, Stoll, M, Santos, T, Sousa, Cd, Chaves, MM & Grant, OM 2002, 'Use of infrared thermography for monitoring stomatal closure in the field: application to grapevine', *Journal of Experimental Botany*, vol. 53, no. 378, pp. 2249-60.

Joshi, AA & Jadhav, B 2016, 'Monitoring and controlling rice diseases using Image processing techniques', in *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, IEEE, pp. 471-6.

Kakani, V, Reddy, K, Zhao, D & Mohammed, A 2003, 'Effects of ultraviolet-B radiation on cotton (Gossypium hirsutum L.) morphology and anatomy', *Annals of Botany*, vol. 91, no. 7, pp. 817-26.

Karimi, Y, Prasher, S, Patel, R & Kim, S 2006, 'Application of support vector machine technology for weed and nitrogen stress detection in corn', *Computers and electronics in agriculture*, vol. 51, no. 1-2, pp. 99-109.

Kaufman, YJ & Tanre, D 1992, 'Atmospherically resistant vegetation index (ARVI) for EOS-MODIS', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 2, pp. 261-70.

Kazan, K & Gardiner, DM 2018, 'Fusarium crown rot caused by Fusarium pseudograminearum in cereal crops: recent progress and future prospects', *Molecular plant pathology*, vol. 19, no. 7, pp. 1547-62.

Kelly, A, Macdonald, B, Percy, C & Davies, P 2016, 'Selection of genotypes for resistance and tolerance to pathogens: a combined statistical analysis of yield and disease response Alison Kelly, Bethany Macdonald, Cassandra Percy & Phil Davies', in *9th Australasian Soilborne Diseases Symposium*: *Proceedings of the 9th Australasian Soilborne Diseases Symposium*, R Falloon (ed.), ASDS, Lincoln University, Christchurch, New Zealand.

REFERENCES

Kempeneers, P, De Backer, S, Debruyn, W, Coppin, P & Scheunders, P 2005, 'Generic wavelet-based hyperspectral classification applied to vegetation stress detection', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 610-4.

Kingma, DP & Ba, J 2014, 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980.*

Klein, ME, Aalderink, BJ, Padoan, R, De Bruin, G & Steemers, TAG 2008, 'Quantitative Hyperspectral Reflectance Imaging', *Sensors (Basel, Switzerland)*, vol. 8, no. 9, pp. 5576-618.

Klein, T, Summerell, B & Burgess, L 1988, 'Influence of stubble-management practices on crown rot of wheat', *Plant Protection Quarterly (Australia)*.

Klein, T, Burgess, L & Ellison, F 1991, 'The incidence and spatial patterns of wheat plants infected by Fusarium graminearum Group 1 and the effect of crown rot on yield', *Australian Journal of Agricultural Research*, vol. 42, no. 3, pp. 399-407.

Knight, NL & Sutherland, MW 2016, 'Histopathological assessment of Fusarium pseudograminearum colonization of cereal culms during crown rot infections', *Plant disease*, vol. 100, no. 2, pp. 252-9.

Knight, NL, Macdonald, B & Sutherland, MW 2017, 'Colonization of durum wheat (Triticum turgidum L. var. durum) culms exhibiting premature senescence (dead heads) associated with Fusarium pseudograminearum crown rot', *Plant Disease*, vol. 101, no. 10, pp. 1788-94.

Knight, NL, Macdonald, B, Percy, C & Sutherland, M 2020, 'Disease responses of hexaploid spring wheat (Triticum aestivum) culms exhibiting premature senescence (dead heads) associated with Fusarium pseudograminearum crown rot', *BioRxiv*.

Knipling, EB 1970, 'Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation', *Remote Sensing of Environment*, vol. 1, no. 3, pp. 155-9.

Koger, CH, Bruce, LM, Shaw, DR & Reddy, KN 2003, 'Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (Ipomoea lacunosa) in soybean (Glycine max)', *Remote Sensing of Environment*, vol. 86, no. 1, pp. 108-19.

Kumar, L, Schmidt, K, Dury, S & Skidmore, A 2002, 'Imaging spectrometry and vegetation science', in *Imaging spectrometry*, Springer, pp. 111-55.

Kuznetsova, A, Rom, H, Alldrin, N, Uijlings, J, Krasin, I, Pont-Tuset, J, Kamali, S, Popov, S, Malloci, M & Duerig, T 2018, 'The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale', *arXiv preprint arXiv:1811.00982.*

Labsphere, 'Reflectance characteristics of Spectralon panels', Reflectance Calibration Laboratory, Sutton, NH, USA. (1998).

Lamprecht, S, Marasas, W, Hardy, M & Calitz, F 2006, 'Effect of crop rotation on crown rot and the incidence of Fusarium pseudograminearum in

wheat in the Western Cape, South Africa', *Australasian Plant Pathology*, vol. 35, no. 4, pp. 419-26.

LeCun, Y, Bottou, L, Bengio, Y & Haffner, P 1998, 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-324.

Lee, GR, Gommers, R, Waselewski, F, Wohlfahrt, K & O'Leary, A 2019, 'PyWavelets: A Python package for wavelet analysis', *Journal of Open Source Software*, vol. 4, no. 36, p. 1237.

Lenthe, J-H, Oerke, E-C & Dehne, H-W 2007, 'Digital infrared thermography for monitoring canopy health of wheat', *Precision agriculture*, vol. 8, no. 1-2, pp. 15-26.

Leslie, JF & Summerell, BA 2006, 'Species Descriptions', in *The Fusarium Laboratory Manual*, Wiley-Blackwell, ch 13, pp. 121-278.

Liu, L 2018, 'Chapter 4 - Biostatistical Basis of Inference in Heart Failure Study', in L Liu (ed.), *Heart Failure: Epidemiology and Research Methods*, Elsevier, pp. 43-82.

Liu, Z-Y, Wu, H-F & Huang, J-F 2010, 'Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis', *Computers and electronics in agriculture*, vol. 72, no. 2, pp. 99-106.

Long, D, McCarthy, C & Jensen, T 2016, 'Row and water front detection from UAV thermal-infrared imagery for furrow irrigation monitoring', in *Advanced Intelligent Mechatronics (AIM), 2016 IEEE International Conference on*, IEEE, pp. 300-5.

López-Granados, F, Peña-Barragán, JM, Jurado-Expósito, M, Francisco-Fernández, M, Cao, R, Alonso-Betanzos, A & Fontenla-Romero, O 2008, 'Multispectral classification of grass weeds and wheat (Triticum durum) using linear and nonparametric functional discriminant analysis and neural networks', *Weed Research*, vol. 48, no. 1, pp. 28-37.

Lottes, P, Behley, J, Milioto, A & Stachniss, C 2018, 'Fully convolutional networks with sequential information for robust crop and weed detection in precision farming', *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2870-7.

Macdonald, B, Percy, C & Kelly, A 2016, 'Efficient sampling for assessing crown rot in wheat', in *9th Australasian Soilborne Diseases Symposium*: *Proceedings of the 9th Australasian Soilborne Diseases Symposium*, R Falloon (ed.), ASDS, Lincoln University, Christchurch, New Zealand p. 76.

Mahlein, A-K 2016, 'Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping', *Plant Disease*, vol. 100, no. 2, pp. 241-51.

Mahlein, A-K, Steiner, U, Dehne, H-W & Oerke, E-C 2010, 'Spectral signatures of sugar beet leaves for the detection and differentiation of diseases', *Precision agriculture*, vol. 11, no. 4, pp. 413-31.

REFERENCES

Mahlein, A-K, Oerke, E-C, Steiner, U & Dehne, H-W 2012, 'Recent advances in sensing plant diseases for precision crop protection', *European Journal of Plant Pathology*, vol. 133, no. 1, pp. 197-209.

Mahlein, A-K, Rumpf, T, Welke, P, Dehne, H-W, Plümer, L, Steiner, U & Oerke, E-C 2013, 'Development of spectral indices for detecting and identifying plant diseases', *Remote Sensing of Environment*, vol. 128, pp. 21-30.

Malligan, CD 2009, 'Crown rot (Fusarium pseudograminearum) symptom development and pathogen spread in wheat genotypes with varying disease resistance', University of Southern Queensland.

McCulloch, WS & Pitts, W 1943, 'A logical calculus of the ideas immanent in nervous activity', *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115-33.

McFarlane, J, Watson, RD, Theisen, AF, Jackson, RD, Ehrler, W, Pinter, P, Idso, SB & Reginato, R 1980, 'Plant stress detection by remote measurement of fluorescence', *Applied Optics*, vol. 19, no. 19, pp. 3287-9.

Meroni, M, Rossini, M, Guanter, L, Alonso, L, Rascher, U, Colombo, R & Moreno, J 2009, 'Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications', *Remote Sensing of Environment*, vol. 113, no. 10, pp. 2037-51.

Mewes, T, Franke, J & Menz, G 2011, 'Spectral requirements on airborne hyperspectral remote sensing data for wheat disease detection', *Precision agriculture*, vol. 12, no. 6, p. 795.

Mitchell, TM 1997, *Machine learning*, McGraw-Hill, Inc., New York, NY, USA.

Mitter, V, Zhang, M, Liu, C, Ghosh, R, Ghosh, M & Chakraborty, S 2006, 'A high-throughput glasshouse bioassay to detect crown rot resistance in wheat germplasm', *Plant pathology*, vol. 55, no. 3, pp. 433-41.

Moore, GK 1979, 'What is a picture worth? A history of remote sensing/Quelle est la valeur d'une image? Un tour d'horizon de télédétection', *Hydrological Sciences Bulletin*, vol. 24, no. 4, pp. 477-85.

Moshou, D, Ramon, H & De Baerdemaeker, J 2002, 'A weed species spectral detector based on neural networks', *Precision agriculture*, vol. 3, no. 3, pp. 209-23.

Moshou, D, Bravo, C, West, J, Wahlen, S, McCartney, A & Ramon, H 2004, 'Automatic detection of 'yellow rust'in wheat using reflectance measurements and neural networks', *Computers and electronics in agriculture*, vol. 44, no. 3, pp. 173-88.

Moshou, D, Bravo, C, Oberti, R, West, J, Bodria, L, McCartney, A & Ramon, H 2005, 'Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using Kohonen maps', *Real-Time Imaging*, vol. 11, no. 2, pp. 75-83.

Moshou, D, Bravo, C, Oberti, R, West, J, Ramon, H, Vougioukas, S & Bochtis, D 2011, 'Intelligent multi-sensor system for the detection and treatment

of fungal diseases in arable crops', *Biosystems Engineering*, vol. 108, no. 4, pp. 311-21.

Murray, G & Brennan, J 2010, 'Estimating disease losses to the Australian barley industry', *Australasian Plant Pathology*, vol. 39, no. 1, pp. 85-96.

Murray, GM & Brennan, JP 2009, 'Estimating disease losses to the Australian wheat industry', *Australasian Plant Pathology*, vol. 38, no. 6, pp. 558-70.

Nasution, M, Sitompul, O & Ramli, M 2018, 'PCA based feature reduction to improve the accuracy of decision tree c4. 5 classification', in *Journal of Physics: Conference Series*, IOP Publishing, p. 012058.

Neumann, M, Hallau, L, Klatt, B, Kersting, K & Bauckhage, C 2014, 'Erosion band features for cell phone image based plant disease classification', in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, IEEE, pp. 3315-20.

Newton, A & Hackett, C 1994, 'Subjective components of mildew assessment on spring barley', *European Journal of Plant Pathology*, vol. 100, no. 6, pp. 395-412.

Nutter, J 1997, 'Disease severity assessment training', *Exercises in plant disease epidemiology*, pp. 1-7.

Nutter Jr, F, Gleason, M, Jenco, J & Christians, N 1993, 'Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems', *Phytopathology*, vol. 83, no. 8, pp. 806-12.

Oerke, E-C, Fröhling, P & Steiner, U 2011, 'Thermographic assessment of scab disease on apple leaves', *Precision agriculture*, vol. 12, no. 5, pp. 699-715.

Oerke, E, Steiner, U, Dehne, H & Lindenthal, M 2006, 'Thermal imaging of cucumber leaves affected by downy mildew and environmental conditions', *Journal of Experimental Botany*, vol. 57, no. 9, pp. 2121-32.

Ophel-Keller, K, McKay, A, Hartley, D & Curran, J 2008, 'Development of a routine DNA-based testing service for soilborne diseases in Australia', *Australasian Plant Pathology*, vol. 37, no. 3, pp. 243-53.

Parry, D, Pettitt, T, Jenkinson, P & Lees, A 1994, 'The cereal Fusarium complex', *Ecology of plant pathogens*, pp. 301-20.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R & Dubourg, V 2011, 'Scikit-learn: Machine learning in Python', *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825-30.

Peiris, KH, Pumphrey, MO & Dowell, FE 2009, 'NIR absorbance characteristics of deoxynivalenol and of sound and Fusarium-damaged wheat kernels', *Journal of Near Infrared Spectroscopy*, vol. 17, no. 4, pp. 213-21.

Penuelas, J, Baret, F & Filella, I 1995, 'Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance', *Photosynthetica*, vol. 31, no. 2, pp. 221-30.

REFERENCES

Percy, C, Wildermuth, G & Sutherland, M 2012, 'Symptom development proceeds at different rates in susceptible and partially resistant cereal seedlings infected with Fusarium pseudograminearum', *Australasian Plant Pathology*, vol. 41, no. 6, pp. 621-31.

Petronaitis, T, Forknall, C & Simpfendorfer, S 2018, 'Microwave radiation reduces survival of Fusarium pseudograminearum in durum wheat stubble', *Australasian Plant Pathology*, vol. 47, no. 4, pp. 375-8.

Pinto, LS, Ray, A, Reddy, MU, Perumal, P & Aishwarya, P 2016, 'Crop disease classification using texture analysis', in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pp. 825-8.

Polder, G, van der Heijden, GW, van Doorn, J & Baltissen, TA 2014, 'Automatic detection of tulip breaking virus (TBV) in tulip fields using machine vision', *Biosystems Engineering*, vol. 117, pp. 35-42.

Prabhakar, M, Prasad, YG, Desai, S, Thirupathi, M, Gopika, K, Rao, GR & Venkateswarlu, B 2013, 'Hyperspectral remote sensing of yellow mosaic severity and associated pigment losses in Vigna mungo using multinomial logistic regression models', *Crop Protection*, vol. 45, pp. 132-40.

Prasad, S, Peddoju, SK & Ghosh, D 2016, 'Multi-resolution mobile vision system for plant leaf disease diagnosis', *Signal, Image and Video Processing*, vol. 10, no. 2, pp. 379-88.

Pujari, JD, Yakkundimath, R & Byadgi, AS 2013a, 'Automatic fungal disease detection based on wavelet feature extraction and PCA analysis in commercial crops', *International Journal of Image, Graphics and Signal Processing*, vol. 6, no. 1, pp. 24-31.

Pujari, JD, Yakkundimath, RS & Byadgi, AS 2013b, 'Statistical methods for quantitatively detecting fungal disease from fruits' images', *International Journal of Intelligent Systems and Applications in Engineering*, vol. 1, no. 4, pp. 60-7.

Purss, G 1969, 'Studies of varietal resistance to crown rot of wheat caused by Fusarium graminearum Schw', *Queensl. J. Agric. Anim. Sci.*, vol. 23, pp. 475-98.

Python version 3.6.8, Python Software Foundation 2019.

Rasti, B, Scheunders, P, Ghamisi, P, Licciardi, G & Chanussot, J 2018, 'Noise reduction in hyperspectral imagery: Overview and application', *Remote Sensing*, vol. 10, no. 3, p. 482.

Rees, S, McCarthy, C, Artizzu, X, Baillie, C & Dunn, M 2009, 'Development of a prototype precision spot spray system using image analysis and plant identification technology', in *Proceedings of the 2009 CIGR International Symposium of the Australian Society for Engineering in Agriculture (SEAg 2009)*, Engineers Australia, pp. 343-9.

Ribeiro da Luz, B & Crowley, JK 2007, 'Spectral reflectance and emissivity features of broad leaf plants: Prospects for remote sensing in the thermal

infrared (8.0–14.0 μm)', *Remote Sensing of Environment*, vol. 109, no. 4, pp. 393-405.

Ricker, N 1943, 'Further developments in the wavelet theory of seismogram structure', *Bulletin of the Seismological Society of America*, vol. 33, no. 3, pp. 197-228.

Rouse Jr, JW, Haas, R, Schell, J & Deering, D 1974, 'Monitoring vegetation systems in the Great Plains with ERTS'.

Rumpf, T, Mahlein, A-K, Steiner, U, Oerke, E-C, Dehne, H-W & Plümer, L 2010, 'Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance', *Computers and electronics in agriculture*, vol. 74, no. 1, pp. 91-9.

Rumpf, T, Römer, C, Weis, M, Sökefeld, M, Gerhards, R & Plümer, L 2012, 'Sequential support vector machine classification for small-grain weed species discrimination with special regard to Cirsium arvense and Galium aparine', *Computers and electronics in agriculture*, vol. 80, pp. 89-96.

Rundel, RD 1983, 'Action spectra and estimation of biologically effective UV radiation', *Physiologia Plantarum*, vol. 58, no. 3, pp. 360-6.

Sá, J, Almeida, A, Pereira da Rocha, B, Mota, M, De Souza, JR & Dentel, L 2016, *Lightning Forecast Using Data Mining Techniques On Hourly Evolution Of The Convective Available Potential Energy*.

Sankaran, S, Mishra, A, Ehsani, R & Davis, C 2010, 'A review of advanced techniques for detecting plant diseases', *Computers and electronics in agriculture*, vol. 72, no. 1, pp. 1-13.

Saremi, H, Okhovvat, S & Ashrafi, S 2011, 'Fusarium diseases as the main soil borne fungal pathogen on plants and their control management with soil solarization in Iran', *African Journal of Biotechnology*, vol. 10, no. 80, pp. 18391-8.

Savitzky, A & Golay, MJ 1964, 'Smoothing and differentiation of data by simplified least squares procedures', *Analytical chemistry*, vol. 36, no. 8, pp. 1627-39.

Seabold, S & Perktold, J 2010, 'Statsmodels: Econometric and statistical modeling with python', in *Proceedings of the 9th Python in Science Conference*, Austin, TX, p. 61.

Shafri, HZM & Yusof, MRM 2009, 'Trends and issues in noise reduction for hyperspectral vegetation reflectance spectra', *Eur. J. Sci. Res*, vol. 29, pp. 404-10.

Sherwood, R, Berg, C, Hoover, M & Zeiders, K 1983, 'Illusions in visual assessment of Stagonospora leaf spot of orchardgrass', *Phytopathology*, vol. 73, no. 2, pp. 173-7.

Shi, Y, Huang, W & Zhou, X 2017, 'Evaluation of wavelet spectral features in pathological detection and discrimination of yellow rust and powdery mildew in winter wheat with hyperspectral reflectance data', *Journal of Applied Remote Sensing*, vol. 11, no. 2, p. 026025.

REFERENCES

Shi, Y, Huang, W, González-Moreno, P, Luke, B, Dong, Y, Zheng, Q, Ma, H & Liu, L 2018, 'Wavelet-Based Rust Spectral Feature Set (WRSFs): A Novel Spectral Feature Set Based on Continuous Wavelet Transformation for Tracking Progressive Host–Pathogen Interaction of Yellow Rust on Wheat', *Remote Sensing*, vol. 10, no. 4, p. 525.

Simpfendorfer, S, Brettell, R & Nicol, J 2012, 'Inter-row sowing reduces crown rot in winter cereals', in *First International Crown Rot Workshop for wheat improvement'. Narrabri.(Organising Committee of the 1st International Crown Rot Workshop)*.

Simpfendorfer, S, McKay, A & Ophel-Keller, K 2019, 'New approaches to crop disease management in conservation agriculture. In (Eds J Pratley and J Kirkegaard) "Australian Agriculture in 2020: From Conservation to Automation" pp 173-188 (Agronomy Australia and Charles Sturt University: Wagga Wagga)', in pp. 173-88.

Skov, J, Lemmens, M & Giese, H 2004, 'Role of a Fusarium culmorum ABC transporter (FcABC1) during infection of wheat and barley', *Physiological and molecular plant pathology*, vol. 64, no. 5, pp. 245-54.

Smiley, RW, Gourlie, JA, Easley, SA, Patterson, L-M & Whittaker, RG 2005, 'Crop damage estimates for crown rot of wheat and barley in the Pacific Northwest', *Plant Disease*, vol. 89, no. 6, pp. 595-604.

Steddom, K, Bredehoeft, M, Khan, M & Rush, C 2005, 'Comparison of visual and multispectral radiometric disease evaluations of Cercospora leaf spot of sugar beet', *Plant Disease*, vol. 89, no. 2, pp. 153-8.

Steinmüller, D & Tevini, M 1985, 'Action of ultraviolet radiation (UV-B) upon cuticular waxes in some crop plants', *Planta*, vol. 164, no. 4, pp. 557-64.

Stellmes, M, Schickling, A, Werner, W & Hill, J 2007, 'Detection of chlorophyll fluorescence by active and passive methods—field experiments with barley', in *Proc. of the 3th Int. Workshop on Remote Sensing of Vegetation Fluorescence*.

Sui, R, Thomasson, JA, Ge, Y & Morgan, C 2008, 'Multispectral sensor for in-situ cotton fiber quality measurement', *Transactions of the ASABE*, vol. 51, no. 6, pp. 2201-8.

Summerell, B, Burgess, L & Klein, T 1989, 'The impact of stubble management on the incidence of crown rot of wheat', *Australian Journal of Experimental Agriculture*, vol. 29, no. 1, pp. 91-8.

Suzuki, Y, Okamoto, H & Kataoka, T 2008, 'Image segmentation between crop and weed using hyperspectral imaging for weed detection in soybean field', *Environmental Control in Biology*, vol. 46, no. 3, pp. 163-73.

Tian, F, Qiu, G, Lü, Y, Yang, Y & Xiong, Y 2014, 'Use of high-resolution thermal infrared remote sensing and "three-temperature model" for transpiration monitoring in arid inland river catchment', *Journal of hydrology*, vol. 515, pp. 307-15.

Thomas, S, Wahabzada, M, Kuska, MT, Rascher, U & Mahlein, A-K 2017, 'Observation of plant–pathogen interaction by simultaneous

hyperspectral imaging reflection and transmission measurements', *Functional Plant Biology*, vol. 44, no. 1, pp. 23-34.

Van der Tol, C, Verhoef, W & Rosema, A 2009, 'A model for chlorophyll fluorescence and photosynthesis at leaf scale', *Agricultural and forest meteorology*, vol. 149, no. 1, pp. 96-105.

Verhulst, P-F 1838, 'Notice sur la loi que la population suit dans son accroissement (Notice on the law of population growth)', *Corresp. Math. Phys.*, vol. 10, pp. 113-26.

Wallwork, H, Butt, M, Cheong, J & Williams, K 2004, 'Resistance to crown rot in wheat identified through an improved method for screening adult plants', *Australasian Plant Pathology*, vol. 33, no. 1, pp. 1-7.

Wang, X, Zhang, M, Zhu, J & Geng, S 2008, 'Spectral prediction of Phytophthora infestans infection on tomatoes using artificial neural network (ANN)', *International Journal of Remote Sensing*, vol. 29, no. 6, pp. 1693-706.

Waskom, M, Botvinnik, O, Hobson, P, Cole, JB, Halchenko, Y, Hoyer, S, Miles, A, Augspurger, T, Yarkoni, T, Megies, T, Coelho, LP, Wehner, D, cynddl, Ziegler, E, diego0020, Zaytsev, YV, Hoppe, T, Seabold, S, Cloud, P, Koskinen, M, Meyer, K, Qalieh, A & Allan, D 2014, *seaborn: v0.5.0 (November 2014)*, 0.5.0, <https://doi.org/10.5281/zenodo.12710>.

Wasserman, PD & Schwartz, T 1988, 'Neural networks. II. What are they and why is everybody so interested in them now?', *IEEE Expert*, vol. 3, no. 1, pp. 10-5.

West, T, Prasad, S & Bruce, LM 2007, 'Multiclassifiers and decision fusion in the wavelet domain for exploitation of hyperspectral data', in *2007 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp. 4850-3.

Wildermuth, G & McNamara, R 1994, 'Testing wheat seedlings for resistance to crown rot caused by Fusarium graminearum Group 1', *Plant disease (USA)*.

Wildermuth, G, Thomas, G, Radford, B, McNamara, R & Kelly, A 1997, 'Crown rot and common root rot in wheat grown under different tillage and stubble treatments in southern Queensland, Australia', *Soil and Tillage Research*, vol. 44, no. 3-4, pp. 211-24.

Williams, DL, Goward, S & Arvidson, T 2006, 'Landsat', *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 10, pp. 1171-8.

Woolley, JT 1971, 'Reflectance and transmittance of light by leaves', *Plant physiology*, vol. 47, no. 5, pp. 656-62.

Wu, D, Feng, L, Zhang, C & He, Y 2008, 'Early detection of Botrytis cinerea on eggplant leaves based on visible and near-infrared spectroscopy', *Transactions of the ASABE*, vol. 51, no. 3, pp. 1133-9.

Yang, C-C, Prasher, S & Goel, P 2004, 'Differentiation of crop and weeds by decision-tree analysis of multi-spectral data', *Transactions of the ASAE*, vol. 47, no. 3, p. 873.

REFERENCES

Yang, C-C, Prasher, SO, Enright, P, Madramootoo, C, Burgess, M, Goel, PK & Callum, I 2003, 'Application of decision tree technology for image classification using remote sensing data', *Agricultural Systems*, vol. 76, no. 3, pp. 1101-17.

Yang, F, Li, J, Gan, X, Qian, Y, Wu, X & Yang, Q 2010, 'Assessing nutritional status of Festuca arundinacea by monitoring photosynthetic pigments from hyperspectral data', *Computers and electronics in agriculture*, vol. 70, no. 1, pp. 52-9.

Zadoks, JC, Chang, TT & Konzak, CF 1974, 'A decimal code for the growth stages of cereals', *Weed Research*, vol. 14, no. 6, pp. 415-21.

Zarco-Tejada, PJ, Berni, JA, Suárez, L, Sepulcre-Cantó, G, Morales, F & Miller, JR 2009, 'Imaging chlorophyll fluorescence with an airborne narrow-band multispectral camera for vegetation stress detection', *Remote Sensing of Environment*, vol. 113, no. 6, pp. 1262-75.

Zhang, J, Pu, R, Huang, W, Yuan, L, Luo, J & Wang, J 2012, 'Using in-situ hyperspectral data for detecting and discriminating yellow rust disease from nutrient stresses', *Field Crops Research*, vol. 134, pp. 165-74.

Zhang, S, Shang, Y & Wang, L 2015, 'Plant disease recognition based on plant leaf image', *J. Anim. Plant Sci*, vol. 25, no. 3, pp. 42-5.

Zheng, Q, Huang, W, Cui, X, Dong, Y, Shi, Y, Ma, H & Liu, L 2019, 'Identification of Wheat Yellow Rust Using Optimal Three-Band Spectral Indices in Different Growth Stages', *Sensors*, vol. 19, no. 1, p. 35.

Вапник, В & Червонекис, А 1964, 'Об одном классе алгоритмов обучения распознаванию образов (On one class of pattern recognition learning algorithms)', *Автоматика и телемеханика*, vol. 25, no. 6, p. 937.

APPENDIX A – GLOSSARY

GLOSSARY

Ascospores: a spore contained or produced within an ascus, specific to ascomycete fungi

Boot: stage of cereal development where pollen forms, head develops and flag leaf is fully emerged

Early Tillering: stage of cereal development where the main shoot has four to five leaves and two or more tillers have formed

Flowering: stage of cereal development when final grain number has been determined identified by freely hanging anthers

Hyperspectral: acquisition of hundreds of contiguous wavebands across the electromagnetic spectrum

Inflorescence: the complete flower of a plant

Macroconidia: a large asexual fungal spore

Multispectral: acquisition of a specific number of spaced wavebands across the electromagnetic spectrum

Near-infrared: spectrum of light from 750–2500 nm

Overfitting: producing an algorithm which too closely fits the problem data, causing a failure to predict further data points

PREDICTA® B: a DNA-based soil test which identifies soil-borne pathogens from soil samples

Resistance: ability of a host crop to combat a pathogen

Solarisation: the use of sunlight to heat soil to a temperature lethal to pests or pathogens

Short-wave Infrared: spectrum of light from 1400–3000 nm

APPENDIX A. GLOSSARY

Spectral Vegetation Index: a transformation of two or more spectral bands to allow improved comparison of spectral phenomena

Spectroscopy: the study of the interaction of matter with electromagnetic radiation

Stem Extension: when cereals begin to extend their stems

Stomata: pores in the epidermis of a plant's tissue, allowing movement of gases into cells

Thermal Infrared: spectrum of light from 8000–15000 nm

Tillering: the creation of lateral shoots from a plant

Tolerance: ability of a host crop to resist yield impact when under pathogen load

Two-leaf Stage: when the third leaf is present, but not yet fully expanded

Ultraviolet: spectrum of light from 10–400 nm

Waveband: a band of adjacent frequencies

Wavelet: an oscillation with an amplitude that begins and ends at zero

Whiteheads: the premature death of plant flowers

APPENDIX B – TRIAL LAYOUT

Chapters 3-6 all used near infrared point sensor data from a total of five trials, which included two glasshouse and one field trial in 2018, and one glasshouse and one field trial in 2019. Chapters 7 and 8 used near infrared camera data from the glasshouse trial in 2019.

B.1 1ST GLASSHOUSE TRIAL

Layout of the first glasshouse trial at the Leslie Research Centre in 2018. Forty-eight inoculated (red) and forty-eight non-inoculated (green) pots of four genotypes (1–4) were placed in a randomised design.
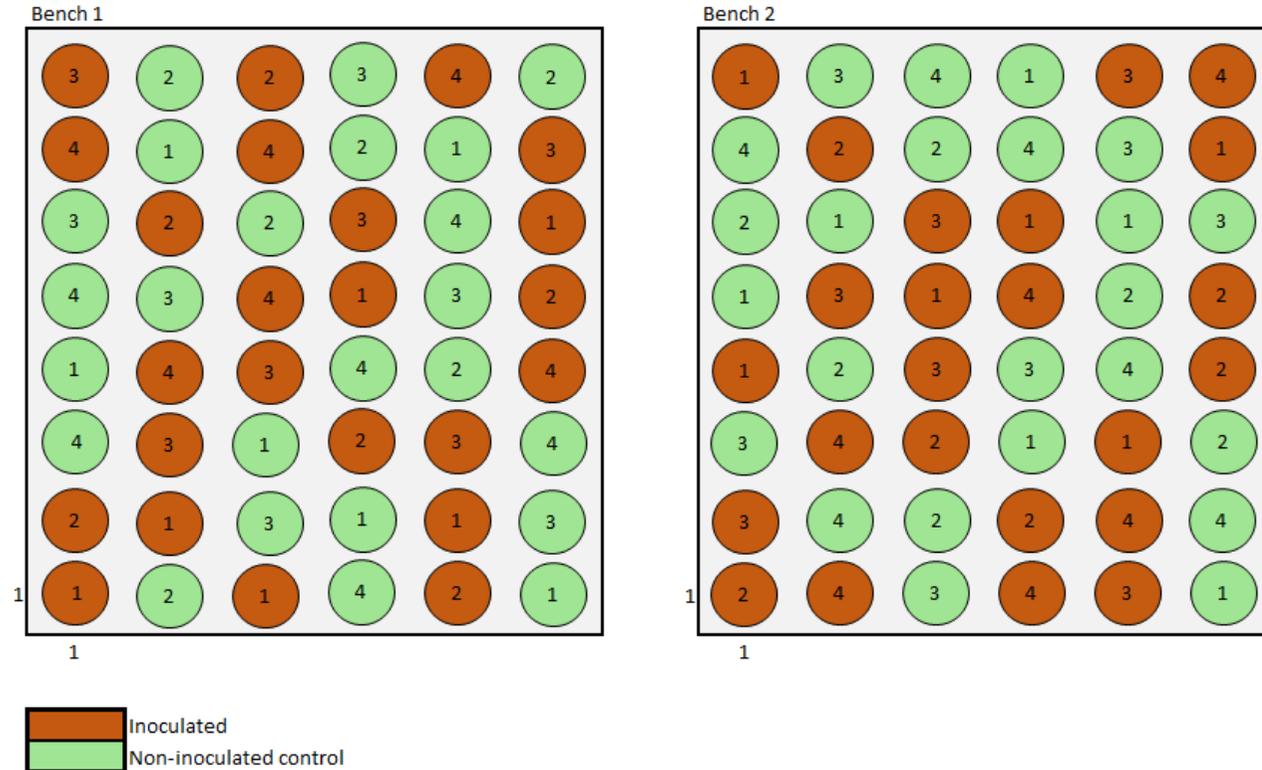
B.2 2ND GLASSHOUSE TRIAL

Layout of the second glasshouse trial at the Leslie Research Centre in 2018. Thirty inoculated (red) and thirty non-inoculated (green) pots of five genotypes (1–5) were placed in a randomised design.
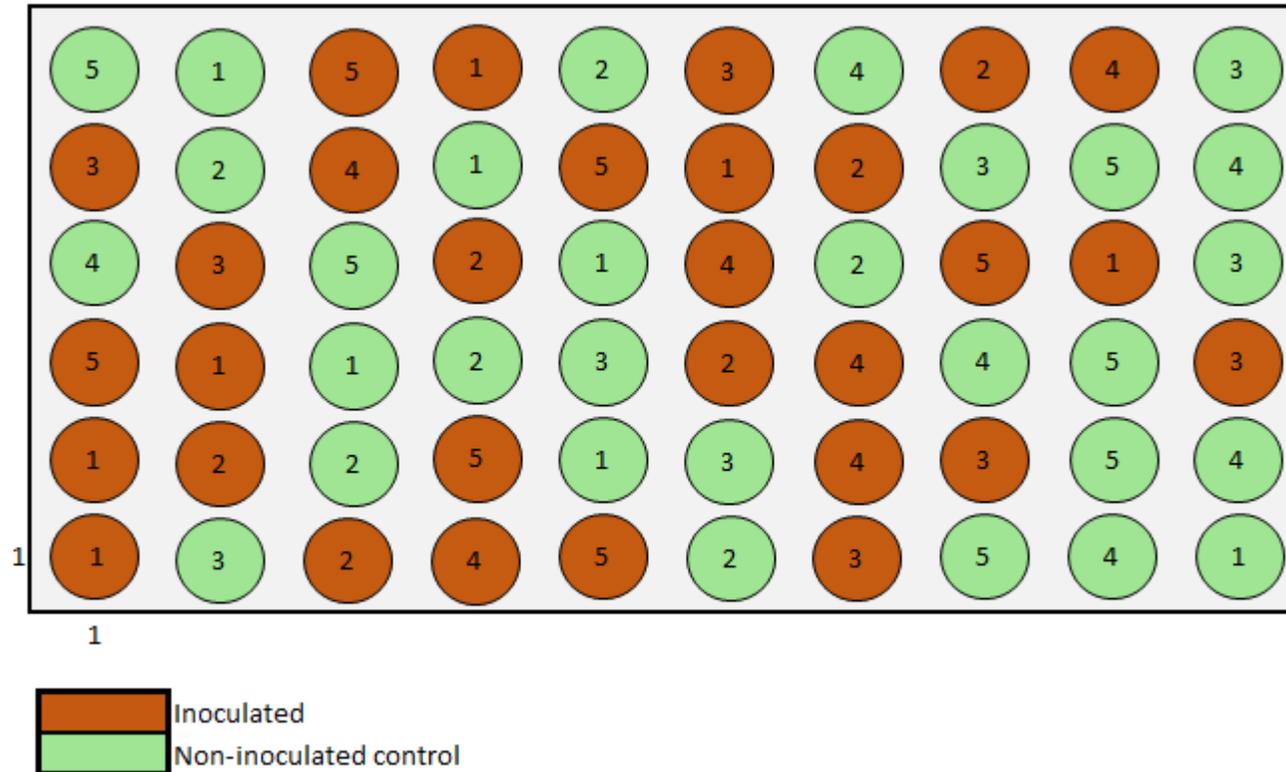
B.3 3ᴿᴰ GLASSHOUSE TRIAL

Layout of the third glasshouse trial at the Centre for Crop Health at USQ in 2019. Thirty inoculated (red) and thirty non-inoculated (green) pots of five genotypes (1–5) were placed in a randomised design.
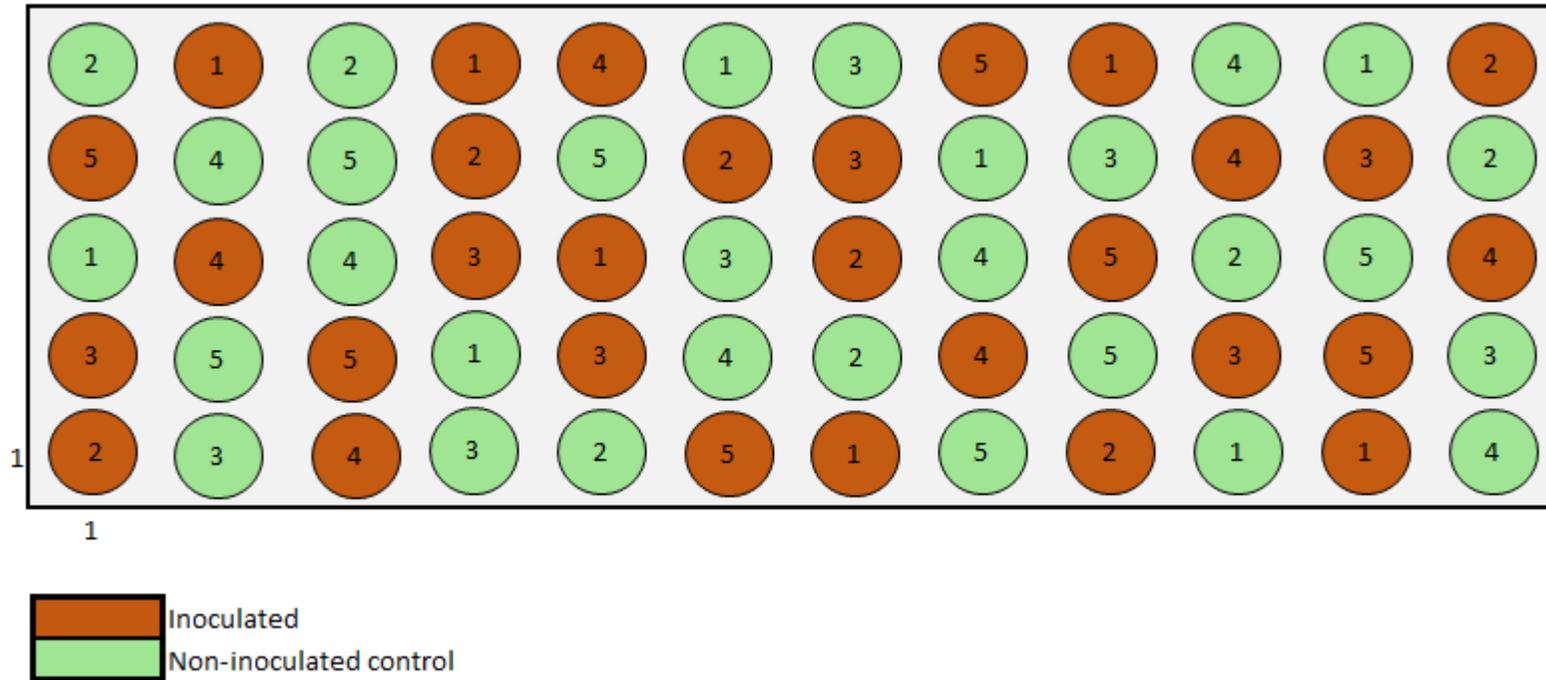
B.4 1ST FIELD TRIAL

Layout of the first field trial at the Tosari Research Station (-27.859964, 151.452766) in 2018. This trial consisted of three replicates across four genotypes (1–4) with inoculated (red) and non-inoculated (green) plots paired and arranged in a randomised strip plot design.

	Replicate	Run	Range 1	Range 2	Range 3	Range 4	Range 5	Range 6	Range 7	Range 8	Range 9	Range 10	Range 11	Range 12
nil	1	1					Genotype 4			Genotype 3			Genotype 1	Genotype 2
plus		2					Genotype 4			Genotype 3			Genotype 1	Genotype 2
nil	2	3			Genotype 3		Genotype 1	Genotype 2			Genotype 4			
plus		4			Genotype 3		Genotype 1	Genotype 2			Genotype 4			
plus	3	5	Genotype 4		Genotype 1	Genotype 2								Genotype 3
nil		6	Genotype 4		Genotype 1	Genotype 2								Genotype 3

<------72 m------>

6 m

<------12m------>

Inoculated	
Non-inoculated control	
Unused	

B.5 2ND FIELD TRIAL

Layout of the second field trial at the Tosari Research Station (-27.859964, 151.452766) in 2019. This trial consisted of three replicates across four genotypes (1–4) with inoculated (red) and non-inoculated (green) plots paired and arranged in a randomised strip plot design.
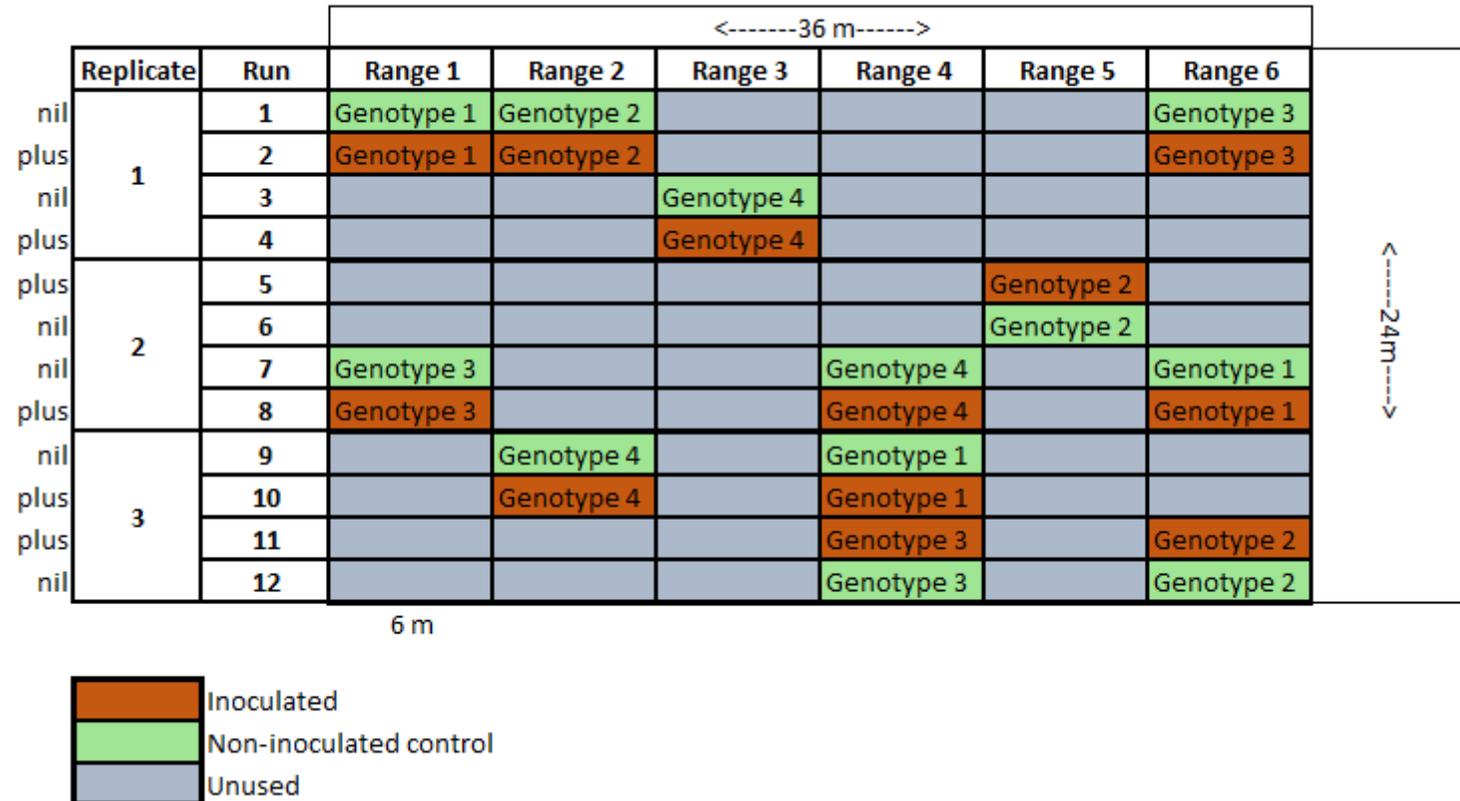
APPENDIX C – SUPPLEMENTARY FIGURES AND TABLES

C.1 Weekly average classification accuracies (%) of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Weeks are weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

Accuracy (20/80 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	week 1	50.63	57.38	48.95	47.95	49.79	47.54	50.21	56.97	49.79	61.07	52.32	48.77
	week 2	60.76	65.23	47.26	49.22	47.26	50.78	59.49	66.41	59.49	67.19	58.65	50.39
	week 3	57.2	61.15	50.8	49.23	50.8	48.08	56	62.69	54.8	61.54	52	48.08
	week 4	57.14	56.98	47.76	49.22	48.57	50.39	53.06	58.91	56.33	58.53	51.02	51.16
	week 5	70.52	55.89	49.8	49.81	62.15	49.81	62.55	65.02	59.76	63.5	60.96	47.91
	week 6	57.2	50.93	49.42	47.96	51.75	48.33	59.14	55.76	56.03	54.28	54.09	48.33
	week 7	56.97	54.92	54.58	56.06	52.99	53.03	54.98	57.58	54.58	56.44	56.97	54.92
	week 8	67.53	63.11	48.92	45.08	48.92	50	60.17	59.43	59.74	57.79	54.98	49.18
	week 9	54.34	60.78	50.68	47.41	46.58	51.29	54.34	60.34	53.88	62.07	52.05	52.16
Glasshouse 2	week 1	99.29	95.14	93.57	90.97	97.86	90.97	90	75	95	84.03	91.43	72.92
	week 2	72.86	78.17	70	81.69	70.71	69.72	72.86	69.01	72.14	71.83	76.43	72.54
	week 3	77.17	73.44	59.78	48.96	63.59	70.83	61.96	69.79	61.41	78.13	61.41	56.25
	week 4	64.54	70.14	61.7	49.31	61.7	56.94	56.03	60.42	54.61	68.06	62.41	54.17
	week 5	86.96	76.39	71.74	56.25	73.19	57.64	64.49	69.44	64.49	72.22	64.49	54.17
	week 6	79.56	75	56.93	53.47	59.85	50.69	62.04	74.31	63.5	67.36	62.04	47.92
	week 7	55.35	53.94	55.35	42.42	55.35	42.42	52.83	58.79	52.2	56.36	55.35	46.06
	week 8	–	–	–	–	–	–	–	–	–	–	–	–
	week 9	78.82	69.44	56.47	50.56	57.65	49.44	56.47	70	58.82	73.89	55.88	52.22

Glasshouse 3	week 1	54.35	52.82	52.17	50	55.07	55.63	57.25	52.11	57.25	56.34	53.62	50
	week 2	92.25	83.45	49.3	48.97	71.83	57.93	66.2	75.86	69.01	84.83	61.97	54.48
	week 3	89.44	86.99	49.3	54.79	60.56	55.48	64.08	78.77	62.68	84.25	70.42	58.9
	week 4	80.82	74.19	65.75	48.39	65.07	59.35	66.44	73.55	71.23	75.48	65.07	58.71
	week 5	65.81	55.76	61.94	50.91	61.94	46.67	62.58	63.03	62.58	59.39	61.94	49.7
	week 6	75.16	87.57	54.04	44.97	53.42	45.56	56.52	76.92	54.66	78.7	62.11	50.3
	week 7	90.59	83.15	57.06	50	57.65	51.12	66.47	81.46	66.47	80.34	64.71	57.87
	week 8	63.28	57.14	48.59	55.49	59.89	53.85	62.71	61.54	60.45	64.84	58.76	53.85
	week 9	73.86	77.6	64.77	52.46	65.91	57.38	63.64	72.13	67.05	72.13	65.34	50.82
Field 1	week 1	60.73	46.8	50.76	48.84	48.94	47.09	56.19	51.16	54.91	51.16	57.7	49.13
	week 2	58.59	51.19	56.75	50.89	57.06	50.89	56.44	55.95	55.79	54.46	56.13	53.87
	week 3	66.46	60.29	55.18	50.14	57.93	47.25	56.1	60.58	53.13	59.42	55.49	51.3
	week 4	61.49	62.32	49.25	48.99	57.31	48.99	53.43	60.87	58.63	60	52.84	51.01
	week 5	67.26	66.76	62.5	59.25	61.9	58.96	55.95	56.07	59.88	63.01	60.12	51.45
	week 6	67.17	67.06	63.53	65.89	62.92	66.47	58.05	56.27	54.17	64.72	59.88	56.85
	week 7	52.08	54.78	50	48.7	50	48.41	54.76	55.94	54.24	55.07	52.08	50.14
	week 8	57.58	50.73	49.39	49.56	56.36	50.73	53.94	52.2	64.2	52.2	53.33	51.03
	week 9	66.27	68.79	49.41	55.49	49.41	54.34	57.69	67.05	56.18	78.03	60.06	59.54
Field 2	week 1	51.95	54.62	51.35	47.98	52.85	50.58	51.35	53.76	50.45	53.47	51.35	49.42
	week 2	55.65	50.29	54.17	49.71	54.46	49.71	56.55	54.05	54.46	50.58	54.17	47.11
	week 3	54.71	54.62	53.5	49.71	54.71	49.13	51.67	55.2	51.98	54.62	50.76	55.78
	week 4	54.73	54.05	52.96	49.71	46.75	48.27	57.1	52.89	55.92	48.27	54.14	52.02
	week 5	61.61	60.98	60.12	53.18	57.44	55.2	58.33	59.83	58.93	65.32	63.99	54.34
	week 6	54.35	50	49.25	50.87	51.95	53.76	54.05	52.02	55.26	52.31	54.05	48.27
	week 7	61.59	54.34	53.05	49.13	51.52	53.76	53.05	56.94	52.74	57.51	52.74	53.76
	week 8	58.21	52.31	53.13	48.55	52.24	49.42	54.93	54.62	56.12	53.47	53.13	48.55
	week 9	53.29	48.84	50	47.69	53.59	48.27	55.09	53.18	52.4	50.29	52.4	48.27
GlasshouseAverage		70.476923	68.334615	56.793462	52.751923	59.617308	54.648846	61.250385	66.355385	61.459615	68.099615	61.016154	53.53
GlasshouseStdError		2.7035649	2.4822135	2.0160741	2.0617744	2.1287472	1.9381706	1.5405505	1.5742022	1.7625924	1.893785	1.7006136	1.3033495
FieldAverage		59.095556	56.042778	53.572222	51.348889	54.296667	51.735	55.259444	56.032222	55.521667	56.883889	55.242222	51.768889

FieldStdError	1.2373721	1.5818511	1.0636162	1.0905149	1.0297074	1.1431991	0.4837425	0.9307887	0.7735755	1.7192047	0.8654213	0.788973
Average	65.820909	63.306136	55.475682	52.177955	57.440682	53.456818	58.799545	62.132273	59.030455	63.511364	58.654091	52.809545
StdError	1.8673394	1.8359848	1.2797205	1.2898887	1.3738488	1.2448518	1.027117	1.2611814	1.1671875	1.5547755	1.1411101	0.8376813
Top GH	18	8	19	7	19	7	6	20	6	20	24	2
%	69.23	30.77	73.08	26.92	73.08	26.92	23.08	76.92	23.08	76.92	92.31	7.69
Top Fld	14	4	14	4	13	5	9	9	8	10	16	2
%	77.78	22.22	77.78	22.22	72.22	27.78	50.00	50.00	44.44	55.56	88.89	11.11
Number of instances as top performer	32	12	33	11	32	12	16	28	14	30	40	4
Percentage of instances as top performer	72.73	27.27	75	25	72.73	27.27	36.36	63.64	31.82	68.18	90.91	9.09

C.2 Weekly average classification accuracies (%) of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Weeks are weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

Accuracy (80/20 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	week 1	63.33	59.02	55	40.98	50	36.07	66.67	65.57	63.33	63.93	63.33	50.82
	week 2	66.67	70.31	68.33	48.44	68.33	60.94	65	73.44	68.33	79.69	68.33	60.94
	week 3	61.9	49.23	50.79	40	57.14	43.08	60.32	58.46	61.9	58.46	55.56	47.69
	week 4	54.84	52.31	53.23	40	53.23	40	50	58.46	48.39	50.77	53.23	33.85
	week 5	68.25	68.18	65.08	45.45	66.67	53.03	57.14	66.67	66.67	75.76	58.73	62.12
	week 6	55.38	66.18	50.77	48.53	50.77	61.76	63.08	55.88	60	63.24	67.69	51.47
	week 7	63.49	46.97	57.14	59.09	55.56	65.15	61.9	62.12	61.9	62.12	58.73	54.55
	week 8	74.14	65.57	50	50.82	53.45	50.82	56.9	63.93	58.62	63.93	58.62	45.9
	week 9	60	60.34	52.73	44.83	49.09	48.28	61.82	62.07	56.36	72.41	56.36	50
Glasshouse 2	week 1	100	100	97.14	100	97.14	97.22	97.14	83.33	97.14	97.22	100	94.44
	week 2	88.57	88.89	91.43	86.11	88.57	83.33	88.57	86.11	88.57	94.44	91.43	77.78
	week 3	69.57	81.25	67.39	68.75	69.57	66.67	67.39	72.92	69.57	81.25	71.74	62.5
	week 4	66.67	77.78	52.78	41.67	52.78	63.89	69.44	75	63.89	69.44	55.56	66.67
	week 5	97.14	97.22	68.57	80.56	77.14	77.78	71.43	72.22	77.14	91.67	82.86	69.44
	week 6	85.71	88.89	62.86	44.44	62.86	69.44	62.86	69.44	62.86	80.56	60	50
	week 7	60	61.9	55	52.38	62.5	45.24	55	57.14	57.5	69.05	65	38.1
	week 8	–	–	–	–	–	–	–	–	–	–	–	–
	week 9	81.4	77.78	58.14	53.33	58.14	55.56	67.44	66.67	65.12	75.56	53.49	55.56
Glasshouse 3	week 1	65.71	75	54.29	63.89	51.43	66.67	54.29	75	54.29	72.22	54.29	50
	week 2	97.22	86.49	72.22	81.08	72.22	89.19	72.22	72.97	66.67	89.19	66.67	62.16
	week 3	97.22	89.19	66.67	67.57	69.44	70.27	63.89	86.49	66.67	83.78	66.67	48.65
	week 4	81.08	87.18	56.76	76.92	56.76	79.49	78.38	79.49	67.57	92.31	62.16	76.92
	week 5	79.49	73.81	53.85	47.62	53.85	57.14	66.67	69.05	56.41	69.05	66.67	66.67

	week 6	90.24	93.02	63.41	55.81	65.85	79.07	73.17	86.05	63.41	93.02	65.85	55.81
	week 7	97.67	97.78	74.42	64.44	74.42	75.56	81.4	86.67	88.37	95.56	83.72	66.67
	week 8	57.78	69.57	55.56	54.35	46.67	52.17	68.89	67.39	66.67	71.74	60	52.17
	week 9	84.09	80.43	72.73	60.87	75	78.26	70.45	78.26	77.27	82.61	75	67.39
	week 1	67.47	61.63	43.37	44.19	44.58	46.51	59.04	61.63	57.83	61.63	60.24	54.65
	week 2	65.85	67.86	54.88	44.05	54.88	57.14	62.2	55.95	65.85	64.29	63.41	54.76
	week 3	69.51	71.26	68.29	43.68	65.85	65.52	58.54	67.82	71.95	77.01	62.2	59.77
	week 4	59.52	79.31	52.38	50.57	58.33	64.37	65.48	64.37	61.9	80.46	61.9	59.77
Field 1	week 5	72.62	67.82	67.86	67.82	66.67	68.97	69.05	63.22	72.62	70.11	69.05	56.32
	week 6	72.29	70.93	62.65	70.93	61.45	70.93	61.45	72.09	61.45	72.09	68.67	62.79
	week 7	57.14	54.02	46.43	43.68	63.1	40.23	61.9	59.77	54.76	65.52	60.71	64.37
	week 8	60.24	58.14	53.01	46.51	57.83	55.81	50.6	61.63	51.81	56.98	54.22	50
	week 9	67.06	75.86	47.06	42.53	52.94	60.92	63.53	78.16	58.82	82.76	57.65	56.32
	week 1	59.52	51.72	53.57	45.98	48.81	47.13	51.19	50.57	46.43	57.47	46.43	45.98
	week 2	60.71	54.02	45.24	47.13	57.14	48.28	59.52	52.87	60.71	45.98	57.14	50.57
	week 3	56.63	59.77	51.81	49.43	54.22	58.62	59.04	60.92	60.24	65.52	54.22	50.57
	week 4	63.53	62.07	58.82	57.47	55.29	59.77	56.47	64.37	57.65	57.47	52.94	42.53
Field 2	week 5	67.86	65.52	67.86	63.22	61.9	58.62	63.1	58.62	71.43	66.67	66.67	57.47
	week 6	63.1	50.57	64.29	51.72	65.48	48.28	58.33	59.77	55.95	58.62	55.95	48.28
	week 7	56.1	59.77	56.1	45.98	58.54	50.57	62.2	60.92	65.85	59.77	58.54	47.13
	week 8	64.29	59.77	54.76	47.13	57.14	52.87	61.9	62.07	67.86	64.37	64.29	55.17
	week 9	54.76	57.47	55.95	47.13	53.57	45.98	52.38	52.87	59.52	54.02	58.33	52.87
GlasshouseAverage		75.6754	75.5496	62.5496	58.3819	63.0223	64.08	67.3638	71.1846	66.7162	76.8838	66.2188	58.395
GlasshouseStdError		2.93404	2.98167	2.34953	3.13142	2.46617	3.07583	2.05743	1.87834	2.18284	2.50239	2.33566	2.55492
FieldAverage		63.2333	62.6394	55.7961	50.5083	57.6511	55.5844	59.7733	61.5344	61.2572	64.4856	59.5867	53.8511
FieldStdError		1.30098	1.93407	1.8503	2.02855	1.38471	2.04404	1.13699	1.57968	1.66441	2.22433	1.36884	1.39181
Average		70.5855	70.2682	59.7868	55.1609	60.825	60.6045	64.2586	67.2368	64.483	71.8118	63.5057	56.5361
StdError		2.02522	2.14389	1.64518	2.09353	1.60056	2.08128	1.40969	1.46261	1.50084	1.95337	1.55708	1.63461

APPENDIX C. SUPPLEMENTARY FIGURES AND TABLES

Top GH	11	14	17	9	11	15	7	19	1	25	21	4
%	44.00	56.00	65.38	34.62	42.31	57.69	26.92	73.08	3.85	96.15	84.00	16.00
Top Fld	11	7	15	3	10	8	8	10	8	10	17	1
%	61.11	38.89	83.33	16.67	55.56	44.44	44.44	55.56	44.44	55.56	94.44	5.56
Number of instances as top performer	22	21	32	12	21	23	15	29	9	35	39	5
Percentage of instances as top performer	50	47.73	72.73	27.27	47.73	52.27	34.09	65.91	20.45	79.55	88.64	11.36

C.3 Weekly average classification F_1 scores of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Weeks are weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

F_1 (20/80 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	week 1	0.6319	0.5873	0.6572	0.6482	0.6572	0.5646	0.5138	0.585	0.5122	0.6215	0.5857	0.3455
	week 2	0.6116	0.6307	0.6418	0.058	0.6418	0.137	0.5761	0.6446	0.5826	0.6038	0.5902	0.0305
	week 3	0.5304	0.5809	0	0.0149	0	0.4	0.5455	0.6008	0.4959	0.5726	0.4495	0.3541
	week 4	0.6464	0.5153	0.6464	0	0.604	0.0588	0.5207	0.5748	0.5158	0.5198	0.4167	0.3298
	week 5	0.7154	0.5538	0	0	0.52	0	0.625	0.6	0.5944	0.5897	0.5487	0.417
	week 6	0.5714	0.472	0	0.0141	0.1389	0.1146	0.5783	0.5448	0.5232	0.4434	0.4636	0.393
	week 7	0.7194	0.5939	0.6833	0.6963	0.5083	0.6242	0.5923	0.6242	0.5581	0.5735	0.605	0.5735
	week 8	0.6835	0.6484	0.657	0.4463	0.657	0.4786	0.6025	0.6148	0.5992	0.5992	0.5772	0.5231
	week 9	0.4898	0.6353	0.0806	0.6433	0.156	0.5498	0.5146	0.6226	0.5073	0.6179	0.4324	0.5316
Glasshouse 2	week 1	0.9929	0.9517	0.9333	0.9065	0.9787	0.9128	0.9054	0.7778	0.9504	0.8553	0.9155	0.7578
	week 2	0.7532	0.8098	0.7407	0.8219	0.7394	0.7425	0.7397	0.75	0.7417	0.7647	0.7442	0.7194
	week 3	0.7742	0.732	0.3273	0	0.4724	0.6667	0.6237	0.7264	0.6011	0.7813	0.6034	0.4615
	week 4	0.6795	0.6993	0.6635	0.6605	0.6584	0.6556	0.6	0.6588	0.5949	0.6974	0.6667	0.5147
	week 5	0.8676	0.7952	0.7111	0.5714	0.7338	0.6164	0.6475	0.7284	0.6423	0.726	0.6667	0.5417
	week 6	0.8079	0.76	0.6845	0.5442	0.6995	0.5235	0.6452	0.7299	0.6575	0.6713	0.6977	0.5033
	week 7	0.6479	0.5096	0.3717	0.3709	0.3932	0.3709	0.4722	0.5062	0.4234	0.52	0.3717	0.3597
	week 8	–	–	–	–	–	–	–	–	–	–	–	–
	week 9	0.7857	0.7027	0.2745	0	0.4559	0.3893	0.5132	0.7404	0.4615	0.7539	0.5763	0.5114
Glasshouse 3	week 1	0.6634	0.5442	0.6207	0.6667	0.6448	0.6802	0.5693	0.5526	0.5693	0.5694	0.6	0.4741
	week 2	0.922	0.8286	0.6604	0.6509	0.726	0.5793	0.6795	0.7482	0.7027	0.8514	0.6405	0.6024
	week 3	0.8993	0.879	0.6604	0.5714	0.6984	0.6012	0.6871	0.805	0.6624	0.8516	0.72	0.6154
	week 4	0.8228	0.7368	0.6875	0.3846	0.6826	0.5882	0.6622	0.7389	0.7	0.7467	0.6667	0.5362
	week 5	0.6878	0.6054	0.6774	0.6747	0.6878	0.5111	0.6588	0.6474	0.6548	0.5868	0.655	0.3852

	week 6	0.7468	0.8852	0.6494	0	0.5486	0.4713	0.6054	0.7914	0.535	0.8043	0.6347	0.494
	week 7	0.9149	0.8315	0.4966	0.2992	0.5	0.3741	0.7107	0.8308	0.6816	0.8043	0.6471	0.5509
	week 8	0.5963	0.5761	0.0215	0.449	0.5799	0.4878	0.6071	0.6277	0.5679	0.6061	0.573	0.3636
	week 9	0.7565	0.7545	0.6667	0.2435	0.6875	0.3906	0.6564	0.7059	0.7071	0.6829	0.663	0.3662
Field 1	week 1	0.5808	0.4246	0.6733	0	0.5806	0.3893	0.5682	0.4898	0.5739	0.4167	0.5905	0.4807
	week 2	0.5574	0.4938	0.5913	0.012	0.5758	0.1951	0.5605	0.5287	0.4863	0.508	0.4604	0.3621
	week 3	0.6605	0.5935	0.3524	0.14	0.4982	0.3893	0.5355	0.5869	0.5085	0.5783	0.5466	0.4474
	week 4	0.6786	0.5724	0	0	0.3182	0	0.5593	0.582	0.5577	0.5071	0.538	0.3475
	week 5	0.6913	0.6724	0.6519	0.5662	0.6049	0.5697	0.5819	0.592	0.5801	0.6364	0.5855	0.5116
	week 6	0.6625	0.6781	0.6571	0.6667	0.6192	0.6547	0.5747	0.5427	0.6106	0.6277	0.6189	0.5595
	week 7	0.4889	0.5329	0	0.4197	0.3058	0.4258	0.5607	0.5658	0.5284	0.5231	0.5091	0.4267
	week 8	0.5745	0.5177	0.0234	0.3857	0.5034	0.4437	0.5543	0.5534	0.5838	0.5119	0.5815	0.5045
	week 9	0.6686	0.6897	0	0.5698	0.2785	0.5183	0.56	0.6812	0.6553	0.7829	0.6266	0.6089
Field 2	week 1	0.5376	0.5651	0.4698	0.3431	0.5257	0.4537	0.5179	0.5413	0.4878	0.5106	0.4894	0.5042
	week 2	0.5842	0.5165	0.5523	0	0.5487	0.4882	0.5799	0.5574	0.5345	0.5128	0.5576	0.4986
	week 3	0.6761	0.5722	0.3139	0	0.4792	0.4465	0.5402	0.5678	0.4626	0.5476	0.4675	0.5714
	week 4	0.5667	0.5364	0.5293	0	0.5135	0.4207	0.565	0.5389	0.5635	0.4241	0.5245	0.4155
	week 5	0.6455	0.6301	0.6398	0.5	0.5903	0.5373	0.5765	0.6032	0.566	0.6703	0.6472	0.5707
	week 6	0.6667	0.4971	0.66	0.4753	0.6667	0.5238	0.5916	0.5493	0.563	0.4665	0.5073	0.4053
	week 7	0.6527	0.4698	0.6486	0.1852	0.615	0.4558	0.5746	0.613	0.5284	0.5836	0.492	0.3939
	week 8	0.5833	0.5385	0.5602	0.4702	0.5092	0.4928	0.5354	0.5423	0.5638	0.5091	0.5399	0.4258
	week 9	0.6667	0.462	0	0.3221	0.4444	0.4207	0.5101	0.5272	0.4897	0.4671	0.3465	0.4781
GlasshouseAverage		0.727635	0.685353846	0.508212	0.397558	0.568081	0.48035	0.617392	0.672208	0.605473	0.6698	0.604277	0.471369
GlasshouseStdError		0.024747	0.025723811	0.054187	0.058892	0.041194	0.042197	0.017411	0.017397	0.021179	0.022596	0.0225	0.028134
FieldAverage		0.619033	0.553488889	0.40685	0.280889	0.50985	0.434744	0.558128	0.564606	0.546883	0.543544	0.534944	0.472911
FieldStdError		0.014062	0.017973513	0.0649	0.056215	0.026293	0.033835	0.00527	0.009879	0.011456	0.021485	0.016922	0.017767
Average		0.683207	0.631409091	0.466745	0.34983	0.544259	0.461693	0.593148	0.628189	0.581505	0.61815	0.575914	0.472
StdError		0.017569	0.019425085	0.041806	0.04218	0.026718	0.028435	0.011331	0.013596	0.013951	0.01843	0.015724	0.017972

Number of instances as top performer	34	10	30	12	35	9	12	32	15	29	40	4
Percentage of instances as top performer	77.27	22.73	68.18	27.27	79.55	20.45	27.27	72.73	34.09	65.91	90.91	9.09

C.4 Weekly average classification F_1 scores of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Weeks are weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

F_1 (80/20 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	3 week	0.6517	0.5763	0.3415	0	0.36	0	0.6	0.6667	0.5769	0.6333	0.4762	0.375
	6 week	0.5833	0.7246	0.5778	0.0571	0.5581	0.4681	0.6441	0.7463	0.6316	0.806	0.6545	0.6032
	9 week	0.5862	0.5479	0	0	0.4255	0.1395	0.6087	0.6316	0.6	0.5846	0.5172	0.4848
	3 week	0.5758	0.5676	0.0645	0	0.3556	0.093	0.4918	0.6111	0.5	0.4138	0.4561	0.2456
	6 week	0.7436	0.6769	0.725	0	0.7042	0.4151	0.6494	0.6933	0.72	0.7714	0.5517	0.6269
	9 week	0.4848	0.6349	0.1579	0	0.2727	0.4091	0.6842	0.5946	0.597	0.6154	0.6038	0.3529
	3 week	0.7089	0.55	0.7273	0.7327	0.7083	0.7356	0.6571	0.6914	0.6479	0.6753	0.64	0.4643
	6 week	0.717	0.6441	0.6353	0.375	0.6216	0.4	0.5614	0.6571	0.5556	0.6944	0.6	0.3529
	9 week	0.5926	0.5965	0.5867	0.1111	0.5672	0.2857	0.6038	0.56	0.5385	0.625	0.4706	0.3556
Glasshouse 2	3 week	1	1	0.9714	1	0.9714	0.9655	0.9697	0.8333	0.9714	0.9677	1	0.9286
	6 week	0.8889	0.9048	0.9143	0.8649	0.8889	0.8421	0.8947	0.878	0.8824	0.95	0.9189	0.7895
	9 week	0.6667	0.8163	0.4828	0.6154	0.5625	0.619	0.6512	0.7347	0.6818	0.8163	0.6341	0.5714
	3 week	0.625	0.75	0.5641	0.5882	0.5143	0.6667	0.6667	0.6667	0.6486	0.7027	0.6	0.6
	6 week	0.9714	0.9677	0.5926	0.7879	0.7143	0.7778	0.7059	0.7222	0.7647	0.9091	0.8125	0.6207
	9 week	0.8571	0.875	0.7234	0.5455	0.6977	0.6857	0.6286	0.625	0.5946	0.7879	0.6111	0.25
	3 week	0.6531	0.6364	0	0.1667	0.4444	0.439	0.5	0.5909	0.5	0.6977	0.65	0.0714
	6 week	–	–	–	–	–	–	–	–	–	–	–	–
	9 week	0.7647	0.7727	0.2759	0.087	0.5263	0.5	0.6111	0.6939	0.5455	0.7556	0.4	0.4444
Glasshouse 3	3 week	0.6538	0.7568	0.6364	0.6486	0.625	0.6842	0.6	0.7429	0.619	0.7222	0.4737	0.4375
	6 week	0.9744	0.8571	0.7059	0.7742	0.7059	0.8889	0.75	0.7222	0.7	0.8889	0.6842	0.6316
	9 week	0.9744	0.8824	0.6842	0.6471	0.7179	0.6857	0.6667	0.8649	0.6667	0.8333	0.6842	0.24
	3 week	0.8108	0.8571	0.5789	0.6667	0.5789	0.75	0.7778	0.8	0.6667	0.9143	0.65	0.7429
	6 week	0.75	0.7179	0.625	0.5769	0.6087	0.6538	0.6977	0.6829	0.5854	0.6667	0.6667	0.5625

	9 week	0.8889	0.9231	0.6512	0.6275	0.6667	0.7273	0.6857	0.8571	0.6341	0.9231	0.65	0.3448
	3 week	0.973	0.9796	0.7755	0.5789	0.7442	0.7442	0.7895	0.875	0.8718	0.96	0.8205	0.6939
	6 week	0.6122	0.6818	0.6667	0.5532	0.5	0.5217	0.6667	0.6667	0.6667	0.6829	0.5909	0.3125
	9 week	0.8293	0.7568	0.7273	0.6087	0.7442	0.7619	0.6818	0.7619	0.7619	0.7778	0.6667	0.6512
	3 week	0.6494	0.6024	0.605	0.6129	0.5741	0.54	0.55	0.5714	0.5714	0.5075	0.5143	0.5185
	6 week	0.6585	0.6667	0.5432	0.6116	0.5432	0.5909	0.6024	0.5647	0.6	0.6429	0.5833	0.4722
	9 week	0.6479	0.6914	0.6076	0.608	0.6216	0.6939	0.5	0.641	0.6349	0.7561	0.5455	0.4444
	3 week	0.6154	0.775	0.6429	0.6325	0.6316	0.6667	0.6265	0.6364	0.6024	0.8	0.5647	0.5455
Field 1	6 week	0.7333	0.6667	0.6897	0.6818	0.6744	0.6897	0.6757	0.5897	0.6944	0.6667	0.675	0.5778
	9 week	0.7013	0.6988	0.6353	0.6988	0.6279	0.6988	0.6024	0.7073	0.5676	0.6923	0.6486	0.5676
	3 week	0.5714	0.5238	0.6341	0.608	0.6517	0.48	0.5556	0.5977	0.525	0.6341	0.56	0.4746
	6 week	0.6292	0.5814	0.6296	0.6349	0.5882	0.6042	0.5057	0.6374	0.4941	0.6022	0.5128	0.3175
	9 week	0.6818	0.7529	0.6034	0.5968	0.6019	0.6047	0.6173	0.7654	0.5882	0.8193	0.5316	0.6042
	3 week	0.6327	0.4884	0.5806	0	0.4819	0.4773	0.5393	0.5	0.5055	0.5287	0.4598	0.3562
	6 week	0.7077	0.5556	0	0	0.5135	0.3284	0.6531	0.5287	0.6118	0.4598	0.5135	0.4941
	9 week	0.7049	0.5647	0.4286	0.2414	0.4865	0.5263	0.6136	0.6222	0.6374	0.6512	0.5682	0.4416
	3 week	0.6353	0.5823	0.4444	0.5195	0.5366	0.5679	0.6022	0.6667	0.5909	0.5843	0.5	0.2647
Field 2	6 week	0.7255	0.6737	0.7523	0.7091	0.6667	0.5714	0.6353	0.6	0.7333	0.6882	0.6957	0.5934
	9 week	0.6737	0.5376	0.7	0.5714	0.6882	0.4828	0.6022	0.6327	0.5934	0.5814	0.5195	0.4706
	3 week	0.7188	0.6237	0.5909	0.5053	0.6136	0.5057	0.6667	0.5854	0.6818	0.5977	0.5405	0.4524
	6 week	0.6809	0.6602	0.5778	0	0.561	0.5393	0.5854	0.6207	0.64	0.6593	0.5783	0.5185
	9 week	0.672	0.5934	0.4638	0	0.5349	0.4598	0.5238	0.5287	0.5854	0.5	0.5934	0.4938
GlasshouseAverage		0.751446	0.755934615	0.553523	0.446665	0.607096	0.571523	0.670935	0.714246	0.6588	0.760592	0.633985	0.490542
GlasshouseStdError		0.02974	0.027557094	0.051042	0.062681	0.031481	0.048322	0.020473	0.018107	0.02274	0.026629	0.027172	0.038993
FieldAverage		0.668872	0.624372222	0.562733	0.457333	0.58875	0.5571	0.592067	0.610894	0.603194	0.631761	0.561372	0.4782
FieldStdError		0.010122	0.018613957	0.038678	0.063903	0.014986	0.022733	0.012444	0.015139	0.014649	0.023437	0.01454	0.021759
Average		0.717666	0.702113636	0.557291	0.45103	0.599591	0.565623	0.63867	0.671966	0.636052	0.707889	0.60428	0.485493
StdError		0.018941	0.020352726	0.033727	0.044843	0.01946	0.029782	0.014283	0.014483	0.015157	0.020634	0.017818	0.024491

Number of instances as top performer	24	17	26	11	19	20	12	29	10	33	38	5
Percentage of instances as top performer	54.55	38.64	59.09	25	43.18	45.45	27.27	65.91	22.73	75	86.36	11.36

C.5 3-week average accuracy (%) scores of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Accuracies for the 3-week average groups were calculated based on the average of the first three, six or nine weeks and described as weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

Accuracy (20/80 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	Average	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	3 week	56.2	61.26	49	48.8	49.28	48.8	55.23	62.02	54.69	63.26	54.32	49.08
	6 week	58.91	57.93	49	48.9	51.72	49.15	56.74	60.96	56.03	61.02	54.84	49.11
	9 week	59.14	58.49	49.8	49.11	50.98	49.92	56.66	60.35	56.04	60.27	54.78	50.1
Glasshouse 2	3 week	83.11	82.25	74.45	73.87	77.39	77.17	74.94	71.27	76.19	77.99	76.42	67.23
	6 week	80.06	78.05	68.95	63.44	71.15	66.13	67.9	69.66	68.53	73.6	69.7	59.66
	9 week	76.82	73.96	65.69	59.2	67.49	61.08	64.58	68.35	65.27	71.48	66.18	57.03
Glasshouse 3	3 week	78.68	74.42	50.26	51.25	62.49	56.35	62.51	68.91	62.98	75.14	62.01	54.46
	6 week	76.3	73.46	55.42	49.67	61.31	53.44	62.18	70.04	62.9	73.16	62.52	53.68
	9 week	76.17	73.18	55.88	50.66	61.26	53.66	62.88	70.6	63.49	72.92	62.66	53.85
Field 1	3 week	61.93	52.76	54.23	49.96	54.64	48.41	56.24	55.9	54.61	55.02	56.44	51.43
	6 week	63.62	59.07	56.33	54	57.68	53.27	56.03	56.82	56.09	58.8	57.03	52.27
	9 week	61.96	58.75	54.09	53.08	55.76	52.57	55.84	57.34	56.79	59.79	56.4	52.7
Field 2	3 week	54.11	53.18	53	49.13	54.01	49.81	53.19	54.34	52.3	52.89	52.09	50.77
	6 week	55.5	54.09	53.56	50.19	53.03	51.11	54.84	54.62	54.5	54.09	54.74	51.16
	9 week	56.23	53.34	53.06	49.61	52.84	50.9	54.68	54.72	54.25	53.98	54.08	50.83
Average	3 week	66.806	64.774	56.188	54.602	59.562	56.108	60.422	62.488	60.154	64.86	60.256	54.594
	6 week	66.878	64.52	56.652	53.24	58.978	54.62	59.538	62.42	59.61	64.134	59.766	53.176
	9 week	66.19967	63.70667	55.862	52.48333	57.88467	53.79167	59.02967	62.29667	59.24167	63.76233	58.97767	52.94767
SE	3 week	5.934068	5.871802	4.659969	4.835524	4.935489	5.458925	3.948923	3.379174	4.401228	5.102655	4.363673	3.276618
	6 week	4.827534	4.716461	3.323973	2.697226	3.488108	2.98275	2.440382	3.200106	2.661855	3.937666	2.857278	1.784849
	9 week	4.354968	4.207092	2.685183	1.848176	3.008767	1.972865	2.003603	3.09313	2.185636	3.654227	2.382864	1.227279
TopPerformer	3 week	4	1	4	1	5	0	2	3	0	5	5	0

APPENDIX C. SUPPLEMENTARY FIGURES AND TABLES

	6 week	5	0	5	0	5	0	1	4	1	4	5	0
	9 week	5	0	5	0	5	0	0	5	1	4	5	0
%	3 week	80	20	80	20	100	0	40	60	0	100	100	0
	6 week	100	0	100	0	100	0	20	80	20	80	100	0
	9 week	100	0	100	0	100	0	0	100	20	80	100	0
Top GH		8	1	8	1	9	0	1	8	0	9	9	0
%		88.89	11.11	88.89	11.11	100.00	0.00	11.11	88.89	0.00	100.00	100.00	0.00
Top Fld		6	0	6	0	6	0	2	4	2	4	6	0
%		100.00	0.00	100.00	0.00	100.00	0.00	33.33	66.67	33.33	66.67	100.00	0.00
TopPerformer		14	1	14	1	15	0	3	12	2	13	15	0
%		93.33	6.67	93.33	6.67	100.00	0.00	20.00	80.00	13.33	86.67	100.00	0.00
Average		66.58267	64.27933	56.18133	53.39133	58.73533	54.78467	59.62933	62.39333	59.644	64.22733	59.614	53.55733
StdDev		10.17091	9.975538	7.310875	6.804487	7.83582	7.604303	5.870379	6.453221	6.465832	8.568986	6.648323	4.595725

C.6 3-week average accuracy (%) scores of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Accuracies for the 3-week average groups were calculated based on the average of the first three, six or nine weeks and described as weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

Accuracy (80/20 split)	Average	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	3 week	63.97	59.52	58.04	43.14	58.49	46.69	63.99	65.82	64.52	67.36	62.41	53.15
	6 week	61.73	60.87	57.2	43.9	57.69	49.15	60.37	63.08	61.44	65.31	61.15	51.15
	9 week	63.11	59.79	55.9	46.46	56.03	51.01	60.31	62.96	60.61	65.59	60.06	50.81
Glasshouse 2	3 week	86.05	90.05	85.32	84.95	85.09	82.41	84.37	80.79	85.09	90.97	87.72	78.24
	6 week	84.61	89	73.36	70.25	74.68	76.39	76.14	76.5	76.53	85.76	76.93	70.14
	9 week	81.13	84.21	69.16	65.91	71.09	69.89	72.41	72.85	72.72	82.4	72.51	64.31
Glasshouse 3	3 week	86.72	83.56	64.39	70.85	64.37	75.38	63.47	78.15	62.54	81.73	62.54	53.6
	6 week	85.16	84.11	61.2	65.48	61.59	73.64	68.1	78.17	62.5	83.26	63.72	60.04
	9 week	83.39	83.61	63.32	63.62	62.85	71.98	69.93	77.93	67.48	83.28	66.78	60.72
Field 1	3 week	67.61	66.92	55.51	43.97	55.1	56.39	59.92	61.8	65.21	67.64	61.95	56.39
	6 week	67.88	69.8	58.24	53.54	58.63	62.24	62.62	64.18	65.27	70.93	64.25	58.01
	9 week	65.75	67.43	55.1	50.44	58.4	58.93	61.31	64.96	61.89	70.09	62.01	57.64
Field 2	3 week	58.95	55.17	50.21	47.51	53.39	51.34	56.58	54.79	55.79	56.32	52.6	49.04
	6 week	61.89	57.28	56.93	52.49	57.14	53.45	57.94	57.85	58.74	58.62	55.56	49.23
	9 week	60.72	57.85	56.49	50.57	56.9	52.23	58.24	58.11	60.63	58.88	57.17	50.06
Average	3 week	72.66	71.044	62.694	58.084	63.288	62.442	65.666	68.27	66.63	72.804	65.444	58.084
	6 week	72.25	72.21	61.39	57.13	61.95	62.97	65.03	67.96	64.90	72.78	64.32	57.71
	9 week	70.82	70.578	59.994	55.4	61.054	60.808	64.44	67.362	64.666	72.048	63.706	56.708
SE	3 week	5.77	6.78	6.10	8.42	5.76	6.98	4.86	4.92	4.91	6.07	5.88	5.17
	6 week	5.275104	6.24829	3.087775	4.750303	3.274817	5.367033	3.245371	3.984285	3.090881	5.186712	3.508301	3.707882
	9 week	4.75	5.67	2.72	3.91	2.77	4.36	2.82	3.55	2.38	4.76	2.70	2.77
TopPerformer	3 week	4.00	1.00	4.00	1.00	3.00	2.00	2.00	3.00	0.00	5.00	5.00	0.00

	6 week	3.00	2.00	4.00	1.00	2.00	3.00	1.00	4.00	1.00	4.00	5.00	0.00
	9 week	2.00	3.00	4.00	1.00	3.00	2.00	1.00	4.00	1.00	4.00	5.00	0.00
%	3 week	80.00	20.00	80.00	20.00	60.00	40.00	40.00	60.00	0.00	100.00	100.00	0.00
	6 week	60.00	40.00	80.00	20.00	40.00	60.00	20.00	80.00	20.00	80.00	100.00	0.00
	9 week	40.00	60.00	80.00	20.00	60.00	40.00	20.00	80.00	20.00	80.00	100.00	0.00
	Top GH	5	4	6	3	5	4	1	8	0	9	9	0
	%	55.56	44.44	66.67	33.33	55.56	44.44	11.11	88.89	0.00	100.00	100.00	0.00
	Top Fld	4	2	6	0	3	3	3	3	2	4	6	0
	%	66.67	33.33	100.00	0.00	50.00	50.00	50.00	50.00	33.33	66.67	100.00	0.00
	TopPerformer	9	6	12	3	8	7	4	11	2	13	15	0
	%	60.00	40.00	80.00	20.00	53.33	46.67	26.67	73.33	13.33	86.67	100.00	0.00
	Average	46.10857	41.30944	44.61686	30.32707	39.0778	38.9983	35.28035	46.30284	32.53662	51.15615	49.88995	26.11116
	StdDev	27.57576	26.91709	22.22229	21.12785	22.45339	23.57414	23.91928	25.93458	24.00637	27.32994	23.35652	20.99061

C.7 3-week average F_1 scores of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Accuracies for the 3-week average groups were calculated based on the average of the first three, six or nine weeks and described as weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

F_1 (20/80 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	Average	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	3 week	0.5913	0.5996	0.433	0.2404	0.433	0.3672	0.5451	0.6101	0.5302	0.5993	0.5418	0.2434
	6 week	0.6179	0.5567	0.3242	0.1225	0.427	0.2125	0.5599	0.5917	0.5373	0.5585	0.5091	0.3117
	9 week	0.6222	0.5797	0.374	0.2801	0.4315	0.3253	0.5632	0.6013	0.5432	0.5713	0.5188	0.3887
Glasshouse 2	3 week	0.8401	0.8312	0.6671	0.5761	0.7302	0.774	0.7563	0.7514	0.7644	0.8004	0.7543	0.6462
	6 week	0.8126	0.7913	0.6767	0.5841	0.7137	0.6862	0.6936	0.7286	0.698	0.7493	0.7157	0.5831
	9 week	0.7886	0.745	0.5883	0.4844	0.6414	0.6097	0.6434	0.7022	0.6341	0.7212	0.6552	0.5462
Glasshouse 3	3 week	0.8282	0.7506	0.6471	0.6297	0.6897	0.6202	0.6453	0.702	0.6448	0.7575	0.6535	0.564
	6 week	0.7904	0.7465	0.6593	0.4914	0.6647	0.5719	0.6437	0.7139	0.6374	0.735	0.6528	0.5179
	9 week	0.7789	0.7379	0.5712	0.4378	0.6395	0.5204	0.6485	0.7164	0.6423	0.7226	0.6444	0.4876
Field 1	3 week	0.5996	0.504	0.539	0.0507	0.5515	0.3245	0.5547	0.5351	0.5229	0.501	0.5325	0.4301
	6 week	0.6385	0.5725	0.4877	0.2308	0.5328	0.3663	0.5633	0.5537	0.5528	0.5457	0.5567	0.4515
	9 week	0.6181	0.575	0.3277	0.3067	0.4761	0.3984	0.5617	0.5692	0.565	0.5658	0.5619	0.4721
Field 2	3 week	0.5993	0.5513	0.4454	0.1144	0.5178	0.4628	0.546	0.5555	0.495	0.5237	0.5049	0.5248
	6 week	0.6128	0.5529	0.5275	0.2197	0.554	0.4784	0.5618	0.5596	0.5296	0.522	0.5323	0.4943
	9 week	0.6199	0.532	0.486	0.2551	0.5436	0.4711	0.5546	0.56	0.5288	0.5213	0.508	0.4737
Average	3 week	0.6917	0.64734	0.54632	0.32226	0.58444	0.50974	0.60948	0.63082	0.59146	0.63638	0.5974	0.4817
	6 week	0.69444	0.64398	0.53508	0.3297	0.57844	0.46306	0.60446	0.6295	0.59102	0.6221	0.59332	0.4717
	9 week	0.68554	0.63392	0.46944	0.35282	0.54642	0.46498	0.59428	0.62982	0.58268	0.62044	0.57766	0.47366
SE	3 week	0.058204	0.061853	0.0489	0.118871	0.05513	0.083328	0.04124	0.041756	0.050255	0.060816	0.046704	0.068971
	6 week	0.044058	0.051593	0.064132	0.088227	0.050671	0.081858	0.027373	0.038082	0.032951	0.049411	0.039161	0.045312
	9 week	0.040129	0.044693	0.05194	0.045552	0.042335	0.048988	0.02116	0.03324	0.023423	0.042319	0.030848	0.025176
	Average	0.69056	0.641747	0.516947	0.334927	0.569767	0.47926	0.60274	0.630047	0.588387	0.626307	0.58946	0.475687

StdDev	0.096261	0.106518	0.11584	0.179318	0.100717	0.147842	0.062479	0.075709	0.074585	0.103081	0.079337	0.09972

C.8 3-week average F_1 scores of CR detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Accuracies for the 3-week average groups were calculated based on the average of the first three, six or nine weeks and described as weeks from Zadoks (Z) 13, 21. Supplementary to Chapter 3.

F_1 (80/20 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	Average	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	3 week	0.6071	0.6163	0.3064	0.019	0.4479	0.2025	0.6176	0.6815	0.6028	0.6746	0.5493	0.4877
	6 week	0.6042	0.6214	0.3111	0.0095	0.446	0.2541	0.613	0.6573	0.6043	0.6374	0.5433	0.4481
	9 week	0.6271	0.6132	0.424	0.1418	0.5081	0.3274	0.6112	0.6502	0.5964	0.6466	0.5522	0.429
Glasshouse 2	3 week	0.8519	0.907	0.7895	0.8267	0.8076	0.8089	0.8385	0.8154	0.8452	0.9114	0.851	0.7632
	6 week	0.8349	0.8856	0.7081	0.7336	0.7248	0.7595	0.7528	0.7433	0.7573	0.8556	0.7628	0.6267
	9 week	0.8034	0.8404	0.5656	0.5819	0.665	0.687	0.7035	0.7181	0.6986	0.8234	0.7033	0.5345
Glasshouse 3	3 week	0.8675	0.8321	0.6755	0.69	0.6829	0.7529	0.6722	0.7766	0.6619	0.8148	0.614	0.4364
	6 week	0.842	0.8324	0.6469	0.6568	0.6505	0.7317	0.6963	0.7783	0.6453	0.8247	0.6348	0.4932
	9 week	0.8296	0.8236	0.6723	0.6313	0.6546	0.7131	0.7018	0.7748	0.6858	0.8188	0.6541	0.513
Field 1	3 week	0.6519	0.6535	0.5853	0.6108	0.5796	0.6083	0.5508	0.5924	0.6021	0.6355	0.5477	0.4784
	6 week	0.6676	0.6835	0.6206	0.6409	0.6121	0.6467	0.5928	0.6184	0.6118	0.6776	0.5886	0.521
	9 week	0.6543	0.6621	0.6212	0.6317	0.6127	0.6187	0.5817	0.6346	0.5865	0.6801	0.5707	0.5025
Field 2	3 week	0.6818	0.5362	0.3364	0.0805	0.494	0.444	0.602	0.5503	0.5849	0.5466	0.5138	0.4306
	6 week	0.68	0.567	0.4843	0.3402	0.5622	0.4923	0.6076	0.5917	0.612	0.5823	0.5428	0.4368
	9 week	0.6835	0.5866	0.5043	0.283	0.5648	0.4954	0.6024	0.5872	0.6199	0.5834	0.5521	0.4539
Average	3 week	0.73204	0.70902	0.53862	0.4454	0.6024	0.56332	0.65622	0.68324	0.65938	0.71658	0.61516	0.51926
	6 week	0.72574	0.71798	0.5542	0.4762	0.59912	0.57686	0.6525	0.6778	0.64614	0.71552	0.61446	0.50516
	9 week	0.71958	0.70518	0.55748	0.45394	0.60104	0.56832	0.64012	0.67298	0.63744	0.71046	0.60648	0.48658
SE	3 week	0.053513	0.069243	0.094519	0.165461	0.065088	0.110159	0.049511	0.051044	0.048252	0.065127	0.061149	0.062008
	6 week	0.047789	0.061033	0.070948	0.134592	0.04659	0.093149	0.030918	0.035883	0.028682	0.053306	0.040796	0.033982
	9 week	0.040771	0.053238	0.043563	0.101509	0.029195	0.071044	0.025975	0.03298	0.023093	0.047769	0.03068	0.019534
	Average	0.725787	0.710727	0.5501	0.458513	0.600853	0.5695	0.649613	0.678007	0.647653	0.714187	0.612033	0.503667

StdDev	0.095423	0.123155	0.145675	0.273064	0.098392	0.185769	0.074094	0.0816	0.070666	0.111776	0.092061	0.085756

C.9 Average number of features chosen, between 1 and 10, optimised for accuracy of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on a 20-80% train-test split. Accuracies were calculated based on the best accuracy score on any given week from Zadoks (Z) 13, 21. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 3.

Number of Features (20/80 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	week 1	4	8	7	1	6	6	6	7	8	6	1	1
	week 2	7	4	7	10	1	1	9	9	8	5	3	10
	week 3	7	1	7	4	8	10	1	1	9	8	2	9
	week 4	9	9	9	9	10	2	4	5	2	8	2	6
	week 5	6	8	8	5	6	5	8	6	10	8	4	7
	week 6	8	4	4	2	1	7	2	8	5	6	1	7
	week 7	8	6	8	3	5	10	9	7	7	5	10	7
	week 8	9	1	7	2	7	7	6	2	3	4	4	9
	week 9	10	1	1	1	1	3	1	2	1	3	3	9
Glasshouse 2	week 1	6	1	9	5	10	1	9	9	10	2	9	9
	week 2	10	6	8	4	9	2	8	4	10	9	2	3
	week 3	2	6	10	8	9	3	3	8	4	4	3	8
	week 4	1	2	4	8	7	9	4	5	4	5	10	8
	week 5	5	10	10	3	10	1	8	1	10	2	9	7
	week 6	9	10	8	8	6	6	8	2	9	2	4	9
	week 7	6	2	9	10	6	1	1	3	1	5	7	8
	week 8	-	-	-	-	-	-	-	-	-	-	-	-
	week 9	4	10	10	7	9	9	9	1	1	8	10	9
Glasshouse 3	week 1	5	10	2	9	8	4	5	4	5	4	1	8
	week 2	7	1	8	8	9	8	8	3	9	5	1	9
	week 3	8	6	9	6	10	6	7	5	8	1	7	8

	week 4	10	5	2	8	2	2	8	10	10	1	3	8
	week 5	4	1	1	6	2	4	3	6	1	5	2	1
	week 6	8	1	1	5	1	6	10	2	3	5	3	7
	week 7	5	5	9	4	2	4	8	7	9	2	8	4
	week 8	7	7	5	8	4	2	10	10	4	4	1	8
	week 9	7	1	6	5	2	7	7	5	9	1	1	9
	week 1	8	3	9	9	1	9	9	8	10	3	9	8
	week 2	9	9	9	9	4	5	7	7	6	5	7	3
	week 3	10	6	10	9	9	9	5	8	10	2	10	8
	week 4	6	5	8	7	7	1	1	6	1	5	2	5
Field 1	week 5	10	8	8	3	9	3	6	1	9	5	5	7
	week 6	8	6	4	6	3	7	3	2	10	9	1	9
	week 7	8	10	4	6	8	8	5	4	6	1	4	9
	week 8	2	10	1	7	9	2	10	1	10	3	2	10
	week 9	9	3	8	2	2	4	5	10	7	10	9	8
	week 1	2	4	5	2	3	7	9	1	6	4	9	7
	week 2	10	9	7	6	1	6	8	6	8	2	4	8
	week 3	2	10	9	4	6	6	2	7	1	6	1	10
	week 4	8	2	4	6	9	8	4	7	9	1	4	10
Field 2	week 5	2	3	8	7	4	10	3	9	9	7	6	9
	week 6	6	4	3	7	10	7	1	6	2	5	2	9
	week 7	10	4	2	9	1	7	8	6	10	3	1	6
	week 8	10	5	4	7	3	5	2	4	2	9	2	1
	week 9	4	7	10	10	8	5	10	9	8	3	2	8
Avg		7	5	6	6	6	5	6	5	6	5	4	7
StdError		0.41	0.48	0.44	0.39	0.49	0.42	0.45	0.43	0.50	0.38	0.48	0.35
Number of instances as top performer		14	27	16	22	19	17	16	23	13	29	30	13
Percentage of instances as top performer		31.11	60	35.56	48.89	42.22	37.78	35.56	51.11	28.89	64.44	66.67	28.89

C.10 Average number of features chosen, between 1 and 10, optimised for accuracy of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on an 80-20% train-test split. Accuracies were calculated based on the best accuracy score on any given week from Zadoks (Z) 13, 21. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 3.

Number of Features (80/20 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	week 1	3	2	10	4	1	5	6	1	9	2	10	9
	week 2	10	7	10	2	4	4	4	7	4	7	8	5
	week 3	9	6	5	2	3	4	3	1	5	8	2	6
	week 4	9	1	3	6	3	6	8	7	1	2	7	4
	week 5	6	1	4	1	6	5	8	2	9	2	10	2
	week 6	7	9	7	4	7	8	8	9	8	4	10	8
	week 7	3	6	7	1	8	4	7	2	10	2	10	9
	week 8	3	8	7	2	9	5	4	7	5	7	4	4
	week 9	9	9	2	9	1	2	8	4	3	9	1	5
Glasshouse 2	week 1	4	1	9	6	9	2	8	1	9	3	8	2
	week 2	5	8	2	5	5	2	9	2	2	1	6	6
	week 3	6	1	9	1	3	2	8	9	8	5	9	1
	week 4	8	3	1	4	8	2	3	4	6	1	4	3
	week 5	6	1	9	4	8	1	7	1	9	7	9	1
	week 6	4	4	9	5	4	2	8	4	8	8	1	1
	week 7	6	7	5	2	1	1	3	4	2	5	1	7
	week 8	-	-	-	-	-	-	-	-	-	-	-	-
	week 9	4	4	1	9	8	8	9	9	7	8	7	3
Glasshouse 3	week 1	7	2	8	8	4	8	2	6	6	4	8	1
	week 2	9	2	6	1	9	7	7	5	7	3	9	2
	week 3	8	1	8	6	9	2	7	9	8	9	9	4

	week 4	7	1	8	4	9	9	9	3	9	7	2	4
	week 5	5	4	7	4	7	8	6	1	7	5	6	4
	week 6	7	9	9	7	2	2	9	8	8	6	6	2
	week 7	6	2	2	6	8	4	9	4	8	5	9	1
	week 8	1	9	2	3	5	3	6	4	7	3	4	1
	week 9	8	5	6	4	8	9	9	5	7	7	8	1
Field 1	week 1	8	8	4	3	3	7	8	4	8	6	6	9
	week 2	9	2	4	6	7	8	3	4	7	1	9	1
	week 3	9	6	1	2	9	3	2	8	2	2	2	7
	week 4	9	1	8	1	8	1	7	7	8	1	3	6
	week 5	4	7	7	7	5	8	7	2	9	4	8	3
	week 6	8	1	2	1	9	9	8	1	5	4	4	1
	week 7	8	8	8	9	3	7	2	4	3	2	7	5
	week 8	9	6	5	2	8	8	9	8	9	4	9	6
	week 9	5	9	1	7	2	9	6	7	1	1	7	2
Field 2	week 1	7	1	9	2	2	8	9	2	3	8	8	1
	week 2	9	4	8	9	1	1	8	1	5	6	5	4
	week 3	2	5	9	1	9	9	2	7	1	5	5	8
	week 4	9	3	3	3	2	5	4	1	2	2	8	9
	week 5	3	4	9	6	9	8	7	6	5	1	6	6
	week 6	2	9	3	7	3	9	8	9	4	6	5	9
	week 7	1	1	9	1	9	7	2	3	1	1	3	8
	week 8	2	9	7	1	9	1	4	3	3	7	9	8
	week 9	2	1	6	8	3	6	1	3	4	6	7	9
Avg		6	5	6	4	6	5	6	5	6	4	6	5
StdError		0.40	0.46	0.44	0.40	0.44	0.43	0.38	0.41	0.42	0.38	0.41	0.43
Number of instances as top performer		11	28	14	27	17	19	17	26	14	25	12	29
Percentage of instances as top performer		24.44	62.22	31.11	60	37.78	42.22	37.78	57.78	31.11	55.56	26.67	64.44

C.11 Average number of features chosen, between 1 and 10, optimised for F_1 of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on a 20-80% train-test split. F_1 was calculated based on the best F_1 score on any given week from Zadoks (Z) 13, 21. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 3.

Number of Features (20/80 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	week 1	4	8	7	1	6	6	6	7	8	6	1	1
	week 2	7	4	7	10	1	1	9	9	8	5	3	10
	week 3	7	1	7	4	8	10	1	1	9	8	2	9
	week 4	9	9	9	9	10	2	4	5	2	8	2	6
	week 5	6	8	8	5	6	5	8	6	10	8	4	7
	week 6	8	4	4	2	1	7	2	8	5	6	1	7
	week 7	8	6	8	3	5	10	9	7	7	5	10	7
	week 8	9	1	7	2	7	7	6	2	3	4	4	9
	week 9	10	1	1	1	1	3	1	2	1	3	3	9
Glasshouse 2	week 1	6	1	9	5	10	1	9	9	10	2	9	9
	week 2	10	6	8	4	9	2	8	4	10	9	2	3
	week 3	2	6	10	8	9	3	3	8	4	4	3	8
	week 4	1	2	4	8	7	9	4	5	4	5	10	8
	week 5	5	10	10	3	10	1	8	1	10	2	9	7
	week 6	9	10	8	8	6	6	8	2	9	2	4	9
	week 7	6	2	9	10	6	1	1	3	1	5	7	8
	week 8	–	–	–	–	–	–	–	–	–	–	–	–
	week 9	10	8	8	9	3	6	9	8	9	8	7	5
Glasshouse 3	week 1	5	10	2	9	8	4	5	4	5	4	1	8
	week 2	7	1	8	8	9	8	8	3	9	5	1	9
	week 3	8	6	9	6	10	6	7	5	8	1	7	8

	week 4	10	5	2	8	2	2	8	10	10	1	3	8
	week 5	4	1	1	6	2	4	3	6	1	5	2	1
	week 6	8	1	1	5	1	6	10	2	3	5	3	7
	week 7	5	5	9	4	2	4	8	7	9	2	8	7
	week 8	7	7	5	8	4	2	10	10	4	4	1	8
	week 9	7	1	6	5	2	7	7	5	9	1	1	9
Field 1	week 1	8	3	9	9	1	9	9	8	10	3	9	8
	week 2	9	9	9	9	4	5	7	7	6	5	7	3
	week 3	10	6	10	9	9	9	5	8	10	2	10	8
	week 4	6	5	8	7	7	1	1	6	1	5	2	5
	week 5	10	8	8	3	9	3	6	1	9	5	5	7
	week 6	8	6	4	6	3	7	3	2	10	9	1	9
	week 7	8	10	4	6	8	8	5	4	6	1	4	9
	week 8	2	10	1	7	9	2	10	1	10	3	2	10
	week 9	9	3	8	2	2	4	5	10	7	10	9	8
Field 2	week 1	2	4	5	2	3	7	9	1	6	4	9	7
	week 2	10	9	7	6	1	6	8	6	8	2	4	8
	week 3	2	10	9	4	6	6	2	7	1	6	1	10
	week 4	8	2	4	6	9	8	4	7	9	1	4	10
	week 5	2	3	8	7	4	10	3	9	9	7	6	9
	week 6	6	4	3	7	10	7	1	6	2	5	2	9
	week 7	10	4	2	9	1	7	8	6	10	3	1	6
	week 8	10	5	4	7	3	5	2	4	2	9	2	1
	week 9	4	7	10	10	8	5	10	9	8	3	2	8
Avg		7	5	6	6	6	5	6	5	7	5	4	7
StdError		3	3	3	3	3	3	3	3	3	2	3	2
Number of instances as top performer		13	27	16	22	19	17	17	22	13	29	30	12
Percentage of instances as top performer		29.55	61.36	36.36	50	43.18	38.64	38.64	50	29.55	65.91	68.18	27.27

C.12 Average number of features chosen, between 1 and 10, optimised for F_1 of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on an 80-20% train-test split. F_1 was calculated based on the best F_1 score on any given week from Zadoks (Z) 13, 21. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 3.

Number of Features (80/20 split)		MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
		DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Glasshouse 1	week 1	3	2	10	4	1	5	6	1	9	2	10	9
	week 2	10	7	10	2	4	4	4	7	4	7	8	5
	week 3	9	6	5	2	3	4	3	1	5	8	2	6
	week 4	9	1	3	6	3	6	8	7	1	2	7	4
	week 5	6	1	4	1	6	5	8	2	9	2	10	2
	week 6	7	9	7	4	7	8	8	9	8	4	10	8
	week 7	3	6	7	1	8	4	7	2	10	2	10	9
	week 8	3	8	7	2	9	5	4	7	5	7	4	4
	week 9	9	9	2	9	1	2	8	4	3	9	1	5
Glasshouse 2	week 1	4	1	9	6	9	2	8	1	9	3	8	2
	week 2	5	8	2	5	5	2	9	2	2	1	6	6
	week 3	6	1	9	1	3	2	8	9	8	5	9	1
	week 4	8	3	1	4	8	2	3	4	6	1	4	3
	week 5	6	1	9	4	8	1	7	1	9	7	9	1
	week 6	4	4	9	5	4	2	8	4	8	8	1	1
	week 7	6	7	5	2	1	1	3	4	2	5	1	7
	week 8	−	−	−	−	−	−	−	−	−	−	−	−
	week 9	6	4	6	6	5	1	4	5	7	3	4	1
Glasshouse 3	week 1	7	2	8	8	4	8	2	6	6	4	8	1
	week 2	9	2	6	1	9	7	7	5	7	3	9	2
	week 3	8	1	8	6	9	2	7	9	8	9	9	4

APPENDIX C. SUPPLEMENTARY FIGURES AND TABLES

	week 4	7	1	8	4	9	9	9	3	9	7	2	4
	week 5	5	4	7	4	7	8	6	1	7	5	6	4
	week 6	7	9	9	7	2	2	9	8	8	6	6	2
	week 7	6	2	2	6	8	4	9	4	8	5	9	1
	week 8	1	9	2	3	5	3	6	4	7	3	4	1
	week 9	8	5	6	4	8	9	9	5	7	7	8	1
Field 1	week 1	8	8	4	3	3	7	8	4	8	6	6	9
	week 2	9	2	4	6	7	8	3	4	7	1	9	1
	week 3	9	6	1	2	9	3	2	8	2	2	2	7
	week 4	9	1	8	1	8	1	7	7	8	1	3	6
	week 5	4	7	7	7	5	8	7	2	9	4	8	3
	week 6	8	1	2	1	9	9	8	1	5	4	4	1
	week 7	8	8	8	9	3	7	2	4	3	2	7	5
	week 8	9	6	5	2	8	8	9	8	9	4	9	6
	week 9	5	9	1	7	2	9	6	7	1	1	7	2
Field 2	week 1	7	1	9	2	2	8	9	2	3	8	8	1
	week 2	9	4	8	9	1	1	8	1	5	6	5	4
	week 3	2	5	9	1	9	9	2	7	1	5	5	8
	week 4	9	3	3	3	2	5	4	1	2	2	8	9
	week 5	3	4	9	6	9	8	7	6	5	1	6	6
	week 6	2	9	3	7	3	9	8	9	4	6	5	9
	week 7	1	1	9	1	9	7	2	3	1	1	3	8
	week 8	2	9	7	1	9	1	4	3	3	7	9	8
	week 9	2	1	6	8	3	6	1	3	4	6	7	9
Avg		6	5	6	4	6	5	6	4	6	4	6	4
StdError		3	3	3	3	3	3	2	3	3	2	3	3
Number of instances as top performer		12	26	13	27	17	19	17	26	13	25	12	28
Percentage of instances as top performer		27.27	59.09	29.55	61.36	38.64	43.18	38.64	59.09	29.55	56.82	27.27	63.64

C.13 Average accuracy (%) scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Models were developed on data from all trial sites from 3-week groupings from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Accuracy (20/80 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Group 1	52.47%	52.12%	48.26%	48.90%	48.09%	49.81%	51.97%	54.72%	51.97%	55.56%	51.49%	50.89%
Group 2	52.20%	55.17%	49.28%	49.52%	49.17%	50.42%	54.16%	55.43%	53.99%	58.54%	54.98%	52.23%
Group 3	50.96%	50.64%	50.31%	49.04%	50.37%	49.06%	53.35%	54.30%	54.23%	54.24%	54.65%	51.29%
Average	51.88%	52.64%	49.28%	49.15%	49.21%	49.76%	53.16%	54.82%	53.40%	56.11%	53.71%	51.47%
SE	0.46%	1.33%	0.59%	0.19%	0.66%	0.39%	0.64%	0.33%	0.72%	1.27%	1.11%	0.40%
StdDev	0.81%	2.31%	1.03%	0.33%	1.14%	0.68%	1.11%	0.57%	1.24%	2.20%	1.93%	0.69%
Number of instances as top performer	2	1	1	2	1	2	0	3	0	3	3	0
Percentage of instances as top performer	66.67	33.33	33.33	66.67	33.33	66.67	0	100	0	100	100	0

C.14 Average accuracy (%) scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Models were developed on data from all trial sites from 3-week groupings from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Accuracy (80/20 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Group 1	53.62%	58.15%	50.50%	49.68%	53.85%	50.97%	53.18%	65.02%	52.84%	68.45%	53.07%	58.69%
Group 2	55.18%	59.81%	51.98%	49.20%	52.53%	50.37%	56.72%	67.13%	57.16%	70.94%	57.27%	64.58%
Group 3	50.96%	52.28%	49.04%	47.40%	50.62%	48.81%	54.45%	68.66%	56.26%	69.96%	54.79%	62.15%
Average	53.25%	56.75%	50.51%	48.76%	52.33%	50.05%	54.78%	66.94%	55.42%	69.78%	55.04%	61.81%
SE	1.23%	2.28%	0.85%	0.69%	0.94%	0.64%	1.04%	1.06%	1.32%	0.72%	1.22%	1.71%
StdDev	2.13%	3.96%	1.47%	1.20%	1.62%	1.11%	1.79%	1.83%	2.28%	1.25%	2.11%	2.96%
Number of instances as top performer	0	3	3	0	3	0	0	3	0	3	0	3
Percentage of instances as top performer	0	100	100	0	100	0	0	100	0	100	0	100

C.15 Average F_1 scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Models were developed on data from all trial sites from 3-week groupings from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

F_1 (20/80 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Group 1	0.6382	0.5091	0.5609	0.3761	0.5452	0.3857	0.5157	0.5405	0.5001	0.5299	0.4943	0.3813
Group 2	0.5193	0.5124	0	0	0.2765	0.2568	0.5361	0.5521	0.5029	0.5567	0.5338	0.3912
Group 3	0.5955	0.4885	0.6662	0.3837	0.6662	0.4306	0.5456	0.5484	0.5417	0.5258	0.5482	0.4356
Average	0.584333	0.503333	0.409033	0.253267	0.495967	0.3577	0.532467	0.547	0.5149	0.537467	0.525433	0.4027
SE	0.034775	0.007478	0.206763	0.126652	0.115159	0.052088	0.00882	0.003421	0.013424	0.009689	0.016112	0.016696
StdDev	0.060231	0.012952	0.358125	0.219368	0.19946	0.09022	0.015278	0.005925	0.023252	0.016782	0.027907	0.028919
Number of instances as top performer	3	0	2	0	3	0	0	3	1	2	3	0
Percentage of instances as top performer	100	0	66.67	0	100	0	0	100	33.33	66.67	100	0

C.16 Average F_1 scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Models were developed on data from all trial sites from 3-week groupings from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

F_1 (80/20 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Group 1	0.5305	0.5717	0.071	0.3772	0.3802	0.4379	0.5381	0.6583	0.5214	0.6762	0.5215	0.5829
Group 2	0.552	0.6251	0.3209	0.4963	0.4771	0.4208	0.5638	0.6828	0.5692	0.7198	0.5745	0.655
Group 3	0.6092	0.5404	0.6581	0.5446	0.6581	0.5336	0.5422	0.6821	0.5571	0.6898	0.5428	0.6202
Average	0.5639	0.579067	0.35	0.4727	0.505133	0.4641	0.548033	0.6744	0.549233	0.695267	0.546267	0.619367
SE	0.023485	0.024727	0.170105	0.049744	0.081438	0.035099	0.007972	0.008053	0.014348	0.01288	0.015398	0.020818
StdDev	0.040677	0.042828	0.29463	0.086159	0.141055	0.060793	0.013807	0.013947	0.024852	0.022308	0.02667	0.036057
Number of instances as top performer	1	2	1	2	2	1	0	3	0	3	0	3
Percentage of instances as top performer	33.33	66.67	33.33	66.67	66.67	33.33	0	100	0	100	0	100

C.17 Average number of features chosen, between 1 and 10, optimised for accuracy of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on a 20-80% train-test split. Models were developed on data from all trial sites from 3-week groupings from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Number of Features (20/80 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Group 1	9	1	1	1	2	1	9	10	6	10	7	1
Group 2	4	7	1	1	1	1	10	4	8	3	8	1
Group 3	5	3	10	1	10	1	9	1	3	9	3	1
Average	6	4	4	1	4	1	9	5	6	7	6	1
SE	3	3	5	0	5	0	1	5	3	4	3	0
StdDev	2	2	3	0	3	0	0	3	1	2	2	0
Number of instances as top performer	1	2	0	1	0	2	1	2	2	1	0	3
Percentage of instances as top performer	33.33	66.67	0	33.33	0	66.67	33.33	66.67	66.67	33.33	0	100

C.18 Average number of features chosen, between 1 and 10, optimised for accuracy of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on a 80-20% train-test split. Models were developed on data from all trial sites from 3-week groupings from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Number of Features (80/20 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
Group 1	3	4	10	1	1	1	5	3	10	4	2	1
Group 2	8	5	2	1	1	1	8	8	7	4	3	1
Group 3	2	10	1	1	2	1	3	6	3	8	3	1
Average	4	6	4	1	1	1	5	6	7	5	3	1
SE	3	3	5	0	1	0	3	3	4	2	1	0
StdDev	2	2	3	0	0	0	1	1	2	1	0	0
Number of instances as top performer	2	1	0	2	0	1	1	1	1	2	0	3
Percentage of instances as top performer	66.67	33.33	0	66.67	0	33.33	33.33	33.33	33.33	66.67	0	100

C.19 Average accuracy (%) scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Models were developed on data from all trial sites from each week for nine weeks from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Accuracy (20/80 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
week 1	57.95%	54.86%	52.62%	49.63%	53.48%	50.71%	54.69%	54.86%	54.51%	56.53%	55.29%	50.87%
week 2	51.57%	52.49%	48.85%	49.39%	50.47%	49.06%	51.31%	56.99%	52.16%	59.04%	52.76%	52.49%
week 3	54.14%	55.30%	49.68%	49.30%	49.68%	54.05%	53.17%	56.62%	53.57%	58.10%	52.11%	52.10%
week 4	52.74%	52.97%	50.17%	49.60%	51.25%	49.92%	52.33%	55.30%	51.83%	58.35%	51.99%	51.28%
week 5	56.10%	56.93%	49.59%	49.25%	51.40%	52.34%	55.11%	58.43%	57.00%	62.79%	59.14%	50.59%
week 6	50.99%	54.02%	49.18%	49.84%	49.92%	51.81%	53.54%	55.68%	56.41%	57.57%	55.10%	52.21%
week 7	50.89%	50.69%	50.64%	49.23%	51.05%	50.69%	50.97%	53.55%	49.36%	55.40%	50.56%	50.15%
week 8	53.45%	50.95%	49.81%	50.05%	51.77%	49.41%	52.15%	52.57%	55.78%	55.36%	53.82%	50.50%
week 9	55.79%	53.40%	49.55%	49.49%	49.55%	49.57%	55.30%	55.74%	58.14%	56.83%	58.30%	50.04%
Average	53.74%	53.51%	50.01%	49.53%	50.95%	50.84%	53.17%	55.53%	54.31%	57.77%	54.34%	51.14%
SE	0.83%	0.68%	0.37%	0.09%	0.41%	0.54%	0.54%	0.59%	0.95%	0.76%	0.97%	0.31%
StdDev	2.48%	2.03%	1.11%	0.28%	1.24%	1.63%	1.61%	1.77%	2.84%	2.27%	2.91%	0.93%
Number of instances as top performer	4	5	6	3	5	4	0	9	2	7	9	0
Percentage of instances as top performer	44.44	55.56	66.67	33.33	55.56	44.44	0	100	22.22	77.78	100	0

C.20 Average accuracy (%) scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Models were developed on data from all trial sites from each week for nine weeks from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Accuracy (80/20 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
week 1	58.08%	83.06%	52.58%	53.16%	51.55%	52.16%	61.17%	64.12%	62.54%	70.76%	61.51%	60.80%
week 2	53.90%	61.76%	48.47%	52.94%	46.10%	46.41%	55.93%	61.76%	56.95%	67.32%	54.92%	56.21%
week 3	54.87%	60.44%	50.65%	48.91%	54.87%	49.53%	51.30%	61.37%	54.55%	63.24%	50.00%	50.78%
week 4	55.81%	51.60%	49.17%	51.92%	50.17%	48.72%	53.16%	59.29%	52.82%	61.54%	52.82%	52.88%
week 5	65.46%	65.51%	60.20%	51.58%	58.55%	56.33%	63.16%	59.81%	62.50%	67.09%	63.16%	53.48%
week 6	59.87%	58.04%	51.32%	50.79%	47.37%	52.37%	58.88%	64.67%	63.49%	67.82%	64.14%	57.41%
week 7	49.84%	51.54%	49.52%	50.00%	49.52%	47.84%	53.38%	57.72%	56.27%	60.49%	58.20%	54.32%
week 8	54.10%	48.20%	54.10%	43.17%	54.10%	43.53%	57.09%	60.07%	55.60%	57.91%	57.84%	52.16%
week 9	60.84%	59.81%	51.13%	46.73%	49.84%	53.89%	62.14%	58.57%	61.17%	65.73%	61.81%	57.32%
Average	56.97%	60.00%	51.90%	49.91%	51.34%	50.09%	57.36%	60.82%	58.43%	64.66%	58.27%	55.04%
SE	1.46%	3.26%	1.13%	1.02%	1.24%	1.26%	1.35%	0.75%	1.26%	1.30%	1.54%	1.00%
StdDev	4.62%	10.31%	3.56%	3.24%	3.91%	3.98%	4.27%	2.38%	4.00%	4.12%	4.89%	3.15%
Number of instances as top performer	4	5	5	4	5	4	2	7	0	9	6	3
Percentage of instances as top performer	44.44	55.56	55.56	44.44	55.56	44.44	22.22	77.78	0	100	66.67	33.33

C.21 Average F_1 scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on a 20-80% train-test split. Models were developed on data from all trial sites from each week for nine weeks from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

F_1 (20/80 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
week 1	0.5989	0.5514	0.6264	0	0.5612	0.4463	0.5569	0.5512	0.5525	0.565	0.4933	0.4918
week 2	0.6556	0.4557	0.6556	0	0.622	0.2337	0.5445	0.5539	0.5186	0.5391	0.5145	0.4184
week 3	0.4987	0.4455	0	0.0031	0	0.3311	0.5066	0.5763	0.4929	0.5325	0.3841	0.4581
week 4	0.5232	0.4903	0.0196	0.0542	0.457	0.329	0.5366	0.5414	0.4974	0.5467	0.4028	0.4604
week 5	0.5525	0.5939	0	0.6305	0.0621	0.5916	0.5539	0.5854	0.5511	0.6438	0.5934	0.4507
week 6	0.5083	0.5716	0	0.5515	0.4437	0.5318	0.5365	0.5619	0.567	0.575	0.5232	0.4694
week 7	0.5523	0.4444	0.6724	0.2129	0.6193	0.2876	0.529	0.5412	0.5023	0.5339	0.485	0.4617
week 8	0.665	0.5385	0.665	0.6671	0.6505	0.5122	0.5496	0.5183	0.58	0.5657	0.5606	0.5226
week 9	0.4406	0.5115	0	0	0.0032	0.0062	0.5392	0.5662	0.5547	0.5455	0.4515	0.4085
Average	0.555011	0.511422	0.293222	0.235478	0.379889	0.363278	0.5392	0.555089	0.535167	0.5608	0.489822	0.460178
SE	0.024611	0.018688	0.11445	0.098314	0.09274	0.060304	0.005064	0.006759	0.010882	0.011523	0.022887	0.011483
StdDev	0.073833	0.056063	0.343349	0.294943	0.278221	0.180912	0.015192	0.020276	0.032647	0.03457	0.068661	0.034449
Number of instances as top performer	6	3	3	5	5	4	2	7	2	7	7	2
Percentage of instances as top performer	66.67	33.33	33.33	55.56	55.56	44.44	22.22	77.78	22.22	77.78	77.78	22.22

C.22 Average F_1 scores of CR detection of machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA), trained on an 80-20% train-test split. Models were developed on data from all trial sites from each week for nine weeks from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

F_1 (80/20 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
week 1	0.609	0.8283	0.5941	0.6158	0.5506	0.5978	0.6143	0.6447	0.6355	0.7059	0.6242	0.604
week 2	0.5466	0.5804	0.6021	0.5636	0.4631	0.4384	0.539	0.6214	0.5436	0.671	0.4789	0.5379
week 3	0.5209	0.6019	0.1739	0.374	0.3966	0.5091	0.5223	0.6287	0.5347	0.638	0.4762	0.4936
week 4	0.6726	0.4752	0.0129	0.537	0.4	0.3333	0.5402	0.614	0.5034	0.6026	0.5085	0.5017
week 5	0.6921	0.6766	0.5353	0.5234	0.5987	0.5	0.6387	0.5897	0.6299	0.6667	0.6316	0.5421
week 6	0.6429	0.5611	0.6445	0.5691	0.6429	0.5572	0.5928	0.5987	0.6364	0.6667	0.5792	0.5485
week 7	0.6624	0.6609	0.6624	0.4873	0.6565	0.5986	0.5559	0.5954	0.5208	0.6049	0.488	0.4599
week 8	0.6145	0.5387	0.6733	0.5153	0.5763	0.453	0.5965	0.5882	0.5673	0.5923	0.5891	0.5461
week 9	0.5849	0.5806	0.6267	0.3736	0.5392	0.5163	0.6061	0.5522	0.5761	0.6452	0.5903	0.5651
Average	0.616211	0.611522	0.5028	0.506567	0.535989	0.500411	0.578422	0.603667	0.571967	0.6437	0.551778	0.533211
SE	0.019275	0.033771	0.079705	0.02789	0.032268	0.028062	0.013381	0.009058	0.017117	0.012662	0.021151	0.014174
StdDev	0.057825	0.101312	0.239114	0.083671	0.096805	0.084186	0.040144	0.027175	0.051352	0.037986	0.063452	0.042523
Number of instances as top performer	6	3	6	3	7	2	3	6	0	9	7	2
Percentage of instances as top performer	66.67	33.33	66.67	33.33	77.78	22.22	33.33	66.67	0	100	77.78	22.22

C.23 Average number of features chosen, between 1 and 10, optimised for accuracy of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on a 20-80% train-test split. Models were developed on data from all trial sites from each week for nine weeks from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Number of Features (20/80 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
week 1	10	2	6	1	1	1	9	9	5	5	10	1
week 2	9	6	7	1	2	1	4	3	10	1	7	5
week 3	9	10	1	1	1	1	8	5	7	8	6	1
week 4	10	9	4	1	6	1	5	6	7	1	4	1
week 5	8	3	1	1	10	1	7	10	3	10	3	1
week 6	6	5	1	1	5	1	4	5	10	3	10	1
week 7	10	9	9	1	7	1	4	9	8	10	1	1
week 8	10	5	1	1	3	1	10	3	7	7	5	1
week 9	5	4	1	1	10	1	10	9	9	10	7	1
Average	9	6	3	1	5	1	7	7	7	6	6	1
SE	2	3	3	0	4	0	3	3	2	4	3	1
StdDev	1	1	1	0	1	0	1	1	1	1	1	0
Number of instances as top performer	1	8	0	4	0	7	4	4	4	3	0	8
Percentage of instances as top performer	11.11	88.89	0	44.44	0	77.78	44.44	44.44	44.44	33.33	0	88.89

C.24 Average number of features chosen, between 1 and 10, optimised for accuracy of CR classification detection of optimised machine learning models pre-processed with the discrete wavelet transform (DWT) or principal component analysis (PCA) on a 80-20% train-test split. Models were developed on data from all trial sites from each week for nine weeks from Zadoks (Z) 13, 21 across all sites. The models utilised were an artificial neural network (MLP), a support vector classifier (SVC), logistic regression, a decision tree classifier, an extreme forest classifier and a k-nearest neighbor classifier. Supplementary to Chapter 4.

Number of Features (80/20 split)	MLP		SVC		Logistic Regression		Decision Tree		Extreme Forest		K-Nearest Neighbor	
	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA	DWT	PCA
week 1	8	1	1	1	3	1	4	4	6	5	4	1
week 2	9	7	3	1	7	1	9	7	8	9	3	1
week 3	1	6	6	1	8	1	5	6	5	2	7	1
week 4	5	5	3	1	6	1	2	1	6	5	1	1
week 5	9	9	3	1	3	1	9	7	4	8	10	1
week 6	10	9	1	1	1	1	10	10	7	7	7	1
week 7	7	2	1	1	7	1	4	7	7	3	9	1
week 8	10	7	4	1	2	1	8	1	2	9	6	1
week 9	9	4	4	1	8	1	2	6	2	1	2	1
Average	8	6	3	1	5	1	6	5	5	5	5	1
SE	3	3	2	0	3	0	3	3	2	3	3	0
StdDev	1	1	1	0	1	0	1	1	1	1	1	0
Number of instances as top performer	1	6	0	5	0	8	3	4	3	5	0	8
Percentage of instances as top performer	11.11	66.67	0	55.56	0	88.89	33.33	44.44	33.33	55.56	0	88.89

C.25 F_1-scores and average mean and standard deviation (std) across machine learning methods for each of the six rating systems (Chapter 6, Fig. 1) for both PCA and DWT preprocessing techniques in the initial glasshouse trial at Leslie Research Centre, 2018. Supplementary to Chapter 6.

	1	2	3	4	5	6
DWT Mean	0.3522	0.5288	0.5594	0.7215	0.7215	0.6252
DWT Std	0.0553	0.0418	0.0477	0.0444	0.0444	0.0839
PCA Mean	0.3398	0.5124	0.5358	0.7133	0.7165	0.6928
PCA Std	0.0497	0.069	0.0332	0.0649	0.0608	0.1954

Rating Group

C.26 F_1-scores and average mean and standard deviation (std) across machine learning methods for each of the six rating systems (Chapter 6, Fig. 1) for both PCA and DWT preprocessing techniques in the second glasshouse trial at Leslie Research Centre, 2018. Supplementary to Chapter 6.

	1	2	3	4	5	6
DWT Mean	0.4867	0.8426	0.8497	0.9302	0.9085	0.8495
DWT Std	0.1033	0.134	0.1192	0.038	0.0864	0.1195
PCA Mean	0.4738	0.7943	0.8155	0.886	0.8668	0.8713
PCA Std	0.0678	0.1104	0.0959	0.0409	0.0647	0.1411
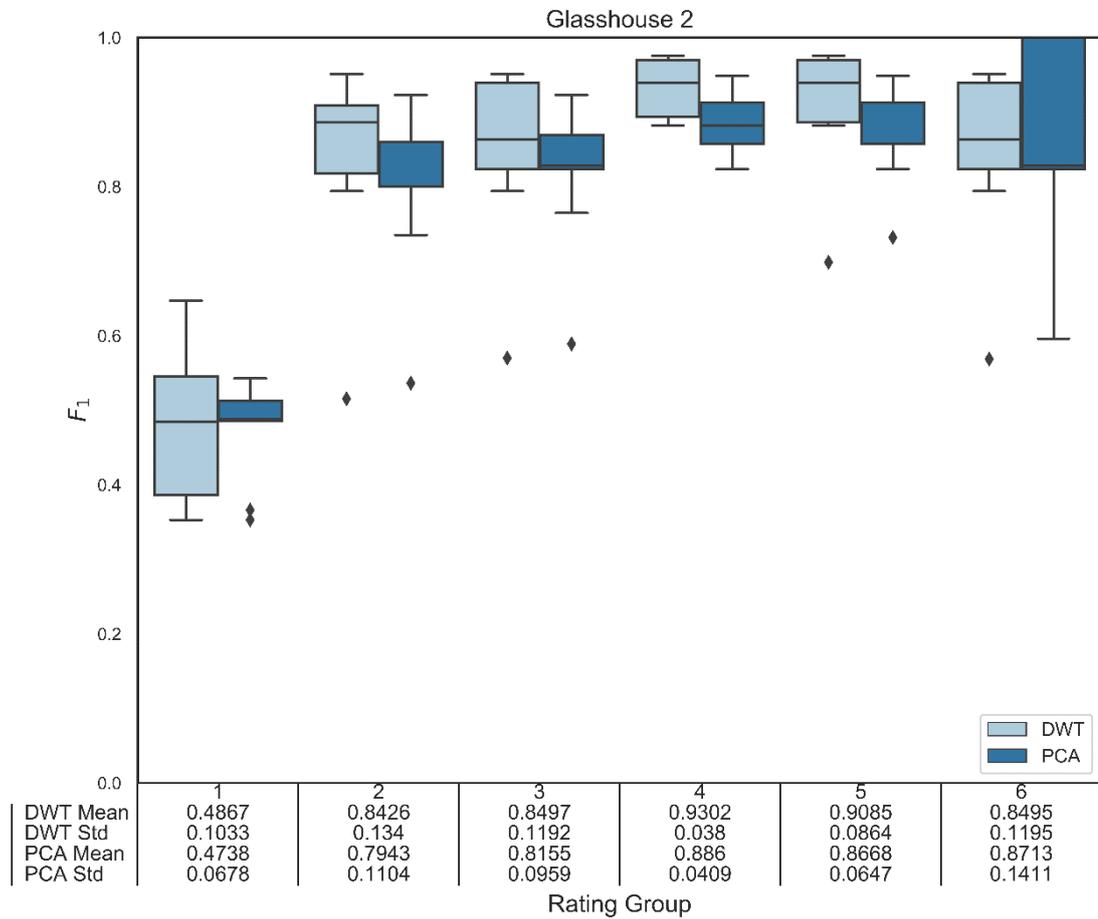
C.27 F_1-scores and average mean and standard deviation (std) across machine learning methods for each of the six rating systems (Chapter 6, Fig. 1) for both PCA and DWT preprocessing techniques in the second glasshouse trial at USQ-CCH, 2019. Supplementary to Chapter 6.

	1	2	3	4	5	6
DWT Mean	0.2204	0.2996	0.3889	0.554	0.5584	0.7353
DWT Std	0.0459	0.0521	0.0764	0.0885	0.0921	0.0581
PCA Mean	0.1514	0.2495	0.3723	0.5806	0.5752	0.8609
PCA Std	0.0714	0.0829	0.0785	0.0783	0.0655	0.1067

C.28 F_1-scores and average mean and standard deviation (std) across machine learning methods for each of the six rating systems (Chapter 6, Fig. 1) for both PCA and DWT preprocessing techniques in the initial field trial at the Tosari Research Station, 2018. Supplementary to Chapter 6.
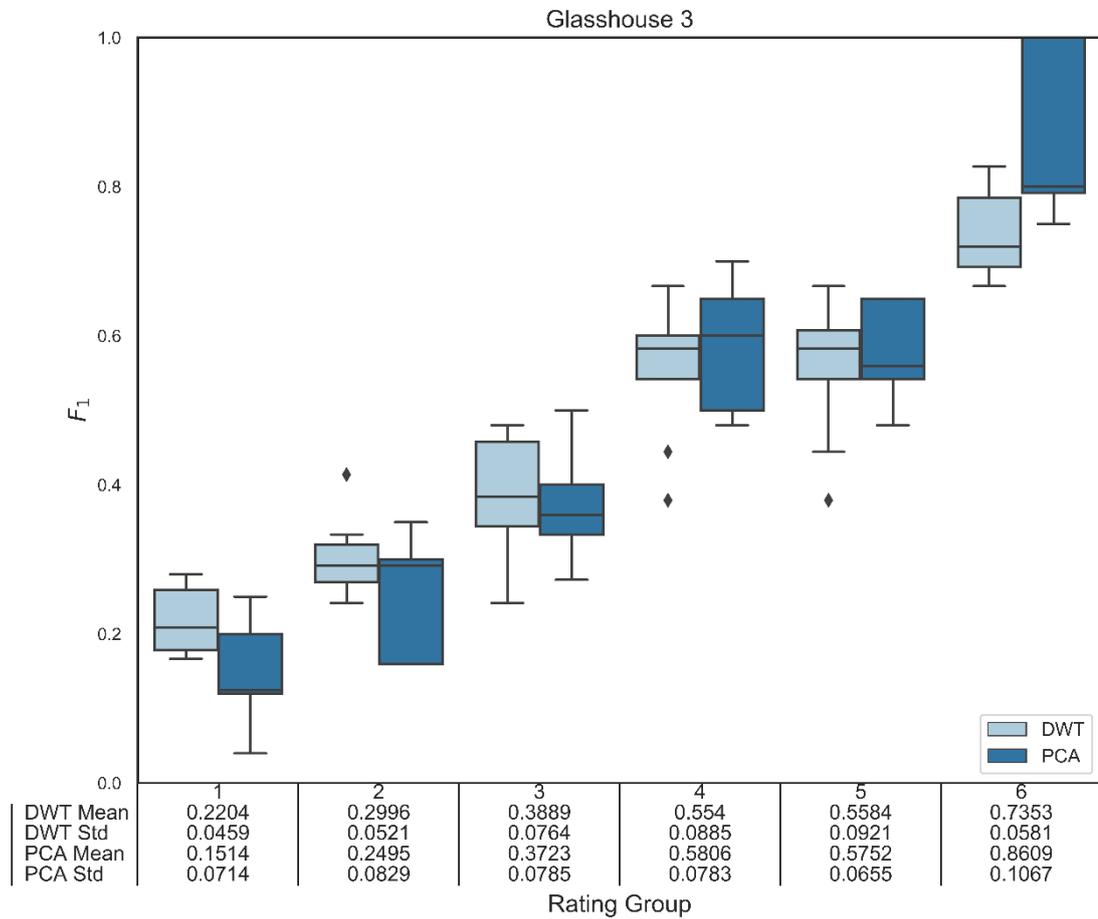

	1	2	3	4	5	6
DWT Mean	0.5859	0.5968	0.6075	0.9652	0.6279	0.6102
DWT Std	0.0223	0.0303	0.0467	0.0138	0.0496	0.0443
PCA Mean	0.5806	0.5806	0.5845	0.9352	0.5988	0.6741
PCA Std	0.0219	0.0222	0.0247	0.0173	0.0266	0.186

C.29 F_1-scores and average mean and standard deviation (std) across machine learning methods for each of the six rating systems (Chapter 6, Fig. 1) for both PCA and DWT preprocessing techniques in the second field trial at the Tosari Research Station, 2019. Supplementary to Chapter 6.
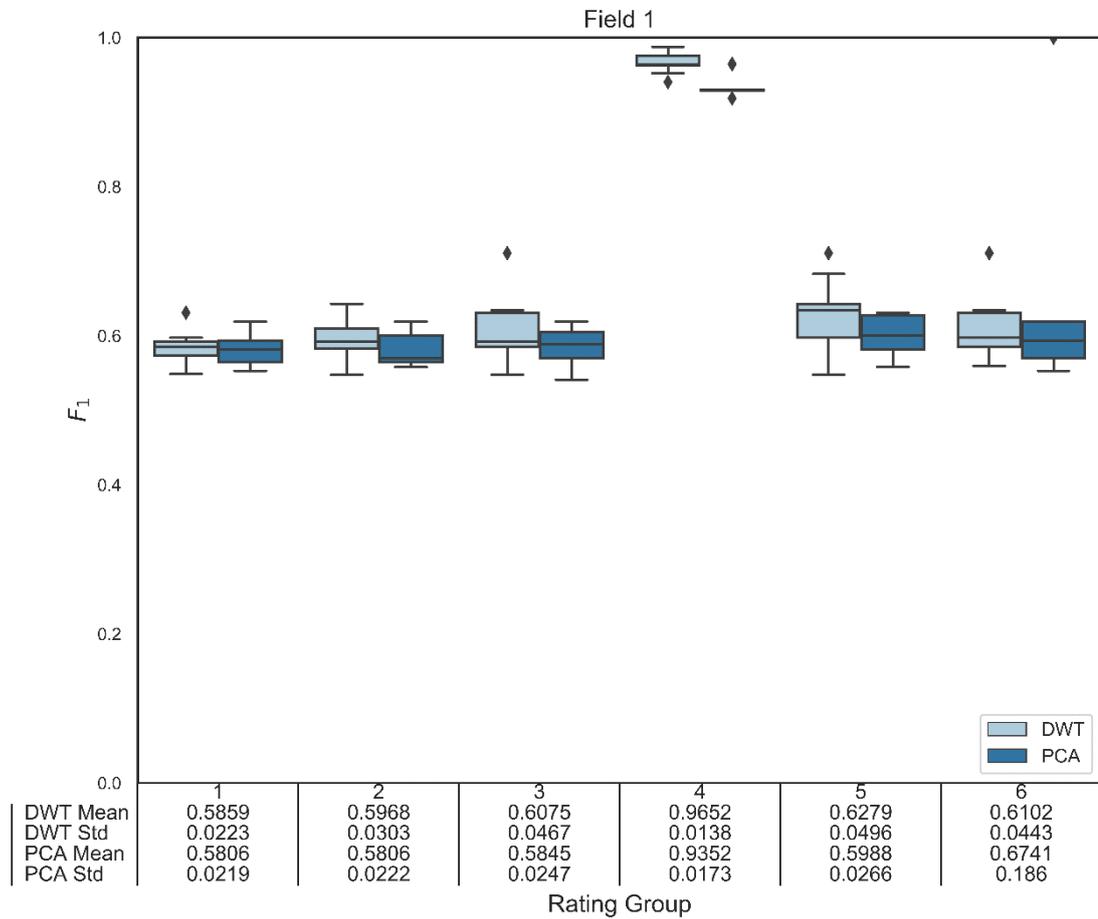

	1	2	3	4	5	6
DWT Mean	0.5172	0.6657	0.732	0.9629	0.8727	0.732
DWT Std	0.033	0.0273	0.023	0.015	0.0174	0.023
PCA Mean	0.5185	0.7356	0.7727	0.954	0.8621	0.8493
PCA Std	0.009	0.0	0.0051	0.0	0.0	0.1131
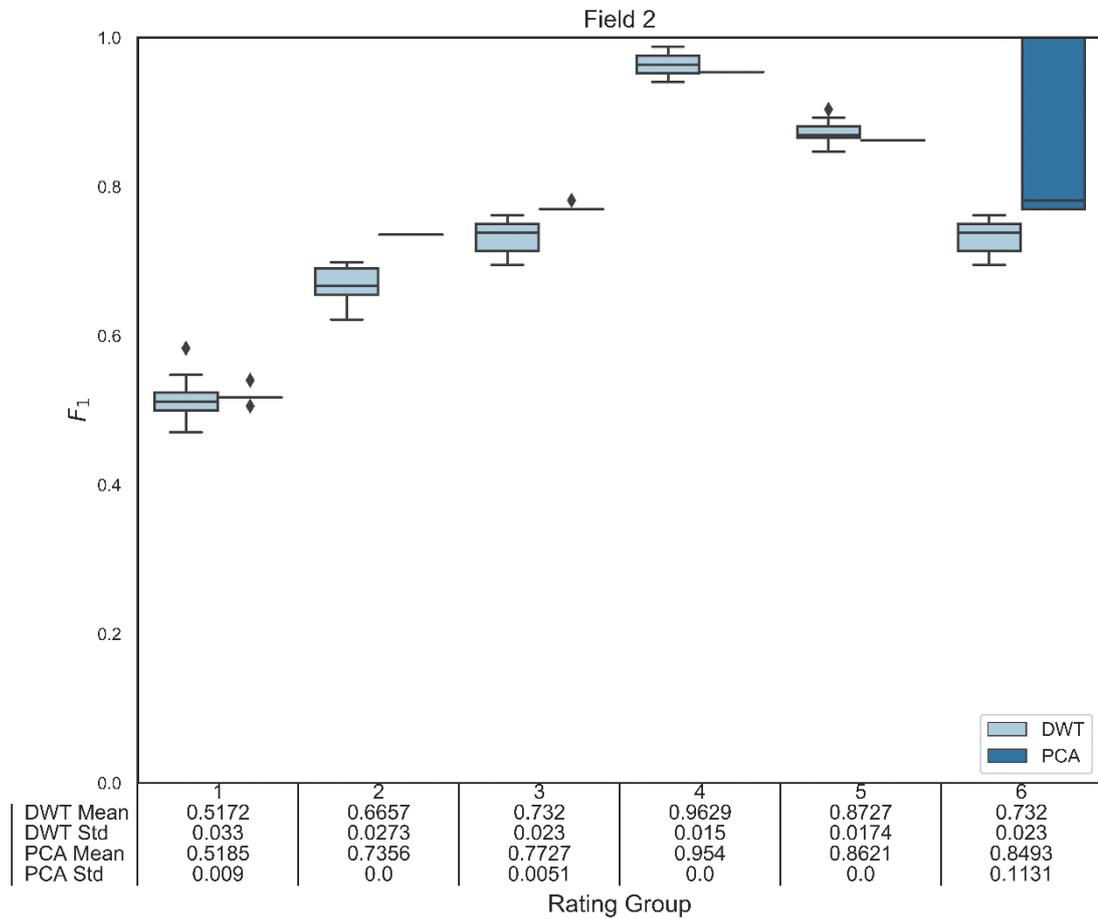
APPENDIX D – DEVELOPED CODE

The stages of model development for both the contact sensor detection and quantification models (D.1) and the camera-based detection and quantification models (D.3, D.4) are presented. Code files are available upon request.

D.1 OVERVIEW OF MODEL DEVELOPMENT (NIRSCAN NANO)

Flow chart representing the stages of model development from raw data recorded by the NIRscan Nano through to final model for both DWT (blue) and PCA-based (green) approaches.
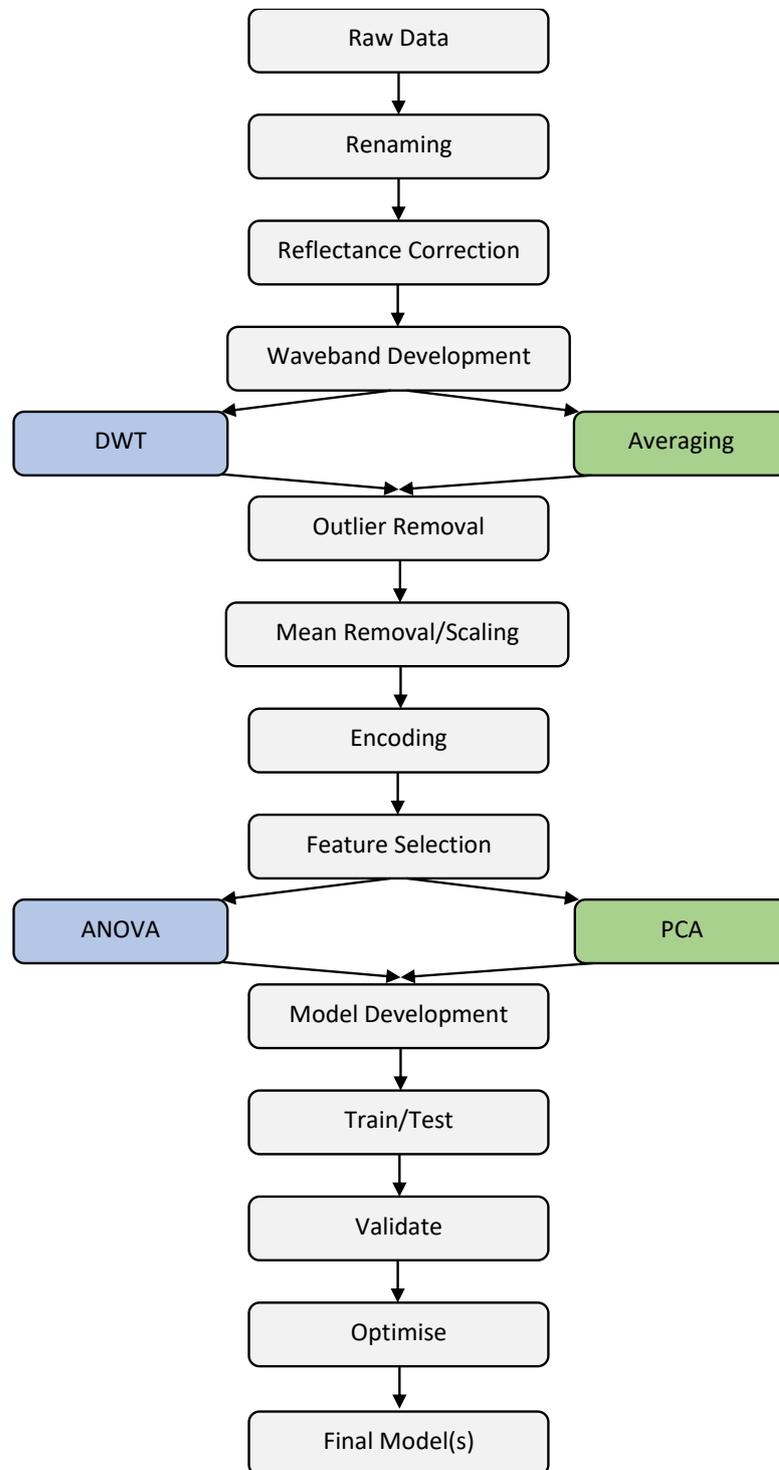
D.2 OVERVIEW OF MODEL DEVELOPMENT (TAU SWIR BINARY)

Flow chart representing the stages of model development for binary classification of CR (+,-) from raw image data recorded by the FLIR Tau SWIR with five narrow-bandpass filters.

```
┌─────────────────────┐
│      Raw Data       │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│      Encoding       │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Training Split   │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│ Greyscale Conversion│
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│      Reshaping      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│       Scaling       │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Concatenation    │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Model Building   │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│      Compiling      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│       Fitting       │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│     Final Model     │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│     Evaluation      │
└─────────────────────┘
```
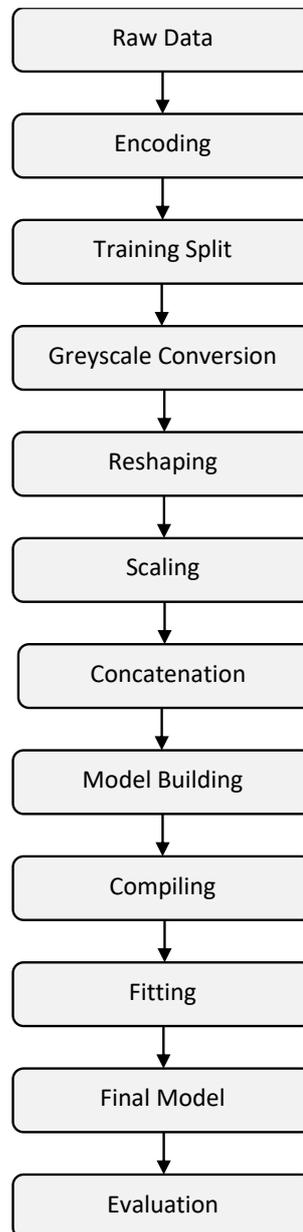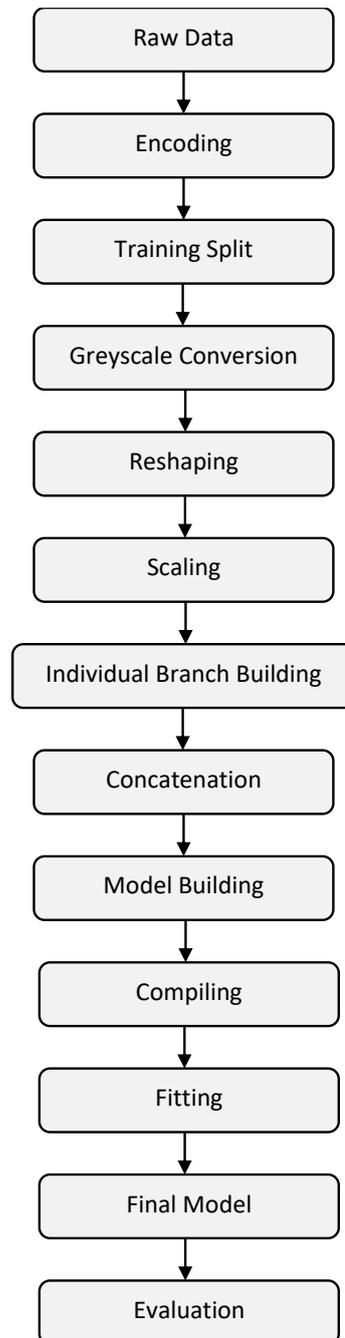
D.3 OVERVIEW OF MODEL DEVELOPMENT (TAU SWIR QUANTIFICATION)

Flow chart representing the stages of model development for quantification classification of CR (+,-), using three scales, from raw image data recorded by the FLIR Tau SWIR with five narrow-bandpass filters.

Raw Data

↓

Encoding

↓

Training Split

↓

Greyscale Conversion

↓

Reshaping

↓

Scaling

↓

Individual Branch Building

↓

Concatenation

↓

Model Building

↓

Compiling

↓

Fitting

↓

Final Model

↓

Evaluation

APPENDIX E – MODEL HYPERPARAMETERS

E.1 OVERVIEW OF MODEL HYPERPARAMETERS USED FOR MODEL DEVELOPMENT

The model and hyperparameters used for each input model from Scikit-learn version 0.21.2 (Pedregosa et al. 2011). All combinations of hyperparameters were evaluated.

Model	Hyperparameters				
	activation function(s)	solver	learning rate	hidden layer size	maximum iterations
MLPClassifier()	identity logistic tanh relu	lbfgs sgd adam	constant invscaling adaptive	100 200 500	100 200 500
	kernel	degree	gamma		
SVC()	linear poly rbf sigmoid	2 3 4 5	auto .001 .01 .1		
LogisticRegression()	default parameters (Scikit-learn version 0.21.2)				
	splitter	criterion			
DecisionTreeClassifier()	best random	gini entropy			
	n_estimators	criterion			
ExtraTreesClassifier()	10 100 150 200	gini entropy			