A MACHINE LEARNING–REMOTE SENSING FRAMEWORK FOR MODELLING WATER STRESS IN SHIRAZ VINEYARDS

KYLE DEVRONNE LOGGENBERG

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in the Faculty of Science at Stellenbosch University.

Supervisor: Mr Nitesh Poona

Co-supervisor: Dr Albert Strever

December 2018

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

The thesis includes two original manuscripts that were published in/submitted to peer-reviewed journals. The manuscripts comprise Chapters 3 and 4 of the thesis, where the nature and scope of my contribution were as follows:

Chapter	Nature of contribution
Chapter 3	This chapter was published as a journal article (Loggenberg, Strever, Greyling & Poona 2018) in *Remote Sensing*, Volume 10, Issue 2 (doi:10.3390/rs10020202) and was co-authored by my supervisors and Berno Greyling. My supervisors contributed to the conceptualisation of the research, data collection, interpretation of results and editing of the manuscript. Berno Greyling aided in data collection and data analysis. I carried out the literature review, data collection, main analysis, and wrote the manuscript.
Chapter 4	This chapter was submitted for peer review in the *International Journal of Remote Sensing*. It was co-authored by my main supervisor who contributed to the conceptualisation of the research, data collection, interpretation of results and editing of the manuscript. I carried out the literature review, data collection, main analysis, and wrote the manuscript.

Date: December 2018

KYLE DEVRONNE LOGGENBERG

SUMMARY

Water is a limited natural resource and a major environmental constraint for crop production in viticulture. The unpredictability of rainfall patterns, combined with the potentially catastrophic effects of climate change, further compound water scarcity, presenting dire future scenarios of undersupplied irrigation systems. Major water shortages could lead to devastating loses in grape production, which would negatively affect job security and national income. It is, therefore, imperative to develop management schemes and farming practices that optimise water usage and safeguard grape production.

Hyperspectral remote sensing techniques provide a solution for the monitoring of vineyard water status. Hyperspectral data, combined with the quantitative analysis of machine learning ensembles, enables the detection of water-stressed vines, thereby facilitating precision irrigation practices and ensuring quality crop yields. To this end, the thesis set out to develop a machine learning–remote sensing framework for modelling water stress in a Shiraz vineyard.

The thesis comprises two components. Component one assesses the utility of terrestrial hyperspectral imagery and machine learning ensembles to detect water-stressed Shiraz vines. The Random Forest (RF) and Extreme Gradient Boosting (XGBoost) ensembles were employed to discriminate between water-stressed and non-stressed Shiraz vines. Results showed that both ensemble learners could effectively discriminate between water-stressed and non-stressed vines. When using all wavebands (p = 176), RF yielded a test accuracy of 83.3% (KHAT = 0.67), with XGBoost producing a test accuracy of 80.0% (KHAT = 0.6).

Component two explores semi-automated feature selection approaches and hyperparameter value optimisation to improve the developed framework. The utility of the Kruskal-Wallis (KW) filter, Sequential Floating Forward Selection (SFFS) wrapper, and a Filter-Wrapper (FW) approach, was evaluated. When using optimised hyperparameter values, an increase in test accuracy ranging from 0.8% to 5.0% was observed for both RF and XGBoost. In general, RF was found to outperform XGBoost. In terms of predictive competency and computational efficiency, the developed FW approach was the most successful feature selection method implemented.

The developed machine learning–remote sensing framework warrants further investigation to confirm its efficacy. However, the thesis answered key research questions, with the developed framework providing a point of departure for future studies.

KEYWORDS

Remote sensing; terrestrial hyperspectral imaging; vineyards; water stress; machine learning; tree-based classifiers; feature selection; hyperparameter value optimisation

OPSOMMING

Water is 'n beperkte natuurlike hulpbron en 'n groot omgewingsbeperking vir gewasproduksie in wingerdkunde. Die onvoorspelbaarheid van reënvalpatrone, gekombineer met die potensiële katastrofiese gevolge van klimaatsverandering, voorspel ‘n toekoms van water tekorte vir besproeiingstelsels. Groot water tekorte kan lei tot groot verliese in druiweproduksie, wat 'n negatiewe uitwerking op werksekuriteit en nasionale inkomste sal hê. Dit is dus noodsaaklik om bestuurskemas en boerderypraktyke te ontwikkel wat die gebruik van water optimaliseer en druiweproduksie beskerm.

Hyperspectrale afstandswaarnemingstegnieke bied 'n oplossing vir die monitering van wingerd water status. Hiperspektrale data, gekombineer met die kwantitatiewe analise van masjienleer klassifikasies, fasiliteer die opsporing van watergestresde wingerdstokke. Sodoende verseker dit presiese besproeiings praktyke en kwaliteit gewasopbrengs. Vir hierdie doel het die tesis probeer 'n masjienleer-afstandswaarnemings raamwerk ontwikkel vir die modellering van waterstres in 'n Shiraz-wingerd.

Die tesis bestaan uit twee komponente. Komponent 1 het die nut van terrestriële hiperspektrale beelde en masjienleer klassifikasies gebruik om watergestresde Shiraz-wingerde op te spoor. Die Ewekansige Woud (RF) en Ekstreme Gradiënt Bevordering (XGBoost) algoritme was gebruik om te onderskei tussen watergestresde en nie-gestresde Shiraz-wingerde. Resultate het getoon dat beide RF en XGBoost effektief kan diskrimineer tussen watergestresde en nie-gestresde wingerdstokke. Met die gebruik van alle golfbande (p = 176) het RF 'n toets akkuraatheid van 83.3% (KHAT = 0.67) behaal en XGBoost het 'n toets akkuraatheid van 80.0% (KHAT = 0.6) gelewer.

Komponent twee het die gebruik van semi-outomatiese veranderlike seleksie benaderings en hiperparameter waarde optimalisering ondersoek om die ontwikkelde raamwerk te verbeter. Die nut van die Kruskal-Wallis (KW) filter, sekwensiële drywende voorkoms seleksie (SFFS) wrapper en 'n Filter-Wrapper (FW) benadering is geëvalueer. Die gebruik van optimaliseerde hiperparameter waardes het gelei tot 'n toename in toets akkuraatheid (van 0.8% tot 5.0%) vir beide RF en XGBoost. In die algeheel het RF beter presteer as XGBoost. In terme van voorspellende bevoegdheid en berekenings doeltreffendheid was die ontwikkelde FW benadering die mees suksesvolle veranderlike seleksie metode.

Die ontwikkelde masjienleer-afstandwaarnemende raamwerk benodig verder navorsing om sy doeltreffendheid te bevestig. Die tesis het egter sleutelnavorsingsvrae beantwoord, met die ontwikkelde raamwerk wat 'n vertrekpunt vir toekomstige studies verskaf.

TREFWOORDE

Afstandswaarneming; terrestriële hiperspektrale beelding; wingerde; waterstres; masjienleer; boom-gebaseerde klassifikasies; veranderlike seleksie; optimalisering van hiperparameter waardes

ACKNOWLEDGEMENTS

I sincerely thank:

- Mr Nitesh Poona, my supervisor, for his continued guidance, support and mentorship throughout the year.
- Dr Albert Strever, my co-supervisor, for his invaluable advice and insight.
- Berne Greyling, who helped immensely with my fieldwork and always contributed valuable input.
- The staff of the Department of Geography and Environmental Studies for helpful comments and constructive criticism.
- The Department of Viticulture and Oenology for providing the data needed to complete the research.
- The SIMERA technology group for providing the hyperspectral sensor.
- The National Research Foundation (NRF) for providing financial support during the duration of my master's degree.
- Winetech for their financial assistance.
- Ms Kelly McDowall for her thorough language editing.
- My fellow masters' students for their camaraderie and willingness to share ideas. You all made the long hours in the lab a little more bearable.
- Ms Juanita February for her willingness to always read my work.
- Ms Maylin Jansen for her solace and encouragement.

And, Most Importantly,

My Mother for Her Unwavering Love and Support.

Dream big. Start small. But most of all, start.

-Simon Sinek

CONTENTS

TABLES

FIGURES

ACRONYMS AND ABBREVIATIONS

AdaBoost	Adaptive Boosting
ANN	Artificial Neural Network
BPNN	Back Propagation Neural Network
CART	Classification and Regression Tree
CCD	Charge Couple Device
DN	Digital Number
EM	Electromagnetic
ENVI	Environment for Visualising Images
FPA	Focal Plane Array
FW	Filter-Wrapper
GBM	Gradient Boosting Machine
GDP	Gross Domestic Product
KNN	K-Nearest Neighbour
KW	Kruskal-Wallis
LAI	Leaf Area Index
LWP	Leaf Water Potential
MCCV	Monte-Carlo Cross Validation
MDA	Mean Decrease Accuracy
MDG	Mean Decrease Gini
MFL	Magnetic Flux Leakage
MMCE	Mean Misclassification Error
NGI	National Geo-Spatial Information
NIR	Near-infrared
OOB	Out-of-Bag

PCA	Principal Component Analysis
PLS	Partial Least Squares
PNN	Probabilistic Neural Network
PRI	Photochemical Reflectance Index
RF	Random Forest
RFE	Recursive Feature Elimination
ROI	Region of Interest
SBS	Sequential Backward Selection
SFFS	Sequential Floating Forward Selection
SFS	Sequential Forward Selection
SI	Spectral Indices
SNR	Signal-to-Noise Ratio
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
SWIR	Shortwave infrared
SWP	Stem Water Potential
UAS	Unmanned Aerial Systems
UV	Ultraviolet
VI	Variable Importance
VIS	Visible
VNIR	Visible and Near-infrared
XGBoost	Extreme Gradient Boosting

CHAPTER 1: INTRODUCTION

This chapter provides an introduction to the thesis. It presents background information to contextualise the study, outlining the research problem, study aim and objectives, research methodology, and research design.

1.1 BACKGROUND TO THIS STUDY

Precision viticulture, a subdivision of precision agriculture, entails the collection and analysis of spatial data to identify anomalies within vineyards (Matese & Di Gennaro 2015). Precision viticulture endeavours to produce site-specific management schemes to improve crop quality, production, and sustainability (Matese & Di Gennaro 2015; Mathews 2013). This increases the economic benefits of vineyard crops and reduces their negative impact on the environment (Mathews 2013; Mulla 2013).

Remote sensing applications in precision viticulture have proven to be a reliable tool for studying the spatial variability within vineyards (Baluja et al. 2012; Bellvert et al. 2014; Matese & Di Gennaro 2015). Due to the limitations in spatial and temporal resolutions often associated with conventional satellite or manned aerial platforms, studies have seen a marked increase in proximal (terrestrial) remote sensing techniques (Del-Moral-Martínez et al. 2016; Reis et al. 2012; Sanz et al. 2013). Proximal remote sensing entails the use of sensors mounted on various mobile or stationary platforms (Mulla 2013). Compared with satellite or manned aerial platforms, proximal remote sensing systems can provide greater spatial resolutions (centimetre resolution), which are less affected by atmospheric conditions (Candiago et al. 2015; Matese & Di Gennaro 2015). Moreover, proximal remote sensing techniques can acquire high temporal resolutions due to their easy in-field deployment and relatively inexpensive operational cost, allowing for real-time, site-specific management of irrigation, fertilisers, and pesticides (Matese & Di Gennaro 2015; Mulla 2013). These advantages of proximal remote sensing can be gainfully employed in precision viticulture and the broader agricultural field, where the monitoring of heterogeneous croplands necessitates short revisit times and high spatial resolutions (Mulla 2013).

Traditionally, remote sensing applications in precision viticulture have concentrated on the measurement of reflected radiation using multispectral sensors (Baluja et al. 2012; Candiago et al. 2015; Matese & Di Gennaro 2015). These sensors are limited in their abilities to detect fine spectral changes in vegetation due to their broad-band (greater than 40 nm) data collection, which primarily focuses on the visible (VIS) and near-infrared (NIR) regions of the electromagnetic (EM) spectrum (Mulla 2013; Wendisch & Brenguier 2013).

Hyperspectral remote sensing (spectroscopy) circumvents many of the challenges faced by traditional multispectral sensors. Hyperspectral remote sensing provides data-collection capabilities across a wider spectral range (typically 350-2500 nm) and at narrower spectral increments (typically 10 nm) (Wendisch & Brenguier 2013). Hyperspectral imaging offers a method for evaluating the spectral and spatial properties of vegetation, providing important variables regarding the biochemical and physiological properties of vegetation (Poona, van Niekerk & Ismail 2016). The continuous narrow-band characteristics of hyperspectral data provide more detailed spectral information, compared with conventional multispectral sensors (Mulla 2013). The increased dimensionality can be exploited to detect spectral differences more proficiently than broad-band multispectral data (Mulla 2013; Poona, van Niekerk & Ismail 2016).

The high dimensionality (spectral, temporal, and spatial) associated with remotely sensed data, such as hyperspectral imagery, presents a unique challenge for data analysis (Singh et al. 2016). However, the rapid growth in computing power experienced in recent years has facilitated the use of machine learning algorithms (Dev et al. 2016), which are capable of efficiently exploiting the information present in these complex datasets (Ali et al. 2015). Machine learning presents scalable and flexible frameworks for data analysis (Dev et al. 2016). These frameworks are adept at identifying patterns in large datasets by simultaneously analysing vast combinations of features, making machine learning approaches more efficient and ideal for vegetative stress detection (Singh et al. 2016). Machine learning approaches have been utilised in a variety of remote sensing applications, such as biomass and soil moisture retrievals (Ali et al. 2015), vegetative disease detection (Poona et al. 2016), and land cover classification (Pedergnana, Marpu & Mura 2013).

Random Forest (RF) (Breiman 2001) is a machine learning algorithm that has been successfully employed for hyperspectral data analysis (Abdel-Rahman et al. 2015; Adam et al. 2017; Poona, van Niekerk & Ismail 2016). RF utilises bootstrap aggregation (bagging) to create training samples, which are used to train an ensemble of independent decision trees (Breiman 2001). This ensemble method of classification has shown to improve model performance by aggregating the outcome of numerous weak decision trees (Belgiu & Drăguţ 2016).

Recently, another ensemble classifier called Extreme Gradient Boosting (XGBoost) (Chen & Guestrin 2016) has been utilised in various classification frameworks (for example, see Möller et al. 2016; Torlay et al. 2017; Xia et al. 2017). XGBoost builds on Gradient Boosting Machines (Friedman 2001) and has produced similar results to RF (Georganos et al. 2018a; Kejela & Rong 2016; Mohite et al. 2017). XGBoost employs boosting samples to iteratively re-train a multitude of decision trees, with each new tree attempting to minimise error by learning from the previously grown tree (Chen & Guestrin 2016).

Moreover, various studies have explored the utility of feature selection as a means to reduce the dimensionality of hyperspectral data (Abdel-Rahman et al. 2015; Pedergnana, Marpu & Mura 2013; Poona et al. 2016). Feature selection approaches aim to produce an optimal subset of wavebands that maximise target relevance and minimise redundant wavebands (Chandrashekar & Sahin 2014). These approaches generally improve model efficiency and lead to decreased computational complexity (Chandrashekar & Sahin 2014). Filter and wrapper approaches are the most common feature selection techniques implemented on hyperspectral datasets (Cao et al. 2017; Lagrange, Fauvel & Grizonnet 2017; Medjahed et al. 2016; Taşkın, Hüseyin & Bruzzone 2017). The filter approach evaluates the relevance and/or importance of wavebands independently from the learning algorithm employed, whereas wrappers are dependent on the feedback information provided by the selected learner (Chandrashekar & Sahin 2014).

1.2 PROBLEM STATEMENT

South Africa is one of the world's largest wine producers, producing 1.05 billion litres of wine in 2016 (SA Wine Industry Information & Systems 2016). Approximately 300 000 people were employed either directly or indirectly by the South African wine industry in 2015, with the industry contributing R36.1 billion to the national gross domestic product (GDP) in the same year (SA Wine Industry Information & Systems 2016). Historically, viticulture has been highly sensitive to changes in climate (Hannah et al. 2013), which is the primary determining factor of agricultural productivity (Nelson et al. 2014). With the effects of global climate change becoming more prominent, greater concern has been expressed regarding the negative impact climate change could have on viticulture production (Nelson et al. 2014).

To safeguard the sustainability and continued growth of the wine sector, it is important to ensure the health of vineyards (Matese & Di Gennaro 2015). This requires the collection of important variables, such as plant water status and plant water potential, which thus far has proven challenging to acquire (Karakizi, Oikonomou & Karantzalos 2016). While direct methods of data acquisition are more precise and accurate, they remain time-consuming and costly (Kalisperakis et al. 2015). Alternatively, remote sensing can provide a faster, less costly method of data acquisition (Mulla 2013).

Remote sensing application in precision viticulture has focused on a vast variety of endeavours, such as vineyard yield estimation (Font et al. 2015), vine variety discrimination (Karakizi, Oikonomou & Karantzalos 2016), and water stress modelling (Zarco-Tejada et al. 2013). The detection of water stress in vineyards is an integral part of many site-specific management systems (Bellvert et al. 2014), with water stress negatively affecting vegetative growth and grape quality (Costa et al. 2016; Kim et al. 2011). Scarce rainfall and high evapotranspiration rates are common in many wine-producing

countries (Baluja et al. 2012; García-Tejero et al. 2016). It is, therefore, imperative to characterise the spatial variability within vineyards to combat against overwatering or unintended water stressing in parts of the vineyard, thereby minimising water wastage (Baluja et al. 2012; Bellvert et al. 2014).

Numerous studies have used remote sensing in precision viticulture (Candiago et al. 2015; Font et al. 2015; Karakizi, Oikonomou & Karantzalos 2016), with a limited number of studies (for example, see Maimaitiyiming et al. 2017; Pôças et al. 2015; Ricci et al. 2016) having used spectroscopic data (field spectroscopy) to model vineyard performance. However, no studies to date have explored the utility of terrestrial hyperspectral imaging in combination with machine learning to model water stress in a Shiraz vineyard.

The need to investigate the use of terrestrial hyperspectral imaging for the proximal remote sensing of vineyard water stress resulted in the following research questions:

1. Can terrestrial hyperspectral imaging be used effectively to model water stress in a Shiraz vineyard?
2. Can the RF and XGBoost algorithms be used to successfully model water stress in a Shiraz vineyard?
3. Can feature selection and algorithm optimisation significantly improve model performance?

1.3 RESEARCH AIM AND OBJECTIVES

The aim of this study is to develop a remote sensing–machine learning framework for modelling water stress in a Shiraz vineyard using terrestrial hyperspectral imaging.

To accomplish this aim, the following objectives were set:

1. Evaluate the utility of terrestrial hyperspectral imaging to discriminate between stressed and non-stressed Shiraz vines.
2. Investigate the efficacy of the RF and XGBoost algorithms for modelling water stress in a Shiraz vineyard.
3. Explore the use of semi-automated algorithm optimisation and feature selection to improve model performance.

1.4 STUDY AREA

The study area, seen in Figure 1.1, is situated on the Welgevallen experimental farm in Stellenbosch (central coordinates: 33°56'38.5"S, 18°52'06.8"E). Stellenbosch forms part of the Cape Winelands situated in the Western Cape Province of South Africa. The region has a Mediterranean climate with warm, dry summers and wet, mild to cold winters (Yelenik, Stock & Richardson 2004). Stellenbosch

receives on average 800 mm of rainfall annually, with temperatures ranging from an average high of 27 ºC during summer and hardly dropping below 7 ºC during winter (Meadows 2003). The region is mountainous and the land is dominated by agricultural farms and residential areas (Conradie et al. 2002). The geology of Stellenbosch consists of sedimentary rock, of the Malmesbury group, with soil deposits comprising rich potassium-containing minerals, making the region conducive to vineyard growth (Conradie et al. 2002). Stellenbosch is home to more than 150 wine cellars, producing approximately 17% of South Africa's wine grape yield (SA Wine Industry Information & Systems 2016).

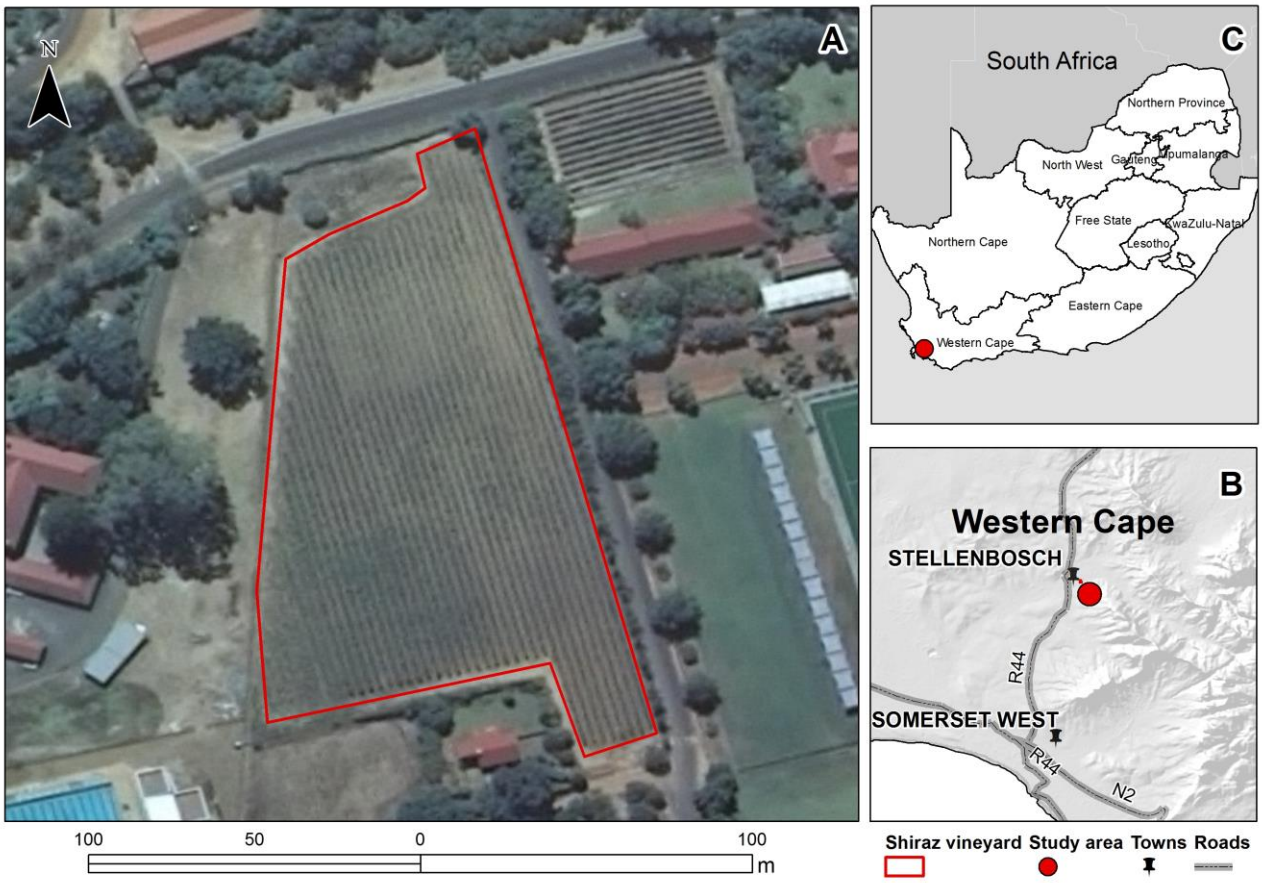

Figure 1.1 The Shiraz vineyard plot (A) situated on the Stellenbosch Welgevallen farm (B), in the Western Cape Province of South Africa (C). Inset map B shows the Shuttle Radar Topography Mission (SRTM) 90 m hillshade as background.

1.5 METHODOLOGY AND RESEARCH DESIGN

The research was conducted in a quantitative manner. Empirical methods were employed to achieve the objectives outlined in Section 1.3. The proposed methods utilise machine learning approaches and remotely sensed data to model water stress in a Shiraz vineyard. An overview of the research design is provided in Figure 1.2. Prior to data analysis, a field campaign was conducted to collect primary data samples. The data acquired consisted of terrestrial hyperspectral imagery, collected for a Shiraz vineyard. The research comprised two components. Component one investigated the utility of

terrestrial hyperspectral imagery, in combination with machine learning, to model vineyard water stress. Component two further explored the efficacy of feature selection and hyperparameter value optimisation to improve model performance.

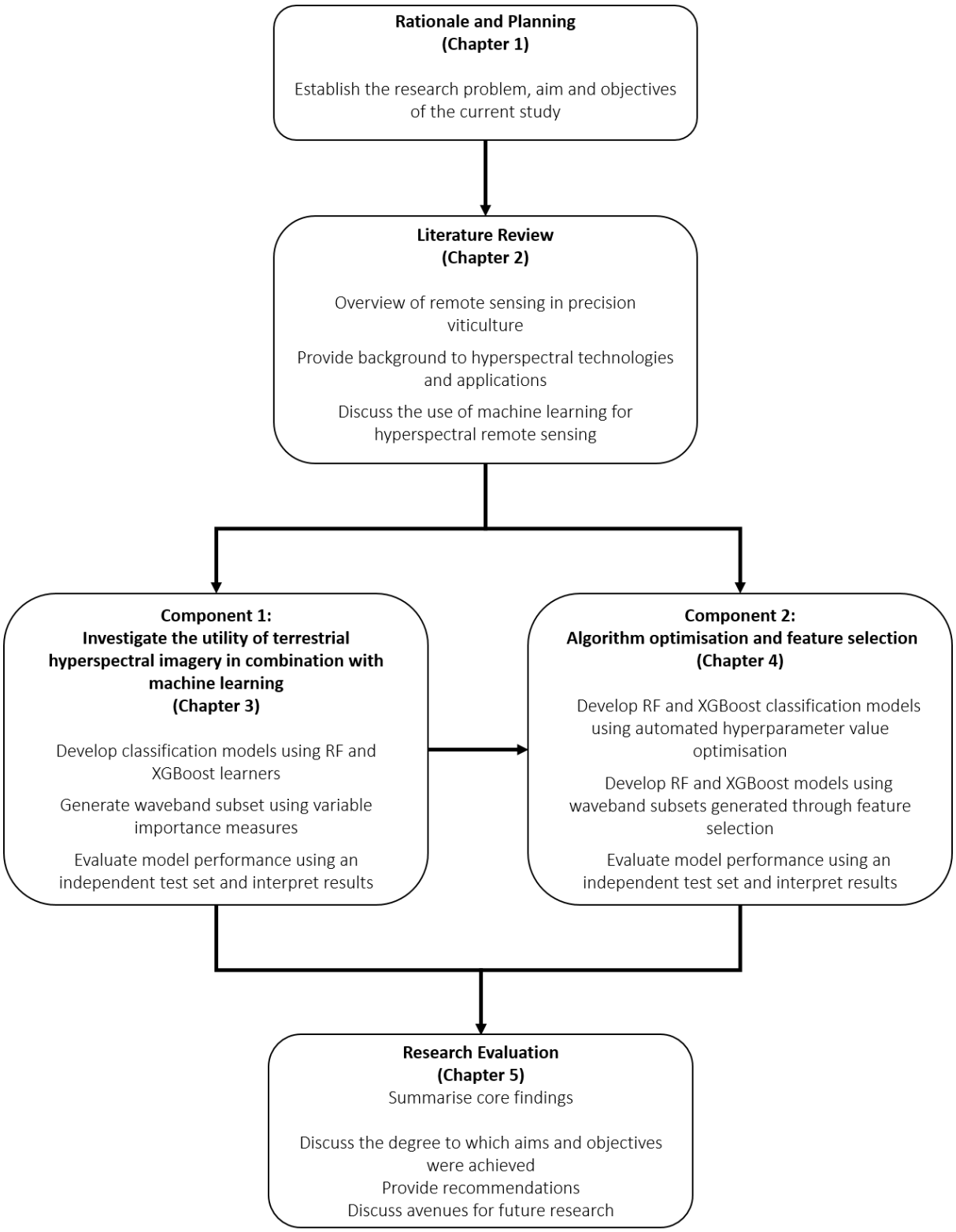

Figure 1.2 Research design for evaluating the utility of terrestrial hyperspectral imagery to model vineyard water stress using machine learning.

1.6 STRUCTURE OF THESIS

The research problem, aim, and objectives have been established in this chapter. The remainder of the thesis is structured as follows:

Chapter 2 outlines the applications of remotely sensed data in precision viticulture, with an emphasis on vineyard water stress. It highlights the benefits and drawbacks of hyperspectral data and provides a brief discussion on the use of machine learning algorithms in hyperspectral remote sensing.

The data collection for the first and second component as well as the methods and findings of component one are detailed in Chapter 3. Chapter 3 aims to develop a remote sensing-machine learning framework to discriminate between stressed and non-stressed Shiraz vines using terrestrial hyperspectral imagery. In so doing, it contributes towards research questions one and two.

Chapter 4 comprises the methods and findings of component two. Chapter 4 builds upon the semi-automated classification framework detailed in Chapter 3 and contributes towards the third research question. Feature selection was performed to determine waveband importance, enabling the creation of optimal waveband subsets. Hyperparameter value optimisation was conducted in an attempt to improve the algorithm accuracies achieved in Chapter 3.

It should be noted that Chapter 3 has been published as a research article in *Remote Sensing*. Chapter 4 was prepared as a manuscript for submission to the *International Journal of Remote Sensing*. Therefore, some similarity might arise in the respective chapters due to the same methods and data being used.

Chapter 5 concludes the thesis by summarising the key findings of both components. Furthermore, it revisits the research aim and objectives outlined in this chapter and provides recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

Remote sensing is a field of study associated with deriving physical information about surface objects from a distance (Eismann 2012). It is a cost-effective method used to capture specific data in a timely manner, thereby facilitating informative decision making (Eismann 2012). This chapter reviews the applications of remote sensing in precision viticulture. Additionally, it will discuss the literature pertaining to hyperspectral data and outline the utility of machine learning in hyperspectral remote sensing.

2.1 ROLE OF REMOTE SENSING IN PRECISION VITICULTURE

Remotely sensed data of plant growth, chlorophyll content, fruit quality, and soil moisture can provide valuable insight for applications in agriculture (Xue & Su 2017). The wine industry is one sector that has benefited immensely from the use of remote sensing in precision viticulture (Smit, Sithole & Strever 2016). The utility of remote sensing has seen it be used to map grapevine vigour (Matese et al. 2015; Matese, Di Gennaro & Berton 2016), monitor vine diseases (Al-Saddik, Simon & Cointault 2017; Di Gennaro et al. 2016), and determine the leaf area index (LAI) of vineyard canopies (Kalisperakis et al. 2015; Mathews & Jensen 2013). In precision viticulture, remote sensing facilitates smarter farming practices where the application of crop productive factors (such as fertilisers, pesticides, and water) are site-specific and applied when necessary (Matese & Di Gennaro 2015). Remote sensing practices are ideal for reducing farming costs and minimising the detrimental impact farming has on the environment.

2.1.1 Spectral response of vegetation

Applications of remote sensing centre on the recording of radiation (i.e. spectral signatures) reflected or emitted from agricultural vegetation or soil (Mulla 2013). Typically, spectral reflectance is captured across the VIS (400 to 700 nm), NIR (700 to 1300 nm), and shortwave infrared (SWIR) (1300 to 2500 nm) portions of the EM spectrum (Wendisch & Brenguier 2013). These spectral signatures can be quantitatively analysed to gain valuable insights into plant health and yield quality (Bioucas-dias et al. 2013).

As plants develop, their growth and exposure to various stressing mechanisms affect their spectral properties (Usha & Singh 2013). Plants that are considered healthy strongly absorb radiation in the VIS region of the EM spectrum (Kim et al. 2011). This absorption of VIS radiation is mainly due to chlorophyll pigments of green leaves, which absorb 70 to90% of radiation (Usha & Singh 2013), and carotenoid pigments, which are known for absorbing radiation in the blue region of the EM spectrum

(Zygielbaum et al. 2009). The absorption of VIS radiation by healthy plants has formed the basis for detecting stressed vegetation (Kim et al. 2011; Usha & Singh 2013). Plants under stress have shown to produce greater reflectance in the VIS region due to decreased concentrations in pigments such as chlorophyll (Kim et al. 2011; Usha & Singh 2013; Zygielbaum et al. 2009). For example, Al-Saddik, Simon & Cointault (2017) exploited the increase in reflected VIS radiation to detect *flavescence dorée* disease in grapevines. The study utilised VIS wavebands and found that these wavebands could accurately detect the *flavescence dorée* disease in grapevines, producing classification accuracies greater than 90%.

In contrast, plants reflect the majority of radiation in the NIR region of the EM spectrum (Usha & Singh 2013). NIR reflectance is mainly due to dense leaf canopies or soil profiles (Usha & Singh 2013). Consequently, the NIR region has been successfully used to assess various attributes pertaining to plant canopies, vineyard soils, and cultivar discrimination. For example, NIR reflectance was utilised by Gutiérrez et al. (2016) to identify grape varieties. Lopo et al. (2018) reported the use of NIR wavebands to classify vineyard soil samples.

Spectral absorption of SWIR radiance is predominantly due to the water content in healthy plant leaves (Gerhards et al. 2016). Several SWIR wavebands have been shown to correlate with in-field measures of plant water status, such as stomatal conductance (Gerhards et al. 2016; Govender et al. 2009; Rodríguez-Pérez et al. 2007). González-Fernández et al. (2015) utilised SWIR reflectance to determine leaf water content in commercial vineyards.

2.1.2 Sensor platforms

Remote sensing applications are generally classified according to the sensor platform used (Mulla 2013). These platforms can be spaceborne (i.e. satellites), airborne (manned or unmanned), or terrestrial (proximal). The use of satellite imagery in precision viticulture has been extensively researched, with applications ranging from estimating spatial patterns in vine growth using 5 m RapidEye imagery (Matese et al. 2015) to using 30 m Landsat-8 imagery for monitoring vineyard evapotranspiration rates (Semmens et al. 2016).

Modern-day satellite platforms, such as GeoEye-1, WorldView-3, and the RapidEye five satellite constellation, have facilitated the collection of high spatial (0.3-6.5 m) and temporal (1-3 days) resolution multispectral imagery. Unfortunately, high-resolution satellite imagery can be quite costly (Matese & Di Gennaro 2015), especially for use in developing countries (Costa et al. 2016). Alternative satellite platforms, such as Sentinel-2, provide free data for the masses. However, the spatial resolution of these satellites is often not sufficient for application in precision viticulture due to the narrow spacing (typically from 1.4 to 2.1 m) of vines (Matese & Di Gennaro 2015).

Furthermore, the use of satellite imagery captured across the VIS and NIR wavelengths is also limited to cloud-free days (Matese & Di Gennaro 2015).

Advancement in technology has seen the use of unmanned aerial systems (UAS) growing in popularity. For example, Mathews & Jensen (2013) utilised UAS imagery to estimate biomass in vineyards. Similarly, Candiago et al. (2015) employed high spatial resolution (0.5-10 cm) UAS imagery (comprising green, red and NIR wavebands) to model vine vigour. UAS technology facilitates inexpensive data collection of very high spatial resolution (sub-meter resolution) imagery (Matese et al. 2015). However, the use of UAS technology remains highly regulated (Costa et al. 2016), which limits its utility to specific locations and applications. Moreover, the small payload and short flight times associated with many UAS platforms limit its implementation in precision viticulture (Matese et al. 2015). Matese et al. (2015) further assert that UAS solutions remain a low-cost source of remote sensing data for small areas (approximately five hectares) only; satellite platforms provide more cost-effective imagery for larger areas.

More recently, there has been a growing interest in real-time, on-the-go monitoring with terrestrial sensors (i.e. proximal remote sensing). Similar to UAS platforms, proximal remote sensing offers cost-effective data-acquisition models that are easily deployable and less restricted in use when compared with UAS platforms. Applications of proximal remote sensing include vine LAI determination using mobile terrestrial laser scanning (Del-Moral-Martínez et al. 2016; Sanz et al. 2013), detection of grape bunches using terrestrial imaging (Reis et al. 2012), and mapping vineyard productivity using videography (Tang et al. 2016).

2.1.3 Vineyard water stress

An important facet of remote sensing in precision viticulture concerns the collection of data pertaining to vine water status (Costa et al. 2016). Traditionally, monitoring vineyard water stress has relied on the acquisition of in-field measurements, such as Stem Water Potential (SWP) (Deloire & Heyms 2011), of specific vines or through analysing soil moisture samples (Rogiers et al. 2012). These conventional methods, though accurate for the sampled vine and/or vineyard zones, are laborious, destructive, and inept for automation (Ihuoma & Madramootoo 2017). Furthermore, the traditional plant-based methods often assume that plant density and transpiration rates are uniform across the field (Matese et al. 2018). This is rarely the case, due to the heterogeneity in soil and vegetation (Ihuoma & Madramootoo 2017). In comparison, remote sensing techniques offer an affordable, less time-consuming alternative that easily lends itself to automation (Chirouze et al. 2014).

Numerous studies have confirmed the utility of remote sensing as a medium for estimating various indirect parameters that are known to be indicative of vine water status. For example, Pôças et al.

(2015) reported the use of leaf reflectance and regression analysis to estimate predawn Leaf Water Potential (LWP) in irrigated vineyards. The study found VIS and NIR (VNIR, 400-1300 nm) wavebands strongly correlated with in-field predawn LWP measurements and could therefore be used to accurately predict LWP. Similarly, Beghi, Giovenzana & Guidetti (2017), Cancela et al. (2017), and Maimaitiyiming et al. (2017) utilised leaf reflectance to predict other well-known indicators of vine water status, such as SWP and stomatal conductance.

Another popular use of remotely sensed spectral data is the development of spectral indices (SI). SIs aim to exploit the contrast in reflectance of two or more wavebands in order to measure the relative abundance of a given substance (i.e. water content or vegetative growth) within a given vineyard (Ihuoma & Madramootoo 2017). Mixed results regarding the effectiveness of SIs have been reported. Zarco-Tejada et al. (2013) utilised the Photochemical Reflectance Index (PRI), one of the most popular indices used in the broader agricultural field, as an indicator of water stress in vineyards. Their study reported that PRI could not accurately track the diurnal dynamics of stomatal conductance and water potential and was, therefore, a poor indicator of water stress in vines. Similar findings were reported by Baluja et al. (2012) and Maimaitiyiming et al. (2017). Ihuoma & Madramootoo (2017) highlighted the utility of various SIs but stated that most well-known SIs are highly sensitive to the confounding absorption of photosynthetic pigments, soil profiles, and canopy structures.

A review of the literature indicated that the majority of studies have concentrated on the reflectance and/or absorbance of radiation by vine leaves to determine vine water status. However, studies have also reported the use of remote sensing techniques to record the emittance of radiation from vine leaves to quantify vineyard water stress. The recording of emitted radiation predominantly concerns the acquisition of infrared thermometry or thermal imagery to detect vine canopy temperatures (Ihuoma & Madramootoo 2017). The use of thermal data to detect water stress is based on the process of evapotranspiration. Canopy temperatures increase as vines absorb solar radiation, but these temperatures decrease as the radiation is used to fuel evapotranspiration (Semmens et al. 2016). Water-stressed vines have lower evapotranspiration rates (Ihuoma & Madramootoo 2017) and therefore emit higher temperatures from their leaves (Semmens et al. 2016). This mechanism of evapotranspiration has been exploited by numerous studies to detect water-stressed vines (Baluja et al. 2012; Bellvert et al. 2014; García-Tejero et al. 2016; Matese et al. 2018; Zarco-Tejada et al. 2013).

2.2 HYPERSPECTRAL REMOTE SENSING

Compared with multispectral remote sensing, hyperspectral remote sensing, also known as imaging spectroscopy, adopts traditional spectroscopy methodologies and merges them with high spatial resolution imaging (Eismann 2012). Hyperspectral data is defined by high spectral resolution

comprising hundreds of narrow (typically < 10 nm) contiguous spectral wavebands (Poona, van Niekerk & Ismail 2016). The narrow bandwidth characteristics of hyperspectral data are what sets it apart from traditional broad-band multispectral sensors. The wide spectral range (350-2500 nm) and narrow wavebands of hyperspectral sensors (Eismann 2012) make it ideal for in-depth examination and discrimination of heterogeneous objects or scenes captured of the Earth's surface (Bioucas-dias et al. 2013). As a result, hyperspectral sensors provide greater utility in terms of use and applications (Eismann 2012; Wendisch & Brenguier 2013).

Typically, hyperspectral sensors utilise a 2D matrix array in the form of a Charge Couple Device (CCD) or Focal Plane Array (FPA) (Wendisch & Brenguier 2013). These sensors record spatial data on a 2D axis (i.e. x and y-axis) and radiance on a third spectral axis (i.e. z-axis), producing what is known as a 3D hypercube (Eismann 2012). A hypercube is generally constructed in a progressive manner, i.e. spatial images are recorded sequentially at different wavelengths or a scene is recorded as sequential swaths that are a pixel wide and multiple pixels long (Eismann 2012; Wendisch & Brenguier 2013). These hyperspectral sensors have been found to produce near-laboratory-quality radiance measures collected predominantly across the VNIR (400-1300 nm) and SWIR (1300-2500 nm) regions of the EM spectrum (Wendisch & Brenguier 2013).

The unprecedented quality of in-field spectral data, coupled with high spatial and temporal resolutions, has enabled a myriad of applications for hyperspectral remote sensing. Its utility has been particularly useful for applications in agriculture and forestry. For example, Abdel-Rahman et al. (2014) detected disease-infected pine trees using VNIR hyperspectral data (bandwidth ranged from 2 to 4 nm). Vélez Rivera et al. (2014) employed 10 nm NIR hyperspectral imaging for the early detection of machine damage in mango crops. Similarly, Carreiro Soares et al. (2016) reported the use of NIR hyperspectral imaging, with a 6 nm spectral resolution, to classify cottonseeds. Rivera-Caicedo et al. (2017) exploited the utility of VNIR and SWIR airborne hyperspectral imaging (bandwidth ranged between 11 and 21 nm) to map crop LAI.

More specifically, within precision viticulture, Zarco-Tejada, González-Dugo & Berni (2012) reported the use of hyperspectral imaging for estimating vineyard water stress. Their study found spectral indices derived from hyperspectral data could moderately predict vine stomatal conductance and water potential, producing r^2 values of 0.66 and 0.67. Kalisperakis et al. (2015) estimated vine LAI by employing UAS hyperspectral imaging. An r^2 value of 0.81 was reported when employing regression analysis on hyperspectral LAI estimates and in-field LAI measurements. Gutiérrez et al. (2016) employed NIR hyperspectral sensing to classify different vine cultivars. The authors reported an average classification accuracy of 88.7% when discriminating between ten grapevine varieties.

2.2.1 Spectral smoothing

The conditions for capturing high-quality hyperspectral data are seldom optimal (Prasad et al. 2015). Variability in solar illumination, atmospheric gases, and aerosols all impact negatively on spectral quality, reducing the signal-to-noise ratio (SNR) of a given sensor (Wendisch & Brenguier 2013). Noise, produced through external atmospheric conditions or self-generated by the sensor (Prasad et al. 2015), is inherently present in spectral signatures (Wendisch & Brenguier 2013). Consequently, hyperspectral data pre-processing has often incorporated an additional spectral smoothing step (Liu et al. 2016; Prasad et al. 2015; Schmidt & Skidmore 2004).

Numerous spectral smoothing algorithms have been explored in the literature. These include median filters (Vélez Rivera et al. 2014), moving averaging filters (Beghi, Giovenzana & Guidetti 2017; Prasad et al. 2015), and wavelet decomposition (Schmidt & Skidmore 2004). The Savitzky-Golay filter (Savitzky & Golay 1964) is the most commonly used spectral smoothing algorithm employed within remote sensing. Savitzky-Golay is a simplified filter that employs least squares convolution[1] for spectral smoothing (Savitzky & Golay 1964). The Savitzky-Golay filter has been successfully used to minimise noise and unwanted light scattering in both laboratory and field-based spectra (Gutiérrez et al. 2016; Liu et al. 2016; Lopo et al. 2018; Prasad et al. 2015).

2.2.2 Statistical challenges

It is evident from the literature that the calibre of high dimensional data provided by hyperspectral remote sensing has enabled greater quantitative analysis of the Earth's surface. However, the high dimensionality of hyperspectral data poses significant challenges to traditional statistical analysis (Camps-Valls et al. 2014). Hyperspectral data is inherently plagued by the so-called "curse of dimensionality" (Poona et al. 2016), which leads to the Hughes phenomenon (Hughes 1968) and ultimately to reduced classification results (Georganos et al. 2018a).

In classification-driven applications, the expansion of spectral dimensions over a finite number of training samples tends to deteriorate classifier accuracy (Georganos et al. 2018a). The collection of training data is application-specific and laborious, which makes it time-consuming and expensive; hence the limited number of training samples available in supervised classification frameworks (Georganos et al. 2018a). Additionally, random variations within high dimensional datasets (Georganos et al. 2018a) and redundancy among the large number of neighbouring wavebands

[1] Convolution is defined as a weighted moving averaging filter, where the weighting is given as a polynomial equation of a given degree (Jung & Ehlers 2016).

(Santara et al. 2017) lead to overfitting the training model, producing models that perform poorly on independent test sets (Pappu & Pardalos 2014). The large number of wavebands also facilitate the creation of complex models, which demand greater computational expense and are often difficult to interpret (Georganos et al. 2018a); hence the need for dimensionality reduction.

2.3 DIMENSIONALITY REDUCTION

Dimensionality reduction methods aim to circumvent the curse of dimensionality by reducing the number of irrelevant and/or redundant wavebands (Thorp et al. 2017) without significantly reducing predictive prowess (Chandrashekar & Sahin 2014). Two main strategies for dimensionality reduction exist, namely feature extraction and feature selection.

2.3.1 Feature extraction

Feature extraction methods reduce dimensionality by transforming the original waveband dataset into a set of new features (Lagrange, Fauvel & Grizonnet 2017). These features are produced by summarising the most informative features in lower dimensional space (Rivera-Caicedo et al. 2017). As such, the need to search for the most relevant wavebands is eliminated and the number of training features is significantly reduced (Lagrange, Fauvel & Grizonnet 2017; Rivera-Caicedo et al. 2017).

Principal Component Analysis (PCA) (Jolliffe 1986), and its various extensions, such as Partial Least Squares (PLS), is one of the most commonly applied feature extraction methods reported in the literature. For example, Rivera-Caicedo et al. (2017) utilised both PCA and PLS for biophysical variable retrieval from hyperspectral data. Similarly, Cheng et al. (2004) utilised PCA to extract hyperspectral wavebands to model cucumber chilling damage. However, the use of PCA is limited, as it is designed to only account for the linear relationship between the features and the target variable (Rivera-Caicedo et al. 2017). Therefore, PCA can produce unsatisfactory results when applied to features that exhibit non-linear relationships (Rivera-Caicedo et al. 2017).

2.3.2 Feature selection

Alternatively, feature selection methods facilitate dimensionality reduction by selecting a subset of input wavebands that have been identified as either relevant or important (Chandrashekar & Sahin 2014). Feature selection methods preserve the originality of the input dataset, unlike feature extraction methods, providing better interpretability for end-users (Lagrange, Fauvel & Grizonnet 2017). Feature selection methods are generally categorised into filter and wrapper approaches.

2.3.2.1 Filters

Filter methods employ feature ranking techniques to filter out the irrelevant wavebands (Lagrange, Fauvel & Grizonnet 2017). A ranking criterion, rather than performance of a given classifier (Chandrashekar & Sahin 2014), is used to measure the correlation between each waveband and a specific output class (Taşkın, Hüseyin & Bruzzone 2017). An importance score or weight is assigned to all the input wavebands based on their usefulness to discriminate between different classes (Chandrashekar & Sahin 2014). A user-defined threshold value is then employed to select wavebands based on their importance scores (Radovic et al. 2017). Filter methods are advantageous as they are computationally inexpensive and produce waveband subsets that can be utilised across multiple classification algorithms (Lagrange, Fauvel & Grizonnet 2017). However, as filter methods are implemented independently from the classifier (i.e. ignores classifier performance), they do not directly optimise classification accuracy (Lagrange, Fauvel & Grizonnet 2017).

Numerous filter methods, such as ReliefF (Robnik-Sikonja & Kononenko 2003), chi-square (Liu & Setiono 1995), Fisher (Jensen, El-Sharkawi & Marks 2001), and information gain (Lewis 1992) have appeared in the literature. Jung & Ehlers (2016) reported the use of ReliefF feature selection to reduce dimensionality in hyperspectral datasets. Mean accuracies ranging from 82.0% to 95.0% (Kappa ranged from 0.79 to 0.94) were reported for the ReliefF produced subsets. Taşkın, Hüseyin & Bruzzone (2017) tested the utility ReliefF, chi-square, Fisher, and information gain methods on different hyperspectral datasets. Their study found no single filter method outperformed the others. The authors concluded that the performance of a given filter is dataset-dependent. To date, no guideline exists for selecting the most appropriate filter method.

2.3.2.2 Wrappers

Wrapper methods aim to produce optimal waveband subsets for a given classification algorithm (Poona et al. 2016). Wrappers utilise classifiers as black box predictors and classifier performance as an objective function[1]. Once the objective function has been defined, feature selection is reduced to a searching problem (Chandrashekar & Sahin 2014), which detects optimal waveband subsets (Jović, Brkić & Bogunović 2015). The predefined classifier then evaluates the subsets (Chandrashekar & Sahin 2014). This process is iterated until a given subset maximises the objective function. Various

[1] A function that evaluates candidate subset performance, based on a given measure of "goodness", e.g. classification accuracy (Chandrashekar & Sahin 2014).

searching algorithms have been developed that can be broadly categorised into exhaustive and heuristic searching methods (Waad, Ghazi & Mohamed 2013).

Exhaustive search methods, also known as complete search methods (Jović, Brkić & Bogunović 2015), find all candidate waveband subsets and evaluate each subset to identify the optimal combination of wavebands (Waad, Ghazi & Mohamed 2013). Exhaustive methods guarantee optimisation of the objective function, as they examine all possible solutions (Datta, Ghosh & Ghosh 2017). However, these methods are computationally expensive, prone to overfitting and become exponentially more impractical as the number of input wavebands increases (Waad, Ghazi & Mohamed 2013).

Heuristic search methods have been proposed for feature selection, as they are less computationally expensive than complete searches (Chandrashekar & Sahin 2014). Heuristic searches evaluate different waveband subsets to optimise the objective function (Chandrashekar & Sahin 2014). However, heuristic searches are deemed suboptimal as they do not evaluate all possible subsets and therefore cannot guarantee the selection of the most optimal waveband subset (Datta, Ghosh & Ghosh 2017). Sequential searches are one of the most popular heuristic methods employed in the literature (Fu et al. 2017; Jung & Ehlers 2016; Lagrange, Fauvel & Grizonnet 2017). Sequential searches incrementally generate waveband subsets in two ways: by adding wavebands to an empty subset one by one, known as sequential forward selection (SFS), or by removing wavebands one by one from the complete set of input data, known as sequential backward selection (SBS) (Chandrashekar & Sahin 2014).

Overall, wrapper methods produce better predictive accuracies when compared with filter methods (Cao et al. 2017; Cen et al. 2016; Medjahed et al. 2016). However, as wrapper methods require the training of a given classifier and a large number of labelled samples (Cao et al. 2017), they are more time-consuming and their complexity necessitates longer processing times (Medjahed et al. 2016). Furthermore, as wrappers are classifier-dependent, their subsets are generally not optimal across different classification algorithms (Chandrashekar & Sahin 2014). Nevertheless, wrapper methods have been successfully employed to reduce the dimensionality of hyperspectral datasets. Recently, Cen et al. (2016) employed the SFS wrapper to select hyperspectral wavebands that are optimal for the detection of chilling injury in cucumbers. The SFS-derived subsets produced classification accuracies above 95.0%. Furthermore, the SFS wrapper reduced dataset dimensionality by more than 90.0%. Poona et al. (2016) reported a testing error of 23.0%, using only 21.0% of the original waveband dataset when employing Recursive Feature Elimination (RFE).

2.4 CLASSIFICATION

Hyperspectral data classification is key to understanding and exploiting the wealth of information provided by high dimensional datasets. Classification algorithms aim to assign unique labels to each image pixel or spectral signature (Bioucas-dias et al. 2013). The high dimensionality, limited training samples, and the non-normal distribution of hyperspectral data (Belgiu & Drăguţ 2016) have rendered traditional parametric classifiers, such as Gaussian maximum likelihood, unreliable and obsolete (Pappu & Pardalos 2014). Consequently, a need for accurate hyperspectral classification frameworks exists. These frameworks should enable practical implementation, be simple to interpret, and effortlessly transferred across various applications.

The rapid increase in computer processing power over the last decade has paved the way for machine learning classifiers to become the standard paradigm for the analysis of remotely sensed hyperspectral data (Dev et al. 2016). Support vector machines (SVM) (Qiao et al. 2018; Wu et al. 2016), k-nearest neighbour (KNN) (Chen et al. 2018; Shuaibu et al. 2018), and artificial neural networks (ANN) (Patteti, Samanta & Chakravarty 2015; Rojas-Moraleda et al. 2017) are popular machine learning classifiers employed on hyperspectral datasets. Although these methods have been shown to produce accurate classification results, they are generally processing-intensive and complex (Belgiu & Drăguţ 2016; Raczko & Zagajewski 2017).

2.4.1 Ensemble learning

Recently, ensemble learning methods have gained considerable recognition in the literature for the classification of hyperspectral data (Abdel-Rahman et al. 2015; Mohite et al. 2017; Pedergnana, Marpu & Mura 2013; Poona, van Niekerk & Ismail 2016). Ensemble methods are supervised learning algorithms that fall within the realm of machine learning. The main premise behind ensemble methods is to combine a multitude of weak learners to produce a classifier that is predictively more accurate and reliable (Poona, van Niekerk & Ismail 2016). Several machine learning ensembles exist; chief among them are the bagging and boosting ensemble methods. Bagging and boosting ensembles often incorporate decision trees as a base learner in classification frameworks.

2.4.1.1 Decision tree ensembles

Decision tree-based ensembles are the most popular machine learning algorithms employed in hyperspectral classification frameworks (Abdel-Rahman et al. 2015; Knauer et al. 2017; Pedergnana, Marpu & Mura 2013; Poona, van Niekerk & Ismail 2016). The classification and regression tree (CART) algorithm (Breiman et al. 1984) is a popular example of a decision tree ensemble. CART is a univariate, non-parametric classifier that iteratively subsets the training data and then sequentially

applies a set of binary rules to discriminate between different classes (Breiman et al. 1984). This binary partitioning of CART models is useful for the identification of key explanatory wavebands (Goel et al. 2003). The tree-based framework of CART has been widely used in hyperspectral remote sensing applications. For example, CART has been implemented to detect weed stress and nitrogen status in corn crops (Goel et al. 2003), identify tree species (Shafri, Suhaili & Mansor 2007), and map wetland weed infestation (Andrew & Ustin 2008).

2.4.1.2 Bagging

Bagging (Breiman 1996) methods, also known as bootstrap aggregation, produce an ensemble learner by training numerous machine learning classifiers on different subsets of the training data (Belgiu & Drăguţ 2016). Bagging resamples the original training data by randomly selecting samples with replacement, i.e. the same sample can be selected for different subsets and duplicated within the same subset (Breiman 1996). The randomisation of the resampling procedure creates a diverse ensemble of classifiers (Breiman 1996). The final ensemble prediction is produced by averaging the results of all the individual classifiers (Breiman 1996). Shuaibu et al. (2018) recently detected fungal disease on apple tree leaves using hyperspectral data. Their study found a bagged ensemble (84.3%) outperform both decision tree (79.8%) and KNN (71.3%) classifiers.

2.4.1.3 Random forest (RF)

RF, developed by Breiman (2001), is an advanced bagging method. RF iteratively trains an ensemble of CART trees, used as weak base learners, on bagging generated subsets (Belgiu & Drăguţ 2016). RF inherently splits the training data into a separate train (2/3 of the input samples) and test set (1/3 of the input samples) (Breiman 2001). This split of the training samples enables RF to estimate an internal measure of model performance, known as the “out-of-bag” (OOB) error. RF has been shown to be insensitive to noise and redundant features, and resistant against overfitting (Belgiu & Drăguţ 2016). These characteristics of RF have empowered the algorithm to become one of the most popular ensembles used in hyperspectral classification frameworks. For example, Harrison, Rivard & Sánchez-Azofeifa (2018) and Maschler, Atzberger & Immitzer (2018) employed RF to discriminate between tree species. Adam et al. (2017) and Poona et al. (2016) utilised RF for vegetative disease detection.

The use of RF in precision viticulture has recently gained recognition in literature. Poblete-Echeverría et al. (2017) exploited the utility of RF to detect vine canopies. The authors observed a classification accuracy of 94.0% (Kappa = 0.91) for RF. Knauer et al. (2017) reported the use of RF and terrestrial hyperspectral imaging to identify Powdery Mildew on grapes. RF yielded an overall accuracy of 87.0%. Similar results were found by Sandika et al. (2016).

2.4.1.4 Boosting

Similar to bagging, boosting (Schapire 1990) attempts to boost prediction accuracy by iteratively training a multitude of learners on different instances of the training data and then combines the output of all the training models (Schapire 1990). However, unlike bagging, boosting iteratively builds models where the training of subsequent learners is dependent on the results of the previous learner (Pappu & Pardalos 2014). Furthermore, boosting weights a model's contribution by its predictive accuracy, rather than assigning equal weights to all models (Schapire 1990). Monteiro et al. (2009) employed a variant of boosting, known as LogitBoost, for hyperspectral classification of ore-bearing rocks and found boosting (97.1%) to outperform SVM (95.0%).

2.4.1.5 Adaptive boosting (AdaBoost)

AdaBoost (Freund & Schapire 1996) employs the boosting algorithm and has been widely used in hyperspectral studies (Chan & Paelinckx 2008; Kawaguchi & Nishii 2007; Xia et al. 2014). AdaBoost assigns a greater weight to samples that have been misclassified in the previous iteration, decreasing the weightings of correctly classified samples (Möller et al. 2016). This enables the new learner in the following iteration to adapt and specifically concentrate on correctly classifying the previously misclassified samples. A final weighted sum is applied to all the predictions to produce the final classification result (Freund & Schapire 1996).

2.4.1.6 Gradient boosting machines (GBM)

GBM (Friedman 2001) combines many weak learners to optimise a user-defined objective function. GBM is a unique boosting variant as it utilises a gradient descent scheme to minimise the loss function, which measures the loss in accuracy brought on by inaccurate predictions made by the base learner (Friedman 2001). GBM differs from AdaBoost as it does not increase the weight of misclassified samples before training a new model; rather, each learner is trained on the remaining error of the previous model (Friedman 2001). GBM classification has been used for hyperspectral applications, such as mapping invasive plant species (Lawrence, Wood & Sheley 2006) and for the discrimination of land cover parcels (Lawrence et al. 2004).

2.4.1.7 Extreme gradient boosting (XGBoost)

XGBoost (Chen & Guestrin 2016) is an advanced implementation of GBM. Similar to GBM, XGBoost follows an iterative scheme where decision trees are grown in each iteration of the boosting algorithm (Chen & Guestrin 2016). However, XGBoost builds on gradient boosting by incorporating regularisation. A regularisation term, which is added to the normal GBM loss function, penalises model complexity and reduces the contribution of individual weak learners to avoid overfitting (Xia

et al. 2017). XGBoost has shown great promise for the classification of high dimensional data (Luo et al. 2018; Martinez-de-Pison et al. 2017; Möller et al. 2016). For example, Georganos et al. (2018a) recently discriminated between various land cover classes using high-dimensional datasets and the XGBoost classifier. The authors reported a 77.8% overall accuracy for XGBoost when using all features (p = 169) as input to classification.

The use of XGBoost in precision viticulture, as well as the broader agricultural field, is an emerging field of study. Mohite et al. (2017) was the first known study to employ XGBoost in precision viticulture. Their study detected pesticide residue on grapes using hyperspectral data and XGBoost classification. Classification accuracies ranging from 81.6% to 87.6% were reported in the study. The findings reported by Mohite et al. (2017) demonstrated the feasibility of XGBoost for the classification of hyperspectral data within the context of precision viticulture.

2.4.2 Hyperparameter optimisation

Hyperparameter value optimisation is an essential component of classification frameworks, as many machine learning algorithms are sensitive to hyperparameter settings (Xia et al. 2017). Manual optimisation solutions have been rendered obsolete with the advent of machine learning algorithms, such as XGBoost, where the optimisation of several hyperparameter values is required. This has led to the development of new hyperparameter value optimisation methods, such as Bayesian optimisation algorithms (Martinez-de-Pison et al. 2017; Xia et al. 2017). To date, the traditional grid search method is one of the most popular optimisation techniques employed in the literature (Abdel-Rahman et al. 2015; Eisavi et al. 2015; Georganos et al. 2018a).

2.5 LITERATURE SUMMARY

Remote sensing provides numerous advantages for precision viticulture; chief among them being the non-destructive acquisition of important productivity variables that aid precise management schemes. Due to the difficulties associated with traditional sources of remote sensing data, such as satellite and aerial platforms, there has been increasing interest in the use of proximal (terrestrial) remote sensing techniques.

According to the literature, methods for modelling vineyard water stress have predominantly concentrated on the use of multispectral data or manual labour. The use of hyperspectral remote sensing offers a unique solution to modelling water stress, as the narrow waveband characteristics enable superior quantitative analysis of a vineyard's physiological response to stress. Furthermore, with the advent of machine learning ensembles and feature selection techniques, more accurate and efficient analysis of hyperspectral data is now possible. However, the combined utility of terrestrial

hyperspectral data and ensemble classification has, to date, not been employed for vineyard water stress modelling, presenting a possible gap in the literature. Therefore, the present study set out to investigate the use of ensemble hyperspectral classification for the modelling of vineyard water stress.

The following chapter, Chapter 3, concentrates on the use of RF and XGBoost ensembles for the classification of terrestrial hyperspectral imagery. Chapter 4 of this thesis focuses on feature selection approaches and hyperparameter value optimisation.

CHAPTER 3: MODELLING WATER STRESS IN A SHIRAZ VINEYARD USING HYPERSPECTRAL IMAGING AND MACHINE LEARNING[1]

3.1 ABSTRACT

The detection of water stress in vineyards plays an integral role in the sustainability of high-quality grapes and prevention of devastating crop loses. Hyperspectral remote sensing technologies, combined with machine learning, provide a practical means for modelling vineyard water stress. In this study, two ensemble learners were utilised, namely Random Forest (RF) and Extreme Gradient Boosting (XGBoost), to discriminate between stressed and non-stressed Shiraz vines using terrestrial hyperspectral imaging (473-708 nm). Additionally, the study evaluated the utility of a spectral subset of wavebands, derived using RF mean decrease accuracy (MDA) and XGBoost gain. The results show that both ensemble learners can effectively analyse the hyperspectral dataset. When using all wavebands (p = 176), RF produced a test accuracy of 83.3% (KHAT = 0.67), and XGBoost a test accuracy of 80.0% (KHAT = 0.6). Using the subset of wavebands (p = 18) produced slight increases in accuracy, ranging from 1.7% to 5.5% for both RF and XGBoost. The study further investigated the effect of smoothing the spectral data using the Savitzky-Golay filter. The results indicated that the Savitzky-Golay filter reduced model accuracies (ranging from 0.7% to 3.3%). The results demonstrate the feasibility of terrestrial hyperspectral imagery and machine learning to create a semi-automated framework for vineyard water stress modelling.

3.2 INTRODUCTION

Water stress in vineyards is a common phenomenon that occurs in the Western Cape of South Africa during the summer (Costa et al. 2016). Water stress promotes stomatal closure (Zarco-Tejada, González-Dugo & Berni 2012), which inhibits photosynthesis and transpiration, leading to an increase in vine leaf temperature (Kim et al. 2011; Maimaitiyiming et al. 2017). Reduced water availability impacts on vine health and productivity, and ultimately on grape quality (Bota et al. 2016). Additionally, under increased climate change scenarios, greater drought periods may be experienced in the near future (Chirouze et al. 2014), with this strain on water resources further inhibiting the development of grapes (Bota et al. 2016). There is consequently an imminent need for real-time monitoring of water stress in vineyards.

---

[1] This chapter was published in *Remote Sensing* and consequently conforms to the prescribed structure of that journal.

Remote sensing provides a fast and cost-effective method for detecting vineyard water stress (Maimaitiyiming et al. 2017), and can thereby help alleviate devastating losses in crop production (Zarco-Tejada et al. 2013) and safeguard high-quality grape yield (González-Fernández et al. 2015). Several studies, for example Baluja et al. (2012) and Zarco-Tejada et al. (2013), have modelled water stress in vineyards using spectral remote sensing techniques. Plant leaves reflect the majority of the NIR spectrum, with the majority of the VIS spectrum (400-680 nm), being absorbed by plant chlorophyll pigments (Kim et al. 2011). Water stress changes the spectral signatures of plants due to decreased photosynthetic absorbance (Kim et al. 2011), resulting in decreased NIR reflectance (Shimada et al. 2012). This phenomenon is known as the "blue-shift", where the red-edge (680-730 nm) shifts toward the VIS end of the spectrum (Govender et al. 2009). Therefore, the red-edge position has subsequently been used to detect water stress in plants (Shimada et al. 2012).

The high spectral resolution of hyperspectral (spectroscopy) data allows for more detailed analysis of plant properties (Govender et al. 2009), and provides a non-destructive approach for assessing vineyard water stress (De Bei et al. 2011). Consequently, application of hyperspectral remote sensing techniques to model vineyard water stress is becoming common practice in precision viticulture (González-Fernández et al. 2015). For example, De Bei et al. (2011) used near infrared (NIR) field spectroscopy to predict the water status of vines using leaf spectral signatures and in-field leaf water potential measurements. Similar studies were conducted by Beghi, Giovenzana & Guidetti (2017) and Diago et al. (2017). All three studies found that wavebands ranging between 1000-2500 nm were ideal for detecting the water stress of vines. Alternatively, studies conducted by Zarco-Tejada et al. (2013) and Pôças et al. (2015) successfully demonstrated the viability of the VIS and red-edge, i.e. 400-730 nm, regions of the electromagnetic (EM) spectrum to predict water stress in vines.

Moreover, the advancement of remote sensing technology in recent years has prompted increased availability of hyperspectral imaging (imaging spectroscopy) sensors. Hyperspectral imaging integrates spectroscopy with the advantages of digital imagery (Carreiro Soares et al. 2016). Each image provides contiguous, narrow-band (typically 10 nm) data, collected across the ultraviolet (UV), VIS, NIR, and shortwave infrared (SWIR) spectrum (typically 350-2500 nm), coupled with high spatial resolutions (typically 1 mm-2 m) (Carreiro Soares et al. 2016; Mulla 2013). A major limitation to the application of hyperspectral data is the inherent "curse of dimensionality" (Poona et al. 2016), which gives rise to the Hughes effect (Hughes 1968) in a classification framework (Pedergnana, Marpu & Mura 2013). High dimensionality can result in reduced classification accuracies (Tong, Xue & Zhang 2014), as the number of wavebands (p) are often many times more than the number of training samples (n), i.e. $p > n$ (Poona & Ismail 2014). However, using variable importance (VI) to create an optimised feature space, i.e. create an optimal subset of input features, has shown to be

effective in reducing the effects of high dimensionality (Abdel-Rahman et al. 2014). For example, Pedergnana, Member & Marpu (2013) exploited the RF mean decrease Gini (MDG) measure of VI to reduce the dimensionality of hyperspectral imagery. The study found that the subset selected based on RF VI produced an increase in accuracy of approximately 1.0%. Alternatively, Abdel-Rahman et al. (2014) utilised the RF mean decrease accuracy (MDA) measure to rank the waveband importance of an AISA Eagle hyperspectral image dataset. The subset produced using MDA VI resulted in a 3.5% increase in accuracy. Contrary to Abdel-Rahman et al. (2014), Corcoran, Knight & Gallant (2013) also utilised RF MDA values to create an optimal subset of features but observed a 4.0% decrease in accuracy. However, in both studies, it was concluded that RF VI could effectively be utilised to increase classification efficiency. Machine learning algorithms, such as Random Forest (RF) (Breiman 2001), have proven to be particularly adept at mitigating the Hughes effect (for example, see Abdel-Rahman et al. 2015; Adam et al. 2017; Poona & Ismail 2014). RF is an ensemble of weak decision trees used for classification and regression (Poona & Ismail 2014). It uses bagging (i.e. bootstrap aggregation) and random variable selection to grow a multitude of unpruned trees from randomly selected training samples (Breiman 2001). RF classification has recently gained significant recognition for its applications in precision viticulture. For example, Sandika et al. (2016) used RF and digital terrestrial imagery to classify Anthracnose, Powdery Mildew, and Downy Mildew diseases within vine leaves. The study found that RF produced the highest accuracy with 82.9%, outperforming Probabilistic Neural Network (PNN), Back Propagation Neural Network (BPNN), and Support Vector Machine (SVM) models. Similar results were found by Knauer et al. (2017) using RF and terrestrial hyperspectral imaging. RF produced an overall accuracy of 87.0% for modelling Powdery Mildew on grapes. Additionally, Knauer et al. (2017) found that dimensionality reduction led to an increase in classification accuracy.

More recently, another tree-based classifier, known as Extreme Gradient Boosting (XGBoost) (Chen & Guestrin 2016), has shown considerable promise in various applications (for example, see Fitriah et al. 2017; Möller et al. 2016; Torlay et al. 2017). XGBoost is an optimised implementation of gradient boosting (Friedman 2001), designed to be fast, scalable, and highly efficient (Ren et al. 2017). Gradient boosting (or boosted trees) combines multiple pruned trees of low accuracies, or weak learners, to create a more accurate model (Friedman 2002). The difference between RF and XGBoost is the way the tree ensemble is constructed. RF grows trees that are independent of one another (Breiman 2001), whereas XGBoost grows trees that are dependent on the feedback information provided by the previously grown tree (Chen & Guestrin 2016). Essentially, each tree in an XGBoost ensemble learns from previous trees and tries to reduce the error produced in subsequent iterations.

Mohite et al. (2017) is the only known study to have employed XGBoost classification in precision viticulture. The study used hyperspectral data to detect pesticide residue on grapes. Four classifiers were compared, i.e. XGBoost, RF, SVM, and Artificial Neural Network (ANN). Additionally, the study investigated the utility of LASSO and Elastic Net feature selection. Results indicated that RF produced the most accurate classification models when using both the LASSO and Elastic Net selected wavebands.

A review of the literature indicated that no study to date has investigated the use of terrestrial hyperspectral imaging to model vineyard water stress. Furthermore, no study has utilised RF or XGBoost classification to detect leaf level water stress in the precision viticulture domain. The aim of the present work was to develop a remote sensing–machine learning framework to model water stress in a Shiraz vineyard. The specific objectives of the study are to evaluate the utility of terrestrial hyperspectral imaging to discriminate between stressed and non-stressed Shiraz vines, and to investigate the efficacy of the RF and XGBoost algorithms for modelling vineyard water stress.

3.3 MATERIALS AND METHODS

3.3.1 Study site

The study was conducted at the Welgevallen experimental farm in Stellenbosch (33°56'38.5"S, 18°52'06.8"E), situated in the Western Cape Province of South Africa (Figure 3.1). Stellenbosch has a Mediterranean climate characterised by dry summers and mild winters, with a mean annual temperature of 16.4 °C (Conradie et al. 2002). Stellenbosch receives low to moderate rainfall, mainly during the winter months (June, July, and August), with an annual average of 802 mm (Conradie et al. 2002), making water scarcity a real threat to irrigated vineyards. Soil deposits in the region comprise rich potassium minerals that are favourable for vineyard growth (Conradie et al. 2002). The Welgevallen experimental farm comprises well-established grape cultivars, including Shiraz and Pinotage; Pinotage being a red cultivar unique to South Africa. Welgevallen is used by Stellenbosch University for research and training, and additionally produces high-quality grapes for commercial use.

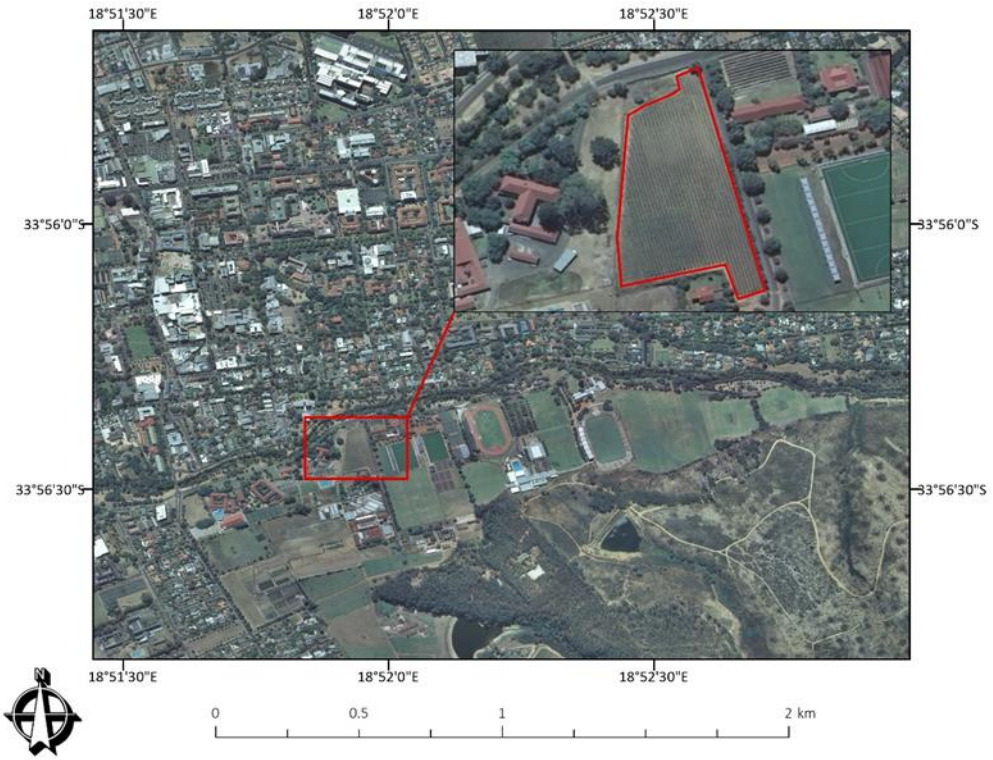

Figure 3.1 Location of the Welgevallen Shiraz vineyard plot used in this study (indicated by red polygon). Background image provided by National Geo-Spatial Information (NGI) (2012).

3.3.2 Data acquisition and pre-processing

To confirm the water stress status of vines, in-field stem water potential (SWP) measurements were captured using a customised pressure chamber (Figure 3.2) as used by Choné et al. (2001) and Deloire & Heyms (2011). Based on the experiments by Deloire & Heyms (2011) and Myburgh, Cornelissen & Southey (2016), vines with SWP values ranging from – 1.0 MPa to – 1.8 MPa were classified as water-stressed, whereas vines with SWP values ≥ – 0.7 MPa were classified as non-stressed. Imaging spectrometer data was subsequently acquired for a water-stressed and non-stressed Shiraz vine. Images were captured between 10:00 and 12:00, on 24 February 2017, to ensure that the side of the vine canopy being captured was fully sunlit.

Figure 3.2 Customised pressure chamber used to measure Stem Water Potential.

Images were captured using the SIMERA HX MkII hyperspectral sensor (SIMERA Technology Group, South Africa). The sensor is a line scanner that captures 340 spectral wavebands across the VIS and NIR (450-1000 nm) with a sensor bandwidth ranging from 0.9 nm to 5 nm. The sensor was mounted on a tripod (Figure 3.3 (A)) to facilitate the collection of terrestrial imagery from a side-on view of the vine canopy. The sensor-tripod assembly was placed at a constant distance of one metre from the vine canopy to ensure that the full canopy of a single vine (approximately 1.4 m W x 1 m H) was captured per image (Figure 3.3 (B)).

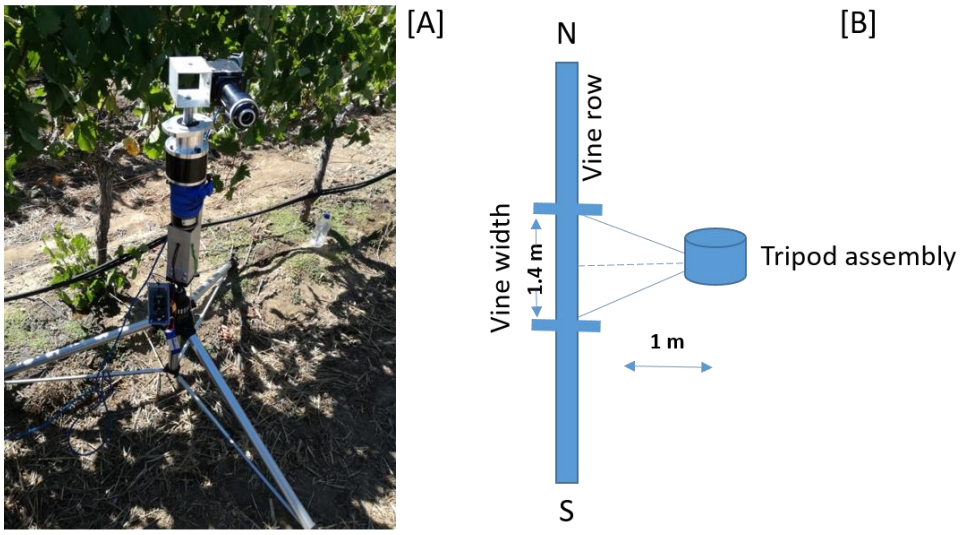

Figure 3.3 The hyperspectral sensor tripod assembly (A), and in-field setup used when collecting terrestrial imagery of the vine canopy (B).

Due to sensor sensitivity and a deteriorating silicon chip, not all the wavebands could be utilised. Spectral subsets were therefore created per image. The spectral subsets consisted of 176 wavebands with a spectral range of 473-708 nm. Thereafter, raw image digital numbers (DNs) were converted to reflectance using the empirical line correction algorithm (Aasen et al. 2015). Empirical line correction uses known field (or reference) reflectance spectra and linear regression to equate DN values to surface reflectance by estimating correction coefficients for each waveband (Aasen et al. 2015). Following Aasen et al. (2015), a white reference panel, positioned in the vine canopy prior to image capture, was used for image correction. Image pre-processing was performed in the Environment for Visualising Images (ENVI) version 5.3.1 software (Exelis Visual Information Solutions 2015). Using a 2x2 pixel region of interest (ROI), a total of 60 leaf spectra were extracted from each image—30 samples per class (stressed and non-stressed)—and used as input for classification.

3.3.3 Spectral smoothing

In-field spectral measurements are often subjected to noise, due to variable sun illumination (Schmidt & Skidmore 2004). Therefore, it is recommended that spectral smoothing be performed in order to

produce a spectral signal that represents the original spectra without the interference of noise (Člupek, Matějka & Volka 2007). The Savitzky-Golay filter (Savitzky & Golay 1964) is a common smoothing technique used in hyperspectral remote sensing (Liu et al. 2016; Prasad et al. 2015; Schmidt & Skidmore 2004). Savitzky-Golay is based on least-squares approximation, which determines smoothing coefficients by applying a polynomial equation of a given degree and cluster size (Savitzky & Golay 1964). The filter is ideal for spectroscopic data, as it minimises signal noise whilst preserving the originality and shape of the input spectra. A second order polynomial filter with a filter size of 15 was applied to the spectral samples prior to classification, following the recommendations of Prasad et al. (2015). The Savitzky-Golay filter was applied using the "signal" package (Ligges, Short & Kienzle 2015) in the R statistical software environment (R Development Core Team 2017). Classification models were produced for both the unsmoothed and smoothed datasets.

3.3.4 Classification

3.3.4.1 Random forest (RF)

The RF ensemble uses a bootstrap sample, i.e. 2/3 of the original dataset (referred to as the "in-bag" sample), to train decision trees. The remaining 1/3 of the data is used to compute an internal measure of accuracy (referred to as the "out-of-bag" or OOB error) (Breiman 2001). To produce the forest of decision trees, two parameters need to be set: the number of unpruned trees to grow, known as *ntree*, and the number of predictor variables (i.e. wavebands) selected, known as *mtry* (Breiman 2001). *Mtry* variables are tested at each node to specify the best split when growing trees. These randomly selected variables produce low correlated trees that prevent over-fitting. In a classification framework, the final classification results are determined by majority vote. For a detailed account of RF, see Belgiu & Drăguţ (2016) and Breiman (2001). RF was implemented using the "randomForest" package (Liaw & Wiener 2002) in the R statistical software environment (R Development Core Team 2017). The default values for *ntree* (*ntree* = 500), and *mtry* (*mtry* = $\sqrt{p}$), were used, following Belgiu & Drăguţ (2016) and Poona, van Niekerk & Ismail (2016).

3.3.4.2 Extreme gradient boosting (XGBoost)

XGBoost, like gradient boosting, is based on three essential elements: (i) a loss function that needs to be optimised, (ii) a multitude of weak decision trees that are used for classification, and (iii) an additive model that combines weak decision trees to produce a more accurate classification model (Möller et al. 2016). XGBoost simultaneously optimises the loss function while constructing the additive model (Chen & Guestrin 2016; Möller et al. 2016). The loss function accounts for the errors

in classification that were introduced by the weak decision trees (Möller et al. 2016). For a detailed account of XGBoost, see Chen & Guestrin (2016). XGBoost was implemented using the "xgboost" package (Chen et al. 2017) in the R statistical software environment (R Development Core Team 2017). XGBoost requires the optimisation of several key parameters (Table 3.1). However, to facilitate fair comparison of RF and XGBoost, the default values for all parameters were used to construct the XGBoost models, with *nrounds* set to 500. Furthermore, to ensure a more robust model and prevent overfitting, a 10-fold cross validation was performed for both RF and XGBoost.

Table 3.1 Key parameters used for XGBoost classification (Chen & Guestrin 2016; Georganos et al. 2018a; Xia et al 2017).

Parameter	Description	Default value
max_depth	Controls the maximum depth of each tree (used to control over-fitting)	6
subsample	Specifies the fraction of observations to be randomly sampled at each tree (adds randomness)	1
eta	The learning rate	0.3
nrounds	The number of trees to be produced (similar to *ntree*)	100-1000
gamma	Controls the minimum loss reduction required to make a node split (used to control over-fitting)	0
min_child_weight	Specifies the minimum sum of instance weight of all the observations required in a child (used to control over-fitting)	1
colsample_bytree	Specifies the number of features to consider when searching for the best node split (adds randomness)	1

3.3.5 Dimensionality reduction

Both RF and XGBoost provide an internal measure of VI. RF provides two measures of VI, namely mean decrease Gini (MDG) and mean decrease accuracy (MDA) (Breiman 2001). MDG quantifies VI by measuring the sum of all decreases in the Gini index, produced by a particular variable. MDA measures the changes in OOB error, which results from comparing the OOB error of the original dataset to that of a dataset created through random permutations of variable values. In this study, MDA was utilised to compute VI following the recommendations of Belgiu et al. (2014); Immitzer, Atzberger & Koukal (2012); and Poona & Ismail (2014). The MDA VI for a waveband X_j is defined by Genuer, Poggi & Tuleau-Malot (2010):

$$VI(X_j) = \frac{1}{ntree}\sum_t (errOOB_{tj} - errOOB_t) \qquad \text{Equation 3.1}$$

where $errOOB_t$ is the misclassification rate of tree t on the OOB_t bootstrap sample not used to construct tree t, and $errOOB_{tj}$ is the error of predictor *t* on the permuted OOB_{tj} sample.

XGBoost ranks VI based on Gain (Chen & Guestrin 2016). Gain measures the degree of improved accuracy brought on by the addition of a given waveband. VI is calculated for each waveband, used for node splitting at a given tree, and then averaged across all trees to produce the final VI per

waveband (Chen & Guestrin 2016). Similar to Abdel-Rahman et al. (2014) and Corcoran, Knight & Gallant (2013), the top 10% (p = 18) of the ranked waveband importance as determined by RF and XGBoost was used to create a subset of important wavebands. RF and XGBoost models were produced for both the original dataset and the subset of 18 wavebands.

3.3.6 Accuracy assessment

To provide an independent estimate of model accuracy, an independent test set was used to evaluate all RF and XGBoost models. Therefore, a second dataset of spectral samples (n = 60) was collected for both stressed (n = 30) and non-stressed (n = 30) vines. Both algorithms were trained using the first dataset of 60 samples and tested using the second dataset. Overall classification accuracies were computed using a confusion matrix (Kohavi & Provost 1998). Additionally, Kappa analysis was used to evaluate model performance. The KHAT statistic (Congalton & Green 2009) provides a measure of the difference between the actual and the chance agreement in the confusion matrix:

$$\widehat{K} = \frac{p_a - p_c}{1 - p_c} \qquad \text{Equation 3.2}$$

where p_a describes the actual agreement and p_c describes the chance agreement. Following Foody (2004); Pedergnana, Marpu & Mura (2013); and Abdel-Rahman et al. (2014), the McNemar's test was employed to determine whether the differences in accuracies yielded by RF and XGBoost were statistically significant. Abdel-Rahman et al. (2014) stated that the McNemar's test could be expressed as the following chi-squared formula:

$$v^2 = \frac{(f_{xgb} - f_{rf})^2}{f_{xgb} + f_{rf}} \qquad \text{Equation 3.3}$$

where f_{xgb} denotes the number of samples misclassified by RF but correctly classified by XGBoost, and f_{rf} denotes the number of samples misclassified by XGBoost but correctly classified by RF. A v^2 value of greater than 3.84, at a 0.05 level of significance, indicates that the results of the two classifiers are significantly different (Abdel-Rahman et al. 2014; Foody 2004).

3.4 RESULTS

3.4.1 Spectral smoothing using the Savitzky-Golay filter

Figure 3.4 shows the results of smoothing the spectral data using the Savitzky-Golay filter. It is evident that the Savitzky-Golay filter produced smoothed spectra without changing the shape of the original spectra. Additionally, the filter successfully preserved the original reflectance values, with the mean difference in reflectance values being less than 0.3% with a standard deviation of 0.003

across all wavebands. All spectra (n = 120) were subsequently smoothed, and the smoothed spectra used as input to classification.

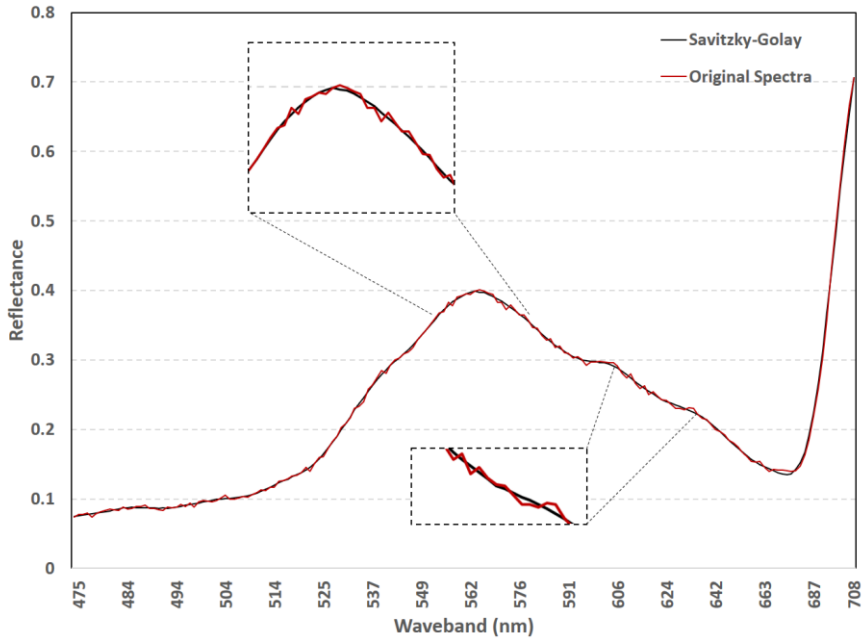

Figure 3.4 Spectra comparison before (red) and after (black) applying the Savitzky-Golay filter.

3.4.2 Important waveband selection

The top 10% (p = 18) of importance wavebands as determined by RF MDA and XGBoost gain are shown in Figure 3.5 (A) and Figure 3.5 (B) respectively. The results in Table 3.2 show that RF selected wavebands across the blue and green (473.92-585.12 nm) regions of the EM spectrum. In comparison, XGBoost selected wavebands across the VIS (473.92-646.04 nm) and red-edge (686.69-708.32 nm) regions. It is evident from Figure 3.5 that the location of the wavebands selected by RF and XGBoost are significantly different. The study attributed the difference in waveband location to the difference in VI measures used for RF and XGBoost. Nevertheless, as illustrated in Figure 3.5 (C), there were common wavebands selected by both RF and XGBoost. The overlapping wavebands (p = 6) were located across blue and green (473.92-585.12 nm) regions. Consequently, those wavebands may be the most important for discriminating between stressed and non-stressed Shiraz vines.

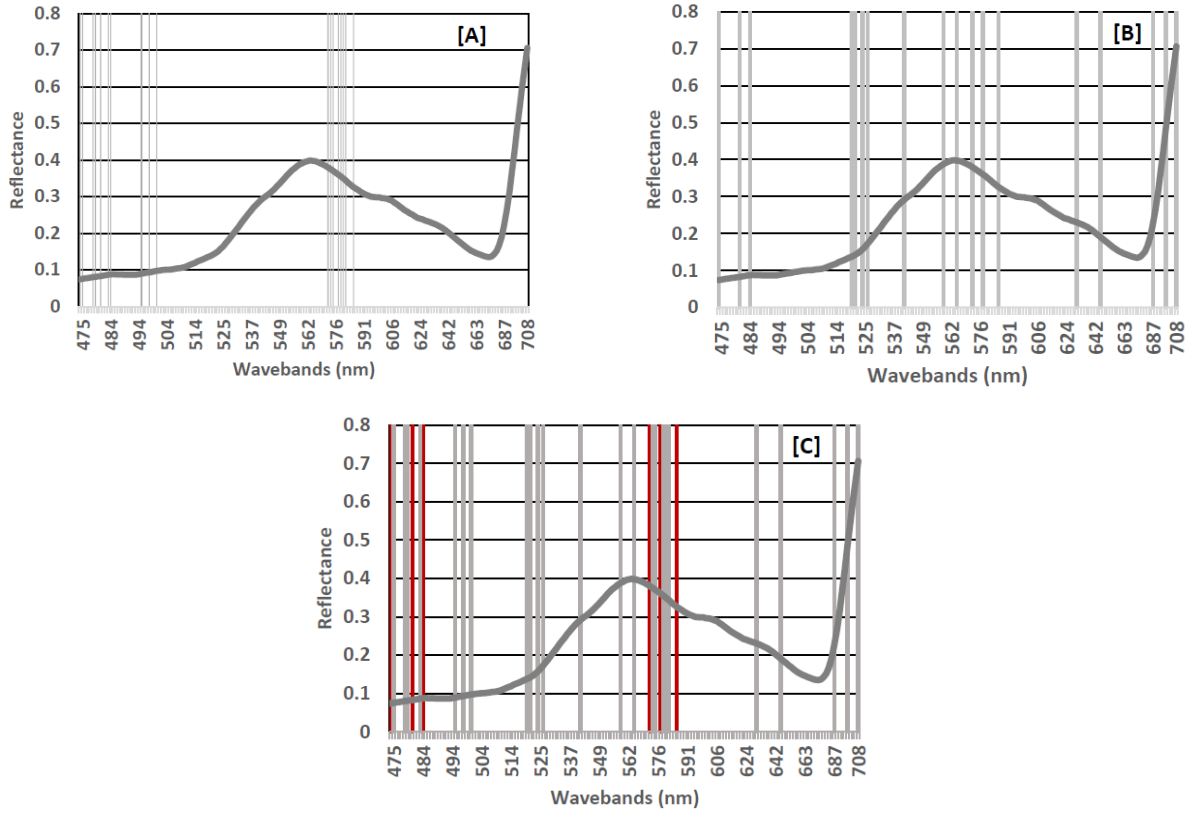

Figure 3.5 The importance wavebands as determined by RF (A); XGBoost (B); and overlapping (C). The grey bars represent the important wavebands selected by RF and XGBoost, respectively. The red bars indicate the overlapping wavebands. The mean spectral signature of a sample is shown as a reference.

Table 3.2 Location of the RF and XGBoost selected important wavebands in the EM spectrum.

	p	**VIS** (473-680 nm)	***p***	**Red-edge** (680-708 nm)
RF	12	474.74, 478.09, 478.94, 483.2, 494.64, 497.36, 500.11, 573.31, 574.59, 578.48, 579.79, 581.11	0	-
XGBoost	9	520.31, 521.32, 524.36, 526.42, 541.34, 558.52, 564.56,630.23, 646.04	3	686.69, 698.39, 708.32
Overlap	6	473.92, 480.63, 484.06 572.04, 577.17, 585.12	0	-

3.4.3 Classification using random forest and extreme gradient boosting

The classification results for RF and XGBoost are shown in Table 3.3. Training accuracies for all models were above 80.0%, with test accuracies ranging from 77.6% to 83.3% (with KHAT values ranging from 0.60 to 0.87). Overall, the results indicate that RF outperformed XGBoost, producing the highest accuracies for all the classification models.

Table 3.3 Classification accuracies of both the RF and XGBoost models constructed using all the wavebands and the subset of important wavebands.

		All Wavebands (p= 176)				Important Wavebands (p = 18)			
		Train		**Test**		**Train**		**Test**	
		Accuracy (%)	Kappa	Accuracy (%)	Kappa	Accuracy (%)	Kappa	Accuracy (%)	Kappa
XGBoost	Unsmoothed	85.0	0.70	78.3	0.57	90.0	0.80	80.0	0.60
	Smoothed	83.3	0.67	77.6	0.53	86.7	0.73	78.3	0.57
RF	Unsmoothed	90.0	0.80	83.3	0.67	93.3	0.87	83.3	0.67
	Smoothed	90.0	0.80	81.7	0.63	91.7	0.83	81.7	0.63

Using the unsmoothed wavebands (p = 176), RF yielded a training accuracy of 90.0% (KHAT = 0.80) and a test accuracy of 83.3% (KHAT = 0.67). In comparison, XGBoost produced significantly lower accuracies with a training accuracy of 85.0% (KHAT = 0.7) and a test accuracy of 78.3% (KHAT = 0.57). These results indicate that the XGBoost ensemble resulted in reduced accuracies (approximately –5.0%) when using all wavebands to classify stressed and non-stressed Shiraz leaves.

Using the subset of important wavebands (p = 18) resulted in an overall improvement in classification accuracies for the RF and XGBoost. Training accuracy for RF increased by 3.3% to 93.3% (KHAT = 0.87). However, the test accuracy remained unchanged. Although XGBoost produced less accurate results, it did experience a greater increase in accuracy (5.0%), producing a training accuracy of 90.0% and a KHAT value of 0.8. The greater increase in accuracy may be attributed to the red-edge wavebands that were only present in the XGBoost subset. Moreover, the XGBoost subset also produced a slight increase (1.7%) in test accuracy (80.0%, KHAT = 0.6). The study attributed the superior performance of the RF algorithm to its use of bootstrap sampling (Breiman 2001), which introduces model stability, and its robustness to noise (Belgiu & Drăguţ 2016).

Classification using the Savitzky-Golay smoothed spectra resulted in reduced accuracies overall. The decrease in accuracy ranged from 0.7% to 3.3% for all models. Furthermore, according to the McNemar's test results, the difference in classifier performance was not statistically significant. For all the classification models, the chi-squared values were less than 3.84 with v^2 values ranging from 0.14 to 1.29.

3.5 DISCUSSION

Ensemble classifiers, like RF and XGBoost, have been widely used to address the classification challenges inherent in high dimensional data (Poona, van Niekerk & Ismail 2016). The present study evaluated the use of terrestrial hyperspectral imaging to model vineyard water stress. More specifically, the study tested the utility of two tree-based ensemble classifiers, namely RF and

XGBoost, to model water stress in a Shiraz vineyard. The experimental results are discussed in further detail in the following sections.

3.5.1 Efficacy of the Savitzky-Golay filter

The Savitzky-Golay filter has become a popular algorithm for smoothing spectroscopic data (Gutiérrez et al. 2016; Prasad et al. 2015; Schmidt & Skidmore 2004). In this study, the filter proved adept at smoothing the hyperspectral signature without significantly altering the originality of the input data. However, the results of this study showed that the filter negatively impacted on classification accuracy, producing reduced accuracies for RF (–1.6%) and XGBoost (–3.3%). The decrease in classification accuracy may be attributed to the specific parameter values used to implement the filter. The study only meant to test the functionality of the Savitzky-Golay filter. Therefore, the filter was implemented using the hyperparameter values as recommended by Prasad et al. (2015). Consequently, the recommended values may not be optimal for the specific dataset used in this study.

Carvalho et al. (2006) utilised the Savitzky-Golay filter to smooth magnetic flux leakage (MFL) signals. Similar to the present study, the authors found that using the smoothed data with an ANN classifier resulted in reduced classification accuracies. It is therefore evident that careful consideration has to be taken when applying the Savitzky-Golay filter.

3.5.2 Classification using all wavebands

Both tree-based ensemble classifiers tested in the study successfully demonstrated their efficiency for analysing hyperspectral data. However, the analysis found the RF bagging ensemble to outperform the boosting-based XGBoost ensemble when using all wavebands (p = 176).

Published comparisons between RF and boosting classifiers, similar to XGBoost, have reported mixed results. For example, Miao et al. (2012) found that RF (93.5%) and AdaBoost (95.3%) produced similar overall accuracies when classifying ecological zones using multi-temporal and multi-sensor data. Contrary to Miao et al. (2012), Xu, Li & Brenning (2014) reported that RF outperformed boosting ensemble classifiers when classifying RADARSAT-1 imagery. Moreover, when directly comparing RF and XGBoost within the context of spectroscopic classification, the findings of the current study contradict the results reported by Möller et al. (2016). Their study reported that XGBoost (96.0 %) yielded significantly better results than RF (87.0 %) when classifying supernovae. However, it should be noted that their study optimised RF and XGBoost parameters. More specifically, within viticulture, the results of the present study compare favourably to those reported by Mohite et al. (2017). The authors found that RF (87.8%) produced improved accuracy

compared to XGBoost (81.6%) when using hyperspectral data in combination with feature selection. A review by Belgiu & Drăguţ (2016) concluded that RF generally achieves greater accuracies compared with boosting methods when used for the classification of high dimensional data, such as hyperspectral imagery.

When comparing the utility of both algorithms, a key advantage shared between them is that RF and XGBoost effectively prevent overfitting (Breiman 2001; Chen & Guestrin 2016). However, given that RF grows trees independently (i.e. parallel to one another), whereas XGBoost grows trees sequentially, it is less complex and therefore less computationally intensive. Furthermore, RF requires the optimisation of only two parameters (Breiman 2001), whereas XGBoost has various parameters that could be optimised for a given dataset (Chen & Guestrin 2016).

3.5.3 Classification using subset of important wavebands

Dimensionality reduction of hyperspectral data using machine learning has been extensively researched (for example, see Abdel-Rahman et al. 2014; Pedergnana, Marpu & Mura 2013; Poona & Ismail 2014). The results of the study indicate the VI ranking provided by RF and XGBoost can successfully be used to select a subset of wavebands for classification. This was evident from the increased accuracies obtained for both RF and XGBoost.

The results of the current study compare favourably to those reported by Pedergnana, Marpu & Mura (2013) and Abdel-Rahman et al. (2014), who demonstrated the feasibility of VI to reduce the high dimensionality of hyperspectral data and improve classification accuracy. The study therefore attributed the improved classification performance to the subset of most important wavebands. Although the subset of important wavebands did not result in massive accuracy gains (accuracy increases of RF ranged from 1.7% to 3.3% and from 0.7% to 3.3% for XGBoost), it did improve classification accuracy using only 10% of the original data. The majority of important wavebands, for RF (p = 9) and XGBoost (p = 10), were located in the green region of the EM spectrum (Table 3.2). The selected wavebands correspond to similar wavebands reported by Zarco-Tejada et al. (2013) and Pôças et al. (2015). The green region (i.e. between 500-600 nm) is highly sensitive to plant chlorophyll absorption (Pôças et al. 2015). Consequently, water stress in plants is closely related to lowered chlorophyll leaf concentrations (Pôças et al. 2015), which presents a possible explanation for the selection of these wavebands.

Moreover, Shimada et al. (2012) reported the use of the blue (490 nm) and red wavebands (620 nm) as indicators of plant water stress. These wavebands correspond to similar wavebands present in the XGBoost subset (484.06 nm and 630.23 nm). In this study, only three red-edge wavebands (Table 3.2) were selected by XGBoost with none selected by RF. These results contradict those reported by

Maimaitiyiming et al. (2017), which found that wavebands in the red-edge region (695-730 nm) were ideal for early water stress detection in vineyards. However, given the overlapping wavebands that occur in the blue and green regions and the results found in the present study, it can be concluded that the red-edge wavebands may not be important for discriminating between stressed and non-stressed Shiraz vines. The results of this study subsequently demonstrate the feasibility of VIS wavebands to model water stress in a Shiraz vineyard.

Various aspects of the current research lend themselves to be operationalised within precision viticulture. For instance, the developed remote sensing–machine learning framework can be readily applied to model vegetative water stress. Furthermore, the identification of important wavebands can potentially lead to the construction of custom multispectral sensors that are less expensive and application-specific.

3.6 CONCLUSION

This study presents a novel remote sensing–machine learning framework for modelling water stress in a Shiraz vineyard using terrestrial hyperspectral imaging. Based on the results of the study, the following conclusions can be drawn:

1. Both RF and XGBoost may be utilised to model water stress in a Shiraz vineyard.
2. Wavebands in the VIS region of the EM spectrum may be used to model water stress in a Shiraz vineyard.
3. It is imperative that future studies carefully consider the impact of applying the Savitzky-Golay filter for smoothing spectral data.
4. The developed framework requires further investigation to evaluate its robustness and operational capabilities.

Given the results obtained in the present study, the employment of RF is recommended (rather than XGBoost) for the classification of hyperspectral data to discriminate stressed from non-stressed Shiraz vines.

CHAPTER 4: A MACHINE LEARNING FRAMEWORK FOR TERRESTRIAL HYPERSPECTRAL IMAGE CLASSIFICATION [1]

4.1 ABSTRACT

The increasing availability of hyperspectral remote sensing data necessitates the development of accurate classification frameworks—either fully or semi-automated—that are computationally inexpensive, flexible, and easily scalable. To this end, the present study employed the Random Forest (RF) and Extreme Gradient Boosting (XGBoost) classifiers, coupled with three feature selection algorithms, to discriminate between water-stressed and non-stressed Shiraz vines using terrestrial hyperspectral imagery. Additionally, the study examined the optimisation of classifier hyperparameter values with respect to classification accuracy and computational expense. The results showed that RF marginally outperformed XGBoost when using all wavebands (p = 176) and optimised hyperparameter values. RF yielded a test accuracy of 83.3% (KHAT = 0.67), whereas XGBoost yielded a test accuracy of 81.7% (KHAT = 0.63). The results further show that optimising hyperparameter values lead to an overall increase in test accuracy, ranging from 0.8% to 5.0%, for both the RF and XGBoost classifiers. Using the Sequential Floating Forward Selection (SFFS) and Filter-Wrapper (FW) derived subsets yielded a test accuracy of 80.0% (KHAT = 0.6) for both RF and XGBoost. However, the SFFS wrapper required longer processing times. Overall, the results highlight the effect of feature selection and optimisation on the performance of machine learning ensembles for modelling vineyard water stress.

4.2 INTRODUCTION

Hyperspectral remote sensing, also known as imaging spectroscopy, provides a wealth of spectral information (Li et al. 2018) by recording narrow-band reflectance across the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) regions of the electromagnetic (EM) spectrum. These narrow, contiguous wavebands convey meaningful spectral variations (Santara et al. 2017; Tong, Xue & Zhang 2014) that can aid in the discrimination of spectrally similar features (Li et al. 2018). Hyperspectral remote sensing combines the high spectral resolution of spectroscopy with the spatial characteristics of 2-dimensional imagery (Mulla 2013). This combination makes imaging

[1] This chapter has been submitted for peer review in the *International Journal of Remote Sensing* and consequently conforms to the prescribed structure of that journal.

spectroscopy ideal for applications in agriculture, as it enables the detection of vegetative stress (Abdel-Rahman et al. 2014) and the quantification of crop spatial variability (Mulla 2013).

However, the complexity and vast volume of information present in high dimensional datasets, such as hyperspectral imagery, have rendered traditional classifiers impractical (Pappu & Pardalos 2014). Traditional classifiers tend to overfit the training model, resulting in decreased classification accuracies (Pappu & Pardalos 2014). Machine learning techniques, such as ensemble learners and feature selection algorithms, provide for quantitative analysis that can enhance the exploitation of hyperspectral data (Tong, Xue & Zhang 2014).

Tree-based ensemble learners have been widely used in classification-driven applications. The Random Forest (RF) tree-based ensemble (Breiman 2001) is one of the most popular algorithms utilised for the classification of hyperspectral data. RF has been successfully employed to discriminate between healthy and stressed vegetation (Adam et al. 2017; Poona et al. 2016), classify ecological zones and ecotopes (Chan & Paelinckx 2008; Miao et al. 2012), and map land cover and land use (Eisavi et al. 2015; Georganos et al. 2018a). RF is a non-parametric classifier that is robust to outliers and noise (Breiman 2001). However, it is the algorithm's resistance to overfitting (Breiman 2001) that makes RF ideal for the classification of hyperspectral datasets (Abdel-Rahman et al. 2014). RF prevents overfitting by growing uncorrelated trees using randomly selected subsets of the input data (Abdel-Rahman et al. 2014; Breiman 2001). Recent remote sensing studies have demonstrated the utility of RF for precision viticulture. For example, Poblete-Echeverría et al. (2017) reported an average classification accuracy of 94.0% (Kappa = 0.91) when utilising RF to discriminate vine canopies from soil profiles. A high classification accuracy (87.0%) was also reported by Knauer et al. (2017) when employing RF for the detection of Powdery Mildew on grapes.

Extreme Gradient Boosting (XGBoost) (Chen & Guestrin 2016) is an alternative to the RF ensemble that has recently gained noteworthy recognition in the literature (for example, see Fitriah et al. 2017; Loggenberg et al. 2018; Torlay et al. 2017). XGBoost is a regularised tree boosting ensemble learner that builds on the Gradient Boosting Machine (GBM) proposed by Friedman (2001). XGBoost has been shown to outperform traditional GBM in both predictive competency and computational cost (Kejela & Rong 2016; Tavares, Mastelini & Barbon Jr. 2017). In classification frameworks, XGBoost has demonstrated considerable promise in a wide variety of applications, ranging from fraud detection of online social media accounts (Tavares, Mastelini & Barbon Jr. 2017) to classifying patients with epilepsy (Torlay et al. 2017). Within precision viticulture, Mohite et al. (2017) utilised XGBoost to detect pesticide residue on vineyard grapes. The authors reported classification accuracies for XGBoost ranging from 81.6% to 87.6%. In a more recent study, Loggenberg et al. (2018) employed the XGBoost algorithm to model water stress in a Shiraz vineyard. Classification accuracies ranged

from 78.0% to 90.0%, with KHAT values ranging from 0.53 to 0.6. Additionally, Mohite et al. (2017) and Loggenberg et al. (2018) compared the utility of XGBoost and RF and found that RF often outperformed XGBoost.

Moreover, neighbouring hyperspectral wavebands are often highly correlated (Thorp et al. 2017), which presents redundancy in a given dataset (Taşkın, Hüseyin & Bruzzone 2017). This inherent multicollinearity, coupled with the curse of dimensionality (Hughes 1968) caused by the number of wavebands (*p)* being greater than the number of training samples (*n*) (i.e. $p > n$), presents a unique challenge for classification frameworks (Li et al. 2018; Tong, Xue & Zhang 2014). As a result, feature selection techniques have been employed to reduce the dimensionality of hyperspectral datasets (for example, see Fu et al. 2017; Lagrange, Fauvel & Grizonnet 2017; Pedergnana, Marpu & Mura 2013; Vélez Rivera et al. 2014). Feature selection reduces dimensionality by removing redundant and/or irrelevant wavebands (Taşkın, Hüseyin & Bruzzone 2017), thereby lessening computational complexity without significantly decreasing classification competency (Chandrashekar & Sahin 2014).

Notably, varied results have been reported regarding the effects of feature selection on model performance. For example, Vélez Rivera et al. (2014) reported a 6.5% decrease in accuracy when classifying mechanical damage in mango fruits using NIR hyperspectral imagery. Li et al. (2011) and Pedergnana, Marpu & Mura (2013) also reported decreased classification accuracies when employing feature selection techniques on hyperspectral datasets. Contrary to these findings, several studies have reported improved model performance after feature selection techniques were employed (for example, see Belgiu et al. 2014; Poona, van Niekerk & Ismail 2016).

Feature selection techniques are broadly categorised into filter and wrapper approaches (Chandrashekar & Sahin 2014; Fu et al. 2017). The filter approach, which functions independently from the classifier (Fu et al. 2017), is used as a pre-processing step to rank wavebands based on their relative importance (Lagrange, Fauvel & Grizonnet 2017). The highly-ranked wavebands, i.e. wavebands that contain the most useful information about a given class, are then selected based on a predefined threshold value and used as input for classification (Chandrashekar & Sahin 2014). In comparison, the wrapper approach evaluates the suitability of waveband subsets based on the performance of a given classifier (Jović, Brkić & Bogunović 2015). Wrappers employ searching strategies to automatically select subsets, with the optimal subset selected based on predictive competency (Chandrashekar & Sahin 2014). Subsequently, wrapper approaches are computationally more expensive when compared with filter approaches (Chandrashekar & Sahin 2014), but have proven to produce subsets that yield improved classification results (Jović, Brkić & Bogunović 2015). For a detailed review of popular feature selection methods, see Chandrashekar & Sahin (2014).

An analysis of the literature indicates that most studies to date have employed machine learning algorithms, such as RF, using default hyperparameter values (for example, see Lagrange, Fauvel & Grizonnet 2017; Poona, van Niekerk & Ismail 2016; Taşkın, Hüseyin & Bruzzone 2017). However, Xia et al. (2017) assert that machine learning algorithms can be highly sensitive to hyperparameter settings, which can greatly affect the algorithm's performance. Furthermore, the hyperparameter values for a given algorithm are dataset-dependent and are consequently rarely optimal across applications (Rodrìguez, Kuncheva & Alonso 2006). Hyperparameter optimisation thus forms an integral addition to machine learning classification frameworks (Martinez-de-Pison et al. 2017), providing an efficient automated method that can greatly lessen the burden of manual hyperparameter tuning (Xia et al. 2017). Poona & Ismail (2014) highlighted that optimised RF hyperparameter values leads to improved classification accuracies when compared with using default hyperparameter values.

Several studies to date have successfully demonstrated the combined utility of hyperspectral data and machine learning in the field of precision viticulture (for example, see Gutiérrez et al. 2016; Loggenberg et al. 2018; Mohite et al. 2017). A frequently identified problem within precision viticulture is that of vineyard water stress. Water stress is an important growth-limiting factor in vineyard crop production (Maimaitiyiming et al. 2017; Pôças et al. 2015). Vineyard water stress negatively influences the grape quality and could potentially lead to devastating crop losses (Costa et al. 2016). Hyperspectral–machine learning frameworks can potentially serve as a practical means to monitor water stress in vineyards and facilitate the implementation of effective management schemes (Genc et al. 2013; Loggenberg et al. 2018).

It is within this context that the current research evaluates the utility of RF and XGBoost to model water stress in a Shiraz vineyard using terrestrial hyperspectral imagery. This study extends the work of Loggenberg et al. (2018) (i.e. Chapter 3) by exploring the utility of three feature selection approaches, namely filter, wrapper, and a filter applied within a wrapper paradigm. All three approaches are examined with the RF and XGBoost ensemble framework. Additionally, the research examines the effect of hyperparameter value optimisation on classification accuracy and computational expense.

4.3 MATERIALS AND METHODS

4.3.1 Experimental design

The study utilised terrestrial imaging spectrometer data to discriminate water-stressed from non-stressed Shiraz vines. The imagery consisted of 176 wavebands ranging from 473 nm to 708 nm, with a bandwidth range of 0.9 nm to 2 nm. The imagery was captured on February 24, 2017, between

10:00 and 12:00, and pre-processed using the Environment for Visualising Images (ENVI) version 5.3.1 software (Exelis Visual Information Solutions 2015). All processing that followed was completed in the R statistical software environment (R Development Core Team 2017) on a contemporary machine running Windows-64 OS, with an i7-4770 CPU @ 3.40GHz and 8 GB RAM. For a detailed account of the study area, data collection, and pre-processing methods, see Loggenberg et al. (2018) (i.e. Chapter 3).

4.3.2 Statistical analysis

Classification models were developed for the RF and XGBoost learners using both default and optimised hyperparameter values. The two tree-based classifiers evaluated the performance of the Kruskal-Wallis (KW) filter (Kruskal & Wallis 1952), Sequential Floating Forward Selection (SFFS) wrapper (Pudil, Novovičová & Kittler 1994), and Filter-Wrapper (FW) (Bischl et al. 2016) feature selection approaches. The predictive competencies of each feature selection approach were compared with using all wavebands, i.e. p = 176. The classifiers were trained using 60 leaf spectra samples—extracted from the terrestrial imagery—with 30 samples collected for each of the stressed and non-stressed classes.

4.3.2.1 Random forest (RF)

RF grows a multitude of classification trees (*ntree*), by randomly selecting training samples with replacement (Breiman 2001). Each tree is maximally grown, i.e. without pruning, on 2/3 of the original data (i.e. bagged sample), with a random subset of wavebands (*mtry*) used to determine the split at each tree node (Breiman 2001). RF produces a complex forest of trees that have high variance and low bias, and applies an arithmetic mean across all the trees grown to produce the final class probability (Belgiu & Drăguţ 2016; Breiman 2001). A detailed review on RF can be found in Breiman (2001) and Belgiu & Drăguţ (2016). RF was implemented using the "randomForest" package (Liaw & Wiener 2002). The default hyperparameter values used for RF were *ntree* = 500 and *mtry* = $\sqrt{p}$, where p is the number of wavebands.

4.3.2.2 Extreme gradient boosting (XGBoost)

Similar to RF, XGBoost transforms weak classification trees into an ensemble of strong predictive capacity. However, unlike RF, where trees are grown independently (Breiman 2001), XGBoost builds trees that learn from the previously grown tree (Chen & Guestrin 2016). XGBoost trains a model in an additive manner and finds the best parameters for the given model by defining an objective function (Chen & Guestrin 2016). The objective function (Equation 4.1) contains a user-defined

predictive term, which measures the predictive competency of the model, and a regularisation term, which controls overfitting and reduces model complexity (Chen & Guestrin 2016):

$$\text{obj}(\theta) = \varphi(\theta) + \gamma(\theta) \qquad \text{Equation 4.1}$$

where φ is the training loss function, and γ is the regularisation term. The XGBoost classifier was employed using default hyperparameter values (Table 4.1) and implemented using the "xgboost" package (Chen et al. 2017). For a more detailed account on XGBoost and its hyperparameter settings, see Chen & Guestrin (2016); Xia et al. (2017); and Loggenberg et al. (2018).

4.3.3 Hyperparameter optimisation

In an attempt to obtain the best classification results, the hyperparameter values for both RF and XGBoost were optimised for all models using the grid search optimisation method. For RF, *ntree* and *mtry* hyperparameter values were optimised following the recommendation of Abdel-Rahman et al. (2014) and Poona et al. (2016). *Ntree* values up to 2500 were evaluated using intervals of 500, with the *mtry* value set to varying multiplicative factors of $\sqrt{p}$ (i.e 1/4, 1/3, 1/2, 1, 1.5, 2, 2.5, 3). Six hyperparameter (see Table 4.1) values were optimised for XGBoost, based on work done by Xia et al. (2017); Georganos et al. (2018a); and Loggenberg et al. (2018). The optimal RF and XGboost hyperparameter values were selected based on model performance, i.e. hyperparameter values that yielded the lowest out-of-bag (OOB) error and loss of accuracy (loss function), for RF and XGBoost respectively. The Grid Search optimisation method was employed utilising the "caret" package (Kuhn et al. 2017). Grid Search optimisation utilises user-specified ranges to iteratively test all possible combinations of hyperparameter values (Kuhn et al. 2017).

Table 4.1 Optimisation ranges tested for XGBoost hyperparameters.

Hyperparameter	Default value	Range	Interval
max_depth	6	3-10	2
subsample	1	0.5-1	0.2
eta	0.3	0.01-0.2	0.2
nrounds	500 (current study)	20-100 50-250 500-2500	20 50 500
min_child_weight	1	0.2-1	0.2
colsample_bytree	1	0.5-1	0.2

4.3.4 Waveband selection

The KW filter, SFFS wrapper, and FW feature selection approaches were all implemented using the "mlr" package (Bischl et al. 2016). The KW filter generated a single waveband subset that served as input to classification for both RF and XGBoost. The SFFS and FW approaches—using default classifier hyperparameters—generated individual subsets for RF and XGBoost, respectively.

4.3.4.1 Filter

The KW filter was employed to rank wavebands based on their respective importance. The KW filter was utilised based on the recommendation of Vora & Yang (2017). The authors reported the effectiveness of the KW filter when applied to high dimensional datasets in a classification framework. KW is a non-parametric filter that is applied to a labelled dataset with x number of classes (Vora & Yang 2017). Wavebands are partitioned based on their class membership, and a KW statistic subsequently calculated (Vora & Yang 2017). The KW statistic is defined as (Zhao et al. 2010):

$$K = (N-1)\frac{\sum_{r=1}^{x} i_r(\varpi_r-\varpi)^2}{\sum_{r=1}^{x}\sum_{k=1}^{i_r}(\omega_{rk}-\varpi)^2} \qquad \text{Equation 4.2}$$

where N is the total number of wavebands across all the classes, ϖ is the average rank of each waveband, i_r is the number of wavebands in class r, ω_r is the average rank of wavebands in class r, and ω_{rk} is the rank of waveband k in class r. The top 10% (p = 18) of the ranked wavebands were then selected and used as input for classification following Abdel-Rahman et al. (2014) and Loggenberg et al. (2018).

4.3.4.2 Wrapper

Sequential wrapper techniques, such as Sequential Forward Selection (SFS), are frequently utilised for dimensionality reduction (Taşkın, Hüseyin & Bruzzone 2017) as they are computationally less expensive than exhaustive search methods (Fu et al. 2017). A more flexible variant of SFS, known as SFFS, was employed in the present study as it has been shown to outperform SFS (Taşkın, Hüseyin & Bruzzone 2017). SFFS builds on its predecessor by integrating feature (i.e. waveband) removal (Chandrashekar & Sahin 2014), which facilitates the prevention of feature nesting (Pudil, Novovičová & Kittler 1994). Similar to SFS, SFFS adds wavebands one at a time to an empty subset (k). The predictive prowess of the subset is then evaluated based on a user-defined objective function (Chandrashekar & Sahin 2014). The subset is then deemed optimal (d) or not; if deemed not optimal, the SFFS algorithm then randomly removes a waveband and the new subset is re-evaluated (see Figure 4.1). This process is iterated until the algorithm converges, i.e. classification accuracy does

not improve more than a given threshold (*alpha*). *Alpha* specifies algorithm stoppage criteria for the SFFS wrapper.

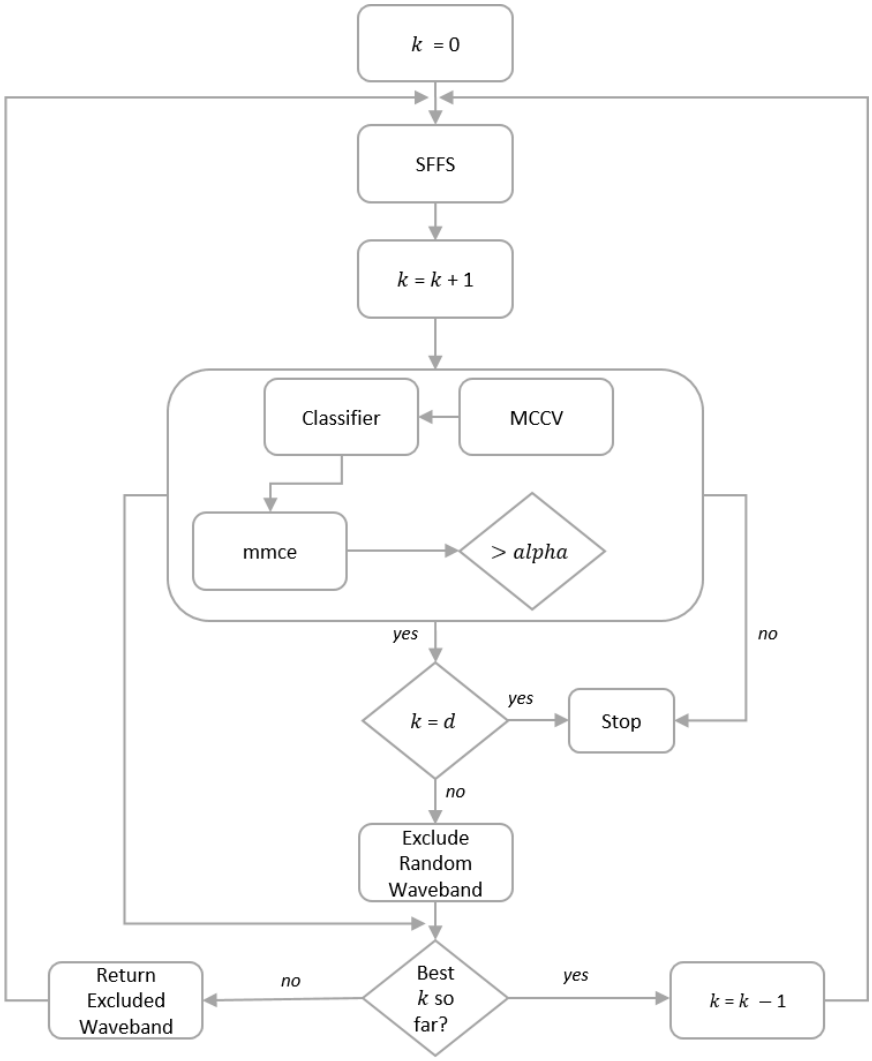

Figure 4.1 SFFS Wrapper workflow (adapted from Chandrashekar & Sahin 2014).

SFFS was implemented with an *alpha* value of 0.02 (Bischl et al. 2016). Monte-Carlo Cross Validation (MCCV) (Xu & Liang 2001) splits the dataset into 2/3 training, with the remaining 1/3 used to test model performance. MCCV was iterated 10-fold and the subset that yielded the lowest Mean Misclassification Error (MMCE) was selected as the optimum subset of wavebands.

4.3.4.3 Filter-Wrapper (FW)

The aim of the FW approach was two-fold: to try and improve the classification accuracy achieved using the KW filter, and to lessen the computational strain often associated with wrapper methods (such as SFFS). Implementation of the FW approach followed a simple workflow (see Figure 4.2). Wavebands are firstly ranked using the KW filter, with a subset then selected based on an optimal threshold value. The classifier then evaluates the predictive competency of the selected subset using a 10-fold MCCV, similar to the SFFS wrapper approach. The entire process is then iterated 10-fold

and the optimum subset selected based on the lowest MMCE. The threshold value for FW was optimised using grid search, with values ranging from 0.02 to 0.2; equivalent to testing subsets of the top 2% to 20% of the ranked wavebands.

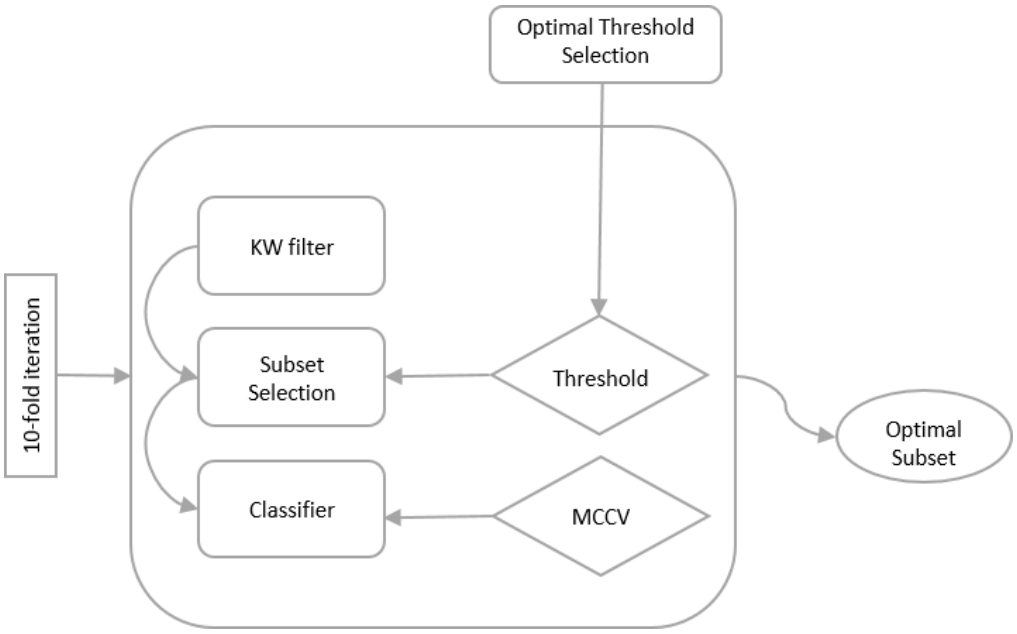

Figure 4.2 Filter-Wrapper workflow.

4.3.5 Accuracy assessment

An independent test set (n = 60), collected for stressed (n = 30) and non-stressed (n = 30) vines, was used to evaluate model performance. The performance of all models was compared based on their measured mean accuracies (Kohavi & Provost 1998), computed using a confusion matrix, and KHAT statistic (Congalton & Green 2009). All classification models were iterated 10-fold to ensure model robustness and to prevent overfitting.

4.4 RESULTS AND DISCUSSION

4.4.1 RF and XGBoost optimisation

Table 4.2 indicates the optimised hyperparameter values for RF and XGBoost. Using all wavebands (p = 176) as input—optimised hyperparameter values for RF were *ntree* = 1 500 and *mtry* = 4—yielded the lowest OOB error of 6.7%. Notably, for all RF models the optimised *mtry* values were smaller than the default values (see Table 4.2 (A)). Although smaller *mtry* values lead to decreased computational expense, as fewer wavebands are considered for node splitting, Goldstein et al. (2010) assert that a smaller *mtry* can lead to biased RF models and decreased accuracies. However, several authors (for example, Abdel-Rahman et al. 2015; Adam et al. 2017) have shown that smaller *mtry* values lead to improved classification performance.

An *ntree* value of 500 was determined as optimal for all three feature selection approaches, i.e. the KW filter, SFFS wrapper, and FW approach. A default *ntree* value of 500 was also found to be optimal in previous studies by Abdel-Rahman et al. (2014); Abdel-Rahman et al. (2015); and Poona et al. (2016). The higher *ntree* value (1 500) obtained using all wavebands may be attributed to the dataset comprising weak predictors, resulting in a model requiring a larger number of trees (Goldstein et al. 2010).

For XGBoost, using all wavebands as input yielded the lowest classification error of 8.3%. All XGBoost models comprised a lesser number of trees (*nrounds*) compared with RF (see Table 4.2 (B)). A similar finding was reported by Georganos et al. (2018a) where RF required more trees (*ntrees = 2000*) than XGBoost (*nrounds* = 600) when optimising models for land cover classification using high dimensional datasets. Xia et al. (2017) noted that an inherent trade-off exists between the number of trees (*nrounds*), tree complexity (*max_depth*), and the learning rate (*eta*). The authors further assert that generally, for a given learning rate, a smaller *max_depth* value leads to a greater number of trees. In this study, learning rates ranging from 0.07 to 0.15 were observed. These relatively small learning rate values indicate that the optimised models would be more robust to overfitting (Xia et al. 2017). However, the slower learning rate (i.e. lower *eta* value) would increase computational expense (Xia et al. 2017).

Table 4.2 Optimised hyperparameter values using grid search.

(A) RF.

Model	**Default**		**Optimised**		
	ntree	*mtry*	*ntree*	*mtry*	**OOB Error**
All wavebands (*p* = 176)	500	13	1500	4	6.70%
KW (*p* = 18)		4	500	2	28.30%
FW (*p* = 35)		6	500	3	11.70%
SFFS (*p* = 4)		2	500	1	8.30%

(B) XGBoost.

Model		*max _depth*	*subsample*	*eta*	*nrounds*	*min_child _weight*	*colsample _bytree*	**Classification Error**
Default		6	1	0.3	500	1	1	-
Optimised	**All wavebands** (*p* = 176)	3	0.7	0.11	250	0.6	0.7	8.30%
	KW (*p* = 18)	5	0.5	0.07	250	1	0.7	27.80%
	FW (*p* = 18)	3	0.7	0.15	100	0.4	0.7	11.20%
	SFFS (*p* = 3)	3	0.9	0.15	80	0.8	0.7	9.30%

4.4.2 Optimal waveband selection

The wavebands selected by the KW filter, SFFS wrapper, and FW approach are shown in Table 4.3, and their respective locations illustrated in Figure 4.3. It is evident from Figure 4.3 that the location of the selected wavebands differs for the three feature selection approaches. For example, the KW filter selected wavebands exclusively in the blue region (473.92-491.95 nm) of the EM spectrum. These wavebands were also present in the subset derived using the FW approach (see Table 4.3). The blue wavebands selected by both the KW and FW approaches were closely related in terms of wavelengths. This could indicate that the KW filter and the FW approach may not be optimal for reducing the multicollinearity within the present dataset.

In comparison, the FW approach selected wavebands across the blue and green regions for both RF (473.92-585.12 nm) and XGBoost (474.74-582.43 nm). These selected wavebands correspond favourably to wavebands reported by Pôças et al. (2015) and Loggenberg et al. (2018), highlighting the feasibility of employing narrow wavebands in the VIS to model vineyard water stress. Notably, the FW approach selected a greater number of wavebands for RF (p = 35) compared with XGBoost (p = 18). Similar results were found by Georganos et al. (2018a). However, the additional features in the FW-RF subset may indicate RF's resistance to the presence of redundant and/or irrelevant wavebands. Moreover, as shown in Table 4.3, the FW approach selected wavebands common to both RF and XGBoost. These common wavebands (p = 18) were located across the blue (474.74-491.95 nm) and green (578.48-582.43 nm) regions of the EM spectrum. Loggenberg et al. (2018) reported similar wavebands when employing internal measures of variable importance to reduce dimensionality. These findings suggest that these wavebands (p = 18) may be the most important for water stress modelling in a Shiraz vineyard. However, this requires further investigation.

Of the three feature selection approaches evaluated in this study, the SFFS wrapper approach yielded the smallest subsets for RF (p = 4) and XGBoost (p = 3). For XGBoost, wavebands were located in the blue (496.45 nm) and green (521.32 nm and 585.12 nm), and in the blue (475.58 nm and 488.41 nm), green (578.48 nm), and red (644.22 nm) for RF. Loggenberg et al. (2018) reported the use of blue (474.74 nm and 497.36 nm) and green (521.32 nm, 578.48 nm and 585.12 nm) wavebands as an indicator of vineyard water status, which correspond to similar wavebands present in the SFFS-XGBoost and SFFS-RF subsets. These wavebands may be significant in modelling crop water stress, as the blue and green regions are highly sensitive to plant pigment (i.e. carotenoid and chlorophyll pigments) absorption (Pôças et al. 2015; Zygielbaum et al. 2009). Vegetative water stress is often expressed as an increase in blue and green reflectance (Zygielbaum et al. 2009).

Table 4.3 RF and XGBoost important wavebands as determined by the KW, FW, and SFFS feature selection approaches. Common wavebands are highlighted in bold.

	Classifier	
	RF (nm)	XGBoost (nm)
KW	473.92, 474.74, 475.58, 476.41, 478.09, 478.94, 479.78, 480.63, 483.20, 484.06, 484.92, 485.79, 487.53, 488.41, 489.29, 490.17, 491.06, 491.95	
FW	473.92, **474.74**, **475.58**, **476.41**, 477.25, **478.09**, **478.94**, **479.78**, 480.63, 481.48, **482.34**, 483.20, **484.06**, 484.92, **485.79**, 486.66, **487.53**, **488.41**, **489.29**, **490.17**, **491.06**, **491.95**, 492.84, 493.74, 494.64, 495.54, 497.36, 504.76, 577.17, **578.48**, 579.79, **581.11**, **582.43**, 583.77, 585.12	**474.74**, **475.58**, **476.41**, **478.09**, **478.94**, **479.78**, **482.34**, **484.06**, **485.79**, **487.53**, **488.41**, **489.29**, **490.17**, **491.06**, **491.95**, **578.48**, **581.11**, **582.43**
SFFS	475.58, 488.41, 578.48, 644.22	496.45, 521.32, 585.12

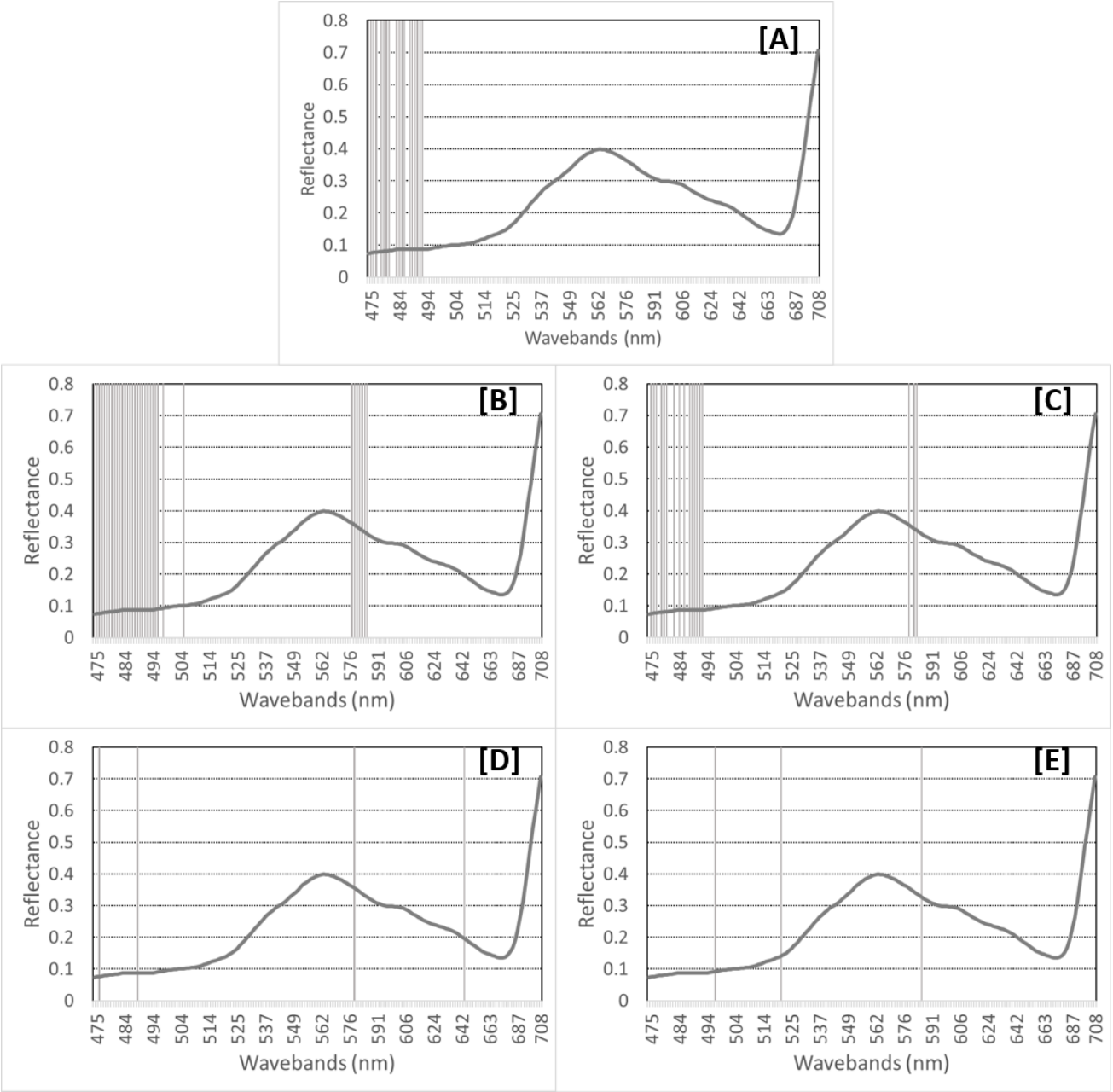

Figure 4.3 The important wavebands as determined by KW (A); FW with RF (B); FW with XGBoost (C); SFFS with RF (D); and SFFS with XGBoost (E). Grey bars indicate important wavebands. The mean spectra of a sample is shown as a reference.

4.4.3 RF and XGBoost classification

Table 4.4 shows the classification results for RF and XGBoost. For all models, the optimised hyperparameter values yielded improved classification accuracies, ranging from 0.8% to 5.5%. Using all wavebands (p = 176) as input yielded the best-performing models overall, producing a test accuracy of 83.3% (KHAT = 0.67) for RF and 81.7% (KHAT = 0.63) for XGBoost. These results compare favourably with work done by Vélez Rivera et al. (2014) and Abdel-Rahman et al. (2015), who found machine learning classifiers to perform best when using all wavebands. These results could indicate that both the RF and XGBoost ensembles are insensitive to the curse of dimensionality.

Furthermore, the results indicate that RF outperformed XGBoost when using all wavebands. Similar findings were reported by Loggenberg et al. (2018) who also showed RF (83.3%, KHAT = 0.67) to outperform XGBoost (78.3%, KHAT = 0.57).

Subsets generated using the KW filter yielded the lowest overall classification accuracies. Test accuracies for all models were found to be less than 60.0%. The decreased accuracies may be attributed to the selected wavebands (Table 4.3), with the subset comprised entirely of wavebands located in the blue region of the EM spectrum. The blue region is most often used in combination with longer wavelengths (Maimaitiyiming et al. 2017; Pôças et al. 2015) for vineyard water stress modelling.

In comparison, the waveband subsets selected by the SFFS wrapper and FW approach yielded higher accuracies, yielding a test accuracy of 80.0% (KHAT = 0.60) for both RF and XGBoost. When compared with using all wavebands, the SFFS wrapper subsets resulted in reduced test accuracies for XGBoost (1.7%) and RF (3.3%). However, the SFFS wrapper subsets achieved these results utilising only 2% (approximately 98% reduction in dimensionality) of the original waveband dataset. Moreover, the FW approach presented here provides flexibility as it can be employed across classifiers and combined with different waveband ranking, i.e. filter, approaches.

Table 4.4 RF and XGBoost classification results. Results for the best-performing and worst-performing models are highlighted in bold.

Feature Selection	Parameters	Dataset	**RF**		**XGB**	
			Accuracy (%)	Kappa	Accuracy (%)	Kappa
All Wavebands (*p = 176*)	**Default**	Train	90.0	0.80	86.7	0.73
		Test	80.0	0.60	78.3	0.57
	Optimised	Train	**93.3**	**0.87**	**91.7**	**0.83**
		Test	**83.3**	**0.67**	**81.7**	**0.63**
KW (*p = 18*)	**Default**	Train	68.3	0.37	66.7	0.33
		Test	56.7	0.13	53.3	0.07
	Optimised	Train	**71.7**	**0.43**	**72.2**	**0.43**
		Test	**57.5**	**0.15**	**58.3**	**0.17**
FW *p = 35 (RF) & p = 18 (XGBoost)*	**Default**	Train	86.7	0.73	86.7	0.73
		Test	78.7	0.57	80.0	0.60
	Optimised	Train	88.3	0.77	88.8	0.77
		Test	80.0	0.60	80.0	0.60
SFFS *p = 4 (RF) & p = 3 (XGBoost)*	**Default**	Train	90.0	0.80	88.3	0.77
		Test	80.0	0.60	80.0	0.60
	Optimised	Train	91.7	0.83	90.7	0.80
		Test	80.0	0.60	80.0	0.60

4.4.4 Comparison of computational expense

Additionally, the study recorded the computational expense of the three feature selection approaches (Table 4.5). The KW filter took 2.59 seconds to run. However, the waveband subset derived using the KW filter yielded the lowest classification accuracies. The SFFS wrapper subsets yielded the highest classification accuracies, but required 1 402.33 seconds (approximately 23 minutes) to run for RF and 5 820.24 seconds (approximately 1.6 hours) for XGBoost. In comparison, the FW approach required less processing time, approximately 2 minutes for RF and 7 minutes for XGBoost, to produce subsets of equivalent predictive competency.

It is evident from these results that algorithm optimisation does improve classification accuracy. However, it can be argued that the implementation of hyperparameter optimisation is application-specific, i.e. marginal increases in accuracy may not always warrant the added computational expense. Additionally, the choice of classifier should be considered when employing optimisation. For example, in the present study, RF optimisation resulted in marginal increases in accuracy, with minimal computational cost; 267.61 seconds using all wavebands (p =176). In comparison, XGBoost optimisation was computationally more expensive; 19 646.84 seconds using all wavebands. Chen & Guestrin (2016) and Xia et al. (2017) recommend the optimisation of XGBoost hyperparameter values. However, the results of the present study indicate that the gain in predictive competency does not justify the greater increase in computational expense.

The longer processing time observed for XGBoost optimisation and feature selection may be attributed to the classifier's use of the greedy search algorithm. Greedy search algorithms selects the best possible solution at each node without considering the overall outcome, hence the term greedy (Xia et al. 2017). XGBoost utilises an exact greedy search algorithm (Chen & Guestrin 2016; Xia et al. 2017) for finding the optimal tree structure. Xia et al. (2017) noted that this method is computationally expensive when employed on high-dimensional datasets, such as hyperspectral data. In contrast, RF uses random coefficients, determined through bagging with replacement and randomisation (Breiman 2001), to find the optimal split for each tree, which is computationally more efficient. Moreover, as asserted by Xia et al. (2017), the longer processing times for the XGBoost algorithm may be explained by the slower learning rates (i.e. low *eta* values) utilised in the present study.

Consequently, this study shows that there is an inherent trade-off between dimensionality reduction, classification accuracy and computational expense. The FW approach presented in this study was the most successful in lessening this trade-off. High classification accuracies were obtained, using only

10% to 20% of the original waveband dataset (equivalent to an 80% to 90% reduction in dimensionality), at minimal computational expense. Although the FW approach warrants further investigation, it has demonstrated its potential operational capability.

Table 4.5 RF and XGBoost computational expense for feature selection and hyperparameter optimisation.

		Processing Time(s)			
		KW	FW	SFFS	All Wavebands
Optimisation	RF	27.8	14.75	31.36	267.61
	XGBoost	17 803.06	17 742.21	18 898.73	19 646.84
Feature Selection	RF	2.59	87.55	1 420.33	-
	XGBoost		402.66	5 820.24	

4.5 CONCLUSION

This study investigated the efficacy of the KW filter, SFFS wrapper, and FW approach for waveband selection of terrestrial hyperspectral imaging data. The predictive competency of the generated subsets was evaluated using the RF and XGBoost machine learning classifiers. The classifiers were employed using both default and optimised hyperparameter values. Based on the findings of this study, the following conclusions are drawn:

1. Both the RF and XGBoost ensemble learners have shown to be insensitive to the curse of dimensionality.
2. The VIS region of the EM spectrum shows promise in discriminating between water-stressed and non-stressed Shiraz vines.
3. Optimising RF and XGBoost hyperparameter values does lead to increased classification accuracies. However, careful consideration should be given to the choice of classifier and the application.
4. The FW approach to feature selection demonstrated considerable promise in both predictive competency and as a means to lessen computational expense.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

This chapter concludes the thesis by providing a summary and critical evaluation of the findings obtained in the present research. The research aim and objectives are revisited, the research potential and limitations are highlighted, and recommendations for future research are made.

5.1 REVISITING THE AIM AND OBJECTIVES

Water is a major limiting factor for viticulture production (Nelson et al. 2014) and requires management of the utmost precision to ensure sustainable vineyard production. Recently, given the unpredictability of climate change and the variability in rainfall patterns, a greater emphasis has been placed on the cultivation of water management schemes and precision irrigation (Hannah et al. 2013). To this end, the present study aimed to develop a remote sensing–machine learning framework to model water stress in a Shiraz vineyard using terrestrial hyperspectral imaging.

The study comprised of two components, which were addressed in Chapter 3 and Chapter 4, respectively. Proximal remote sensing techniques and a hyperspectral imaging spectrometer were utilised to collect data for water-stressed and non-stressed Shiraz vines. These vines were identified using in situ stem water potential measurements.

Objectives 1 and 2 set out to evaluate the utility of terrestrial hyperspectral imaging, in combination with the RF and XGBoost ensemble learners, to model water stress in a Shiraz vineyard. These objectives were addressed in Chapter 3 by generating classification models for RF and XGBoost to discriminate water-stressed from non-stressed vines. In so doing, objectives 1 and 2 were fully satisfied.

Objective 3, addressed in Chapter 4, aimed to improve the results achieved in Chapter 3 by incorporating semi-automated feature selection approaches and hyperparameter value optimisation. Chapter 4 investigated the KW filter, the SFFS wrapper, and the FW approach as a means for semi-automated feature selection. It further compared the predictive competency of default and optimised hyperparameter values for the RF and XGBoost ensembles, thereby fulfilling Objective 3.

5.2 KEY FINDINGS AND POTENTIAL OF TECHNIQUES

In Chapter 3, the results indicated that both the RF and XGBoost classifiers could effectively be used to discriminate water-stressed from non-stressed Shiraz vines. Chapter 3 also demonstrated the feasibility of terrestrial hyperspectral imagery to model vineyard water stress. The effectiveness of using internal measures of VI for dimensionality reduction, provided by both the RF and XGBoost classifiers, was also confirmed in Chapter 3. The VI selected waveband subsets did not lead to notable

gains in test accuracy. However, when compared with using all wavebands (p = 176), the VI subsets produced similar accuracies using only 10% (p = 18) of the original waveband dataset.

The findings of Chapter 4 indicated that although hyperparameter value optimisation can be processing-intensive, it does lead to improved predictive competency. However, the marginal increase in predictive competency gained from optimising hyperparameter values may not justify the added computational expense. The KW filter, the SFFS wrapper, and the FW feature selection approach yielded no increase in classification accuracy when compared with using all wavebands (p = 176). However, they did improve classification efficiency. Furthermore, the developed FW feature selection approach produced encouraging results and was significantly less computationally expensive when compared with the more commonly used SFFS wrapper. The FW approach could hold great potential for dimensionality reduction in operational frameworks as its computational efficiency would save valuable time and, therefore, be more cost-effective.

A key finding of both Chapters 3 and 4 was that wavebands in the VIS region of the EM spectrum could successfully be utilised to model water stress in a Shiraz vineyard. The use of the VIS region to model water stress holds great potential for viticulture and the broader agricultural field. Traditionally, vegetative water stress has been modelled using NIR/SWIR (Sai-Sesha et al. 2016; Zhang et al. 2017) and thermal sensors (García-Tejero et al. 2016; Gerhards et al. 2016). These sensors are often significantly more costly than VIS sensors. Consequently, the use of less expensive VIS sensors could make accurate water stress modelling, and by extension accurate precision irrigation, more accessible to smaller farms.

The selection of important wavebands, as demonstrated in Chapters 3 and 4, lends itself to the development of customised sensors that are less expensive and specifically designed to model vineyard water stress. Moreover, the developed framework provides a low-cost method for water stress detection that can easily be utilised across different vineyard cultivars and could potentially be deployed on different agricultural crops. The machine learning–remote sensing framework can also be easily scaled to work with airborne and satellite imagery to detect not only water stress but also other forms vegetative stress, such as disease infestation (Adam et al. 2017) and pesticide residue (Mohite et al. 2017).

5.3 LIMITATIONS, RECOMMENDATIONS AND FUTURE RESEARCH

The research demonstrated great potential. However, as the study utilised data for a single cultivar captured for a single day, the results remain limited to specific environmental conditions. As such, the use of multi-temporal data captured for different cultivars is highly recommended for future

studies. The addition of multi-temporal data would help determine what growing phase is optimal for water stress detection and, in so doing, help prevent crop loses and ensure high-quality grape yield.

The study utilised in-field measurements to determine water-stressed and non-stressed Shiraz vines. However, the effects of other stressing mechanisms (such as diseases and potassium deficiency) were not ruled out, which presents a degree of uncertainty within the present study. Additionally, the range of wavebands available also limited the present study to only using the VIS region of the EM spectrum. By incorporating the full EM spectra (i.e. 350-2500 nm), the predictive prowess of VIS, NIR, and SWIR wavebands could be compared and, thereby, help determine which region of the EM spectrum is most optimal for vineyard water stress modelling.

The use of spectral smoothing as a pre-processing step for hyperspectral data also warrants further investigation. The present work only meant to test the functionality of the Savitzky-Golay filter. The filter was therefore only employed using recommended hyperparameter values. Optimising hyperparameter values for the Savitzky-Golay filter could potentially have led to improved classification results.

An important facet of the developed framework is the optimisation of hyperparameter values, which was performed for both classifiers before and after feature selection. A review of the literature (Abdel-Rahman et al. 2014; Eisavi et al. 2015; Georganos et al. 2018b; Poona et al. 2016; Taşkın, Hüseyin & Bruzzone 2017) indicated that the effects of re-optimising hyperparameter values after feature selection was mainly unexplored. Therefore, future research could investigate whether re-optimisation is indeed warranted after performing feature selection. Furthermore, a comparison study of alternative hyperparameter optimisation methods, such as Bayesian (Xia et al. 2017) and Gradient-based (Maclaurin, Duvenaud & Adams 2015) optimisation, for the RF and XGBoost ensembles would be greatly beneficial to the broader research community.

A major limiting factor inhibiting the operationalisation of the developed framework is the availability of hyperspectral sensors. To date, hyperspectral sensors are still very costly (Camps-Valls et al. 2014; Wendisch & Brenguier 2013), which makes the acquisition of hyperspectral data problematic. However, with the rapid advancement of technology, cheaper hyperspectral sensors could be available in the near future.

5.4 CONCLUSION

Vineyard water stress poses a great threat to the quality and sustainability of grape production. It is therefore imperative to develop frameworks that are capable of detecting water-stressed vines. These frameworks can prevent devastating crop loses and help better manage water resources. Traditionally,

water stress has been modelled using thermal datasets, vegetative indices or identified through laborious in-field measurements.

This thesis presents a novel hyperspectral–machine learning framework for the non-destructive identification of water-stressed vines. The results indicated the viability of narrow VIS wavebands to model vineyard water stress and established the utility of machine learning ensembles within the domain of viticulture. Furthermore, the findings also confirmed the robustness of the RF and XGBoost classifiers to the curse of dimensionality. The study provides a point of departure for the operationalisation of future machine learning–remote sensing frameworks for water stress monitoring. In so doing, this will contribute to the sustainability and continued growth of the wine sector.

Word count: 21 620

REFERENCES

Aasen H, Burkart A, Bolten A & Bareth G 2015. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. *ISPRS Journal of Photogrammetry and Remote Sensing* 108: 245-259.

Abdel-Rahman EM, Makori DM, Landmann T, Piiroinen R, Gasim S, Pellikka P & Raina SK 2015. The utility of AISA eagle hyperspectral data and random forest classifier for flower mapping. *Remote Sensing* 7, 10: 13298-13318.

Abdel-Rahman EM, Mutanga O, Adam E & Ismail R 2014. Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing* 88: 48-59.

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* 2017: 1-8.

Al-Saddik H, Simon JC & Cointault F 2017. Development of spectral disease indices for "flavescence dorée" grapevine disease identification. *Sensors* 17, 12: 1-25.

Ali I, Greifeneder F, Stamenkovic J, Neumann M & Notarnicola C 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing* 7, 12: 16398-16421.

Andrew ME & Ustin SL 2008. The role of environmental context in mapping invasive plants with hyperspectral image data. *Remote Sensing of Environment* 112, 12: 4301-4317.

Baluja J, Diago MP, Balda P, Zorer R, Meggio F, Morales F & Tardaguila J 2012. Assessment of vineyard water status variability by thermal and multispectral imagery using an unmanned aerial vehicle (UAV). *Irrigation Science* 30, 6: 511-522.

Beghi R, Giovenzana V & Guidetti R 2017. Better water use efficiency in vineyard by using visible and near infrared spectroscopy for grapevine water status monitoring. *Chemical Engineering Transactions* 58: 691-696.

De Bei R, Cozzolino D, Sullivan W, Cynkar W, Fuentes S, Dambergs R, Pech J & Tyerman SD 2011. Non-destructive measurement of grapevine water potential using near infrared spectroscopy. *Australian Journal of Grape and Wine Research* 17, 1: 62-71.

Belgiu M & Drăguţ L 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24-31.

Belgiu M, Tomljenovic I, Lampoltshammer TJ, Blaschke T & Höfle B 2014. Ontology-based classification of building types detected from airborne laser scanning data. *Remote Sensing* 6, 2: 1347-1366.

Bellvert J, Zarco-Tejada PJ, Girona J & Fereres E 2014. Mapping crop water stress index in a Pinot-noir vineyard: Comparing ground measurements with thermal remote sensing imagery from an unmanned aerial vehicle. *Precision Agriculture* 15, 4: 361-376.

Bioucas-dias JM, Plaza A, Camps-valls G, Scheunders P, Nasrabadi NM & Chanussot J 2013. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine* 1, 2: 6-36.

Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, Casalicchio G & Jones Z 2016. mlr: Machine learning in R. *Journal of Machine Learning Research* 17, 170: 1-5.

Bota J, Tomás M, Flexas J, Medrano H & Escalona JM 2016. Differences among grapevine cultivars in their stomatal behavior and water use efficiency under progressive water stress. *Agricultural Water Management* 164: 91-99.

Breiman L, Friedman J, Stone C & Olshen R 1984. *Classification and regression trees*. 1st ed. Wadsworth, Belmon: CRC Press.

Breiman L 2001. Random forests. *Machine Learning* 45, 1: 5-32.

Breiman L 1996. Bagging predictors. *Machine Learning* 24, 2: 123-140.

Camps-Valls G, Tuia D, Bruzzone L & Benediktsson JA 2014. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine* 31, 1: 45-54.

Cancela JJ, Fandiño M, Rey BJ, Dafonte J & González XP 2017. Discrimination of irrigation water management effects in pergola trellis system vineyards using a vegetation and soil index. *Agricultural Water Management* 183: 70-77.

Candiago S, Remondino F, Giglio M De, Dubbini M & Gattelli M 2015. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sensing* 7, Vi: 4026-4047.

Cao X, Wei C, Han J & Jiao L 2017. Hyperspectral band selection using improved classification map. *IEEE Geoscience and Remote Sensing Letters* 14, 11: 2147-2151.

Carreiro Soares SF, Medeiros EP, Pasquini C, de Lelis Morello C, Harrop Galvão RK & Ugulino Araújo MC 2016. Classification of individual cotton seeds with respect to variety using near-

infrared hyperspectral imaging. *Analytical Methods* 8, 48: 8498-8505.

Carvalho AA, Rebello JMA, Sagrilo LVS, Camerini CS & Miranda IVJ 2006. MFL signals and artificial neural networks applied to detection and classification of pipe weld defects. *NDT and E International* 39, 8: 661-667.

Cen H, Lu R, Zhu Q & Mendoza F 2016. Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification. *Postharvest Biology and Technology* 111: 352-361.

Chan JCW & Paelinckx D 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment* 112, 6: 2999-3011.

Chandrashekar G & Sahin F 2014. A survey on feature selection methods. *Computers and Electrical Engineering* 40, 1: 16-28.

Chen C, Jiang F, Yang C, Rho S, Shen W, Liu S & Liu Z 2018. Hyperspectral classification based on spectral–spatial convolutional neural networks. *Engineering Applications of Artificial Intelligence* 68: 165-171.

Chen T & Guestrin C 2016. *XGBoost: A scalable tree boosting system*. Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA. ACM. 785-794.

Chen T, He T, Benesty M, Khotilovich V & Tang Y 2017. xgboost: Extreme gradient boosting.

Cheng X, Chen YR, Tao Y, Wang CY, Kim MS & Lefcourt AM 2004. A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. *Transactions of the ASAE* 47, 4: 1313-1320.

Chirouze J, Boulet G, Jarlan L, Fieuzal R, Rodriguez JC, Ezzahar J, Er-Raki S, Bigeard G, Merlin O, Garatuza-Payan J, Watts C & Chehbouni G 2014. Intercomparison of four remote-sensing-based energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate. *Hydrology and Earth System Sciences* 18, 3: 1165-1188.

Choné X, Van Leeuwen C, Dubourdieu D & Gaudillère JP 2001. Stem water potential is a sensitive indicator of grapevine water status. *Annals of Botany* 87, 4: 477-483.

Člupek M, Matějka P & Volka K 2007. Noise reduction in Raman spectra: Finite impulse response filtration versus Savitzky–Golay smoothing. *Journal of Raman Spectroscopy* 38, 9: 1174-1179.

Congalton RG & Green K 2009. *Assessing the accuracy of remotely sensed data: principles and*

practices. 2nd ed. Boco Raton, FL, USA: CRC press.

Conradie WJ, Carey VA, Bonnardot V, Saayman D & Van Schoor LH 2002. Effect of different environmental factors on the performance of sauvignon blanc grapevines in the Stellenbosch / Durbanville districts of South Africa. I. Geology, Soil, Climate, Phenology and Grape Composition. *South African Journal for Enology and Viticulture* 23, 2: 78-91.

Corcoran JM, Knight JF & Gallant AL 2013. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota. *Remote Sensing* 5, 7: 3212-3238.

Costa JM, Vaz M, Escalona J, Egipto R, Lopes C, Medrano H & Chaves MM 2016. Modern viticulture in southern Europe: Vulnerabilities and strategies for adaptation to water scarcity. *Agricultural Water Management* 164: 5-18.

Datta A, Ghosh S & Ghosh A 2017. Unsupervised band extraction for hyperspectral images using clustering and kernel principal component analysis. *International Journal of Remote Sensing* 38, 3: 850-873.

Del-Moral-Martínez I, Rosell-Polo JR, Company J, Sanz R, Escolà A, Masip J, Martínez-Casasnovas JA & Arnó J 2016. Mapping vineyard leaf area using mobile terrestrial laser scanners: Should rows be scanned on-the-go or discontinuously sampled? *Sensors* 16, 1: 1-13.

Deloire A & Heyms D 2011. The leaf water potentials: principles, method and thresholds. *WineLand Magazine South Africa*: 119-121.

Dev S, Wen B, Lee YH & Winkler S 2016. Machine learning techniques and applications for ground-based image analysis. *IEEE Geoscience and Remote Sensing Magazine* 4, 2: 79-93.

Diago MP, Bellincontro A, Scheidweiler M, Tardaguila J, Tittmann S & Stoll M 2017. Future opportunities of proximal near infrared spectroscopy approaches to determine the variability of vineyard water status. *Australian Journal of Grape and Wine Research* 23, 3: 409-414.

Eisavi V, Homayouni S, Yazdi AM & Alimohammadi A 2015. Land cover mapping based on random forest classification of multitemporal spectral and thermal images. *Environmental Monitoring and Assessment* 187, 5: 187-291.

Eismann MT 2012. *Hyperspectral remote sensing.* Bellingham, WA: SPIE.

Exelis Visual Information Solutions 2015. Boulder, Colorado.

Fitriah N, Wijaya SK, Fanany MI, Badri C & Rezal M 2017. *EEG channels reduction using PCA to increase XGBoost's accuracy for stroke detection*. Proceedings of the 2nd International

Symposium on Current Progress in Mathematics and Sciences (2016). Depok, Jawa Barat, Indonesia. AIP Publishing.

Font D, Tresanchez M, Martínez D, Moreno J, Clotet E & Palacín J 2015. Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors* 15, 4: 8284-8301.

Foody GM 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing* 70, 5: 627-633.

Freund Y & Schapire RE 1996. *Experiments with a new boosting algorithm*. Proceedings of the Thirteenth International Conference on Machine Learning. Bari, BA, Italy. Kaufmann. 148-156.

Friedman J 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*: 1189-1232.

Friedman JH 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38, 4: 367-378.

Fu Y, Zhao C, Wang J, Jia X, Yang G, Song X & Feng H 2017. An improved combination of spectral and spatial features for vegetation classification in hyperspectral images. *Remote Sensing* 9, 3: 1-16.

García-Tejero IF, Costa JM, Egipto R, Durán-Zuazo VH, Lima RSN, Lopes CM & Chaves MM 2016. Thermal data to monitor crop-water status in irrigated Mediterranean viticulture. *Agricultural Water Management* 176: 80-90.

Genc L, Inalpulat M, Kizil U, Mirik M, Smith SE & Mendes M 2013. Determination of water stress with spectral reflectance on sweet corn (*Zea mays L.*) using classification tree (CT) analysis. *Zemdirbyste-Agriculture* 100, 1: 81-90.

Di Gennaro SF, Battiston E, Di Marco S, Facini O, Matese A, Nocentini M, Palliotti A & Mugnai L 2016. Unmanned aerial vehicle (UAV)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex. *Phytopathologia Mediterranea* 55, 2: 262-275.

Genuer R, Poggi J-M & Tuleau-Malot C 2010. Variable selection using random forests. *Pattern Recognition Letters* 31, 14: 2225-2236.

Georganos S, Grippa T, Vanhuysse S, Lennert M, Shimoni M, Kalogirou S & Wolff E 2018a. Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience and Remote Sensing* 55, 2: 221-242.

Georganos S, Grippa T, Vanhuysse S, Lennert M, Shimoni M & Wolff E 2018b. Very high resolution object-based land use-land cover urban classification using extreme gradient boosting. *IEEE Geoscience and Remote Sensing Letters* 15, 4: 607-611.

Gerhards M, Rock G, Schlerf M & Udelhoven T 2016. Water stress detection in potato plants using leaf temperature, emissivity, and reflectance. *International Journal of Applied Earth Observation and Geoinformation* 53: 27-39.

Goel PK, Prasher SO, Patel RM, Landry JA, Bonnell RB & 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* 39, 2: 67-93.

Goldstein BA, Hubbard AE, Cutler A & Barcellos LF 2010. An application of random forests to a genome-wide association dataset: methodological considerations & new findings. *BMC genetics* 11, 49: 1-13.

González-Fernández AB, Rodríguez-Pérez JR, Marcelo V & Valenciano JB 2015. Using field spectrometry and a plant probe accessory to determine leaf water content in commercial vineyards. *Agricultural Water Management* 156: 43-50.

Govender M, Dye P, Witokwski E & Ahmed F 2009. Review of commonly used remote sensing and ground based technologies to measure plant water stress. *Water SA* 35, 5: 741-752.

Gutiérrez S, Tardaguila J, Fernández-Novales J & Diago MP 2016. Data mining and NIR spectroscopy in viticulture: Applications for plant phenotyping under field conditions. *Sensors* 16, 2: 1-15.

Hannah L, Roehrdanz PR, Ikegami M, Shepard A V, Shaw MR, Tabor G, Zhi L, Marquet PA & Hijmans RJ 2013. Climate change, wine, and conservation. *PNAS* 110, 17: 6907-6912.

Harrison D, Rivard B & Sánchez-Azofeifa A 2018. Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest. *International Journal of Applied Earth Observation and Geoinformation* 66: 93-105.

Hughes GF 1968. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory* 14, 1: 55-63.

Ihuoma SO & Madramootoo CA 2017. Recent advances in crop water stress detection. *Computers and Electronics in Agriculture* 141: 267-275.

Immitzer M, Atzberger C & Koukal T 2012. Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. *Remote Sensing* 4, 9: 2661-2693.

Jensen CA, El-Sharkawi MA & Marks RJ 2001. Power system security assessment using neural networks: feature selection using Fisher discrimination. *IEEE Transactions on Power Systems* 16, 4: 757-763.

Jolliffe IT 1986. Principal component analysis and factor analysis. In *Principal Component Analysis. Springer Series in Statistics*, 115-128. New York, NY: Springer.

Jović A, Brkić K & Bogunović N 2015. *A review of feature selection methods with applications*. Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2015). Opatija, Croatia. MIPRO. 1200-1205.

Jung R & Ehlers M 2016. Comparison of two feature selection methods for the separability analysis of intertidal sediments with spectrometric datasets in the German Wadden Sea. *International Journal of Applied Earth Observation and Geoinformation* 52: 175-191.

Kalisperakis I, Stentoumis C, Grammatikopoulos L & Karantzalos K 2015. *Leaf area index estimation in vineyards from UAV hyperspectral data, 2D image mosaics and 3D canopy surface models*. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Proccedings of the 3rd International Conference on Unmanned Aerial Vehicles in Geomatics. Toronto, Canada. ISPRS. 299-303.

Karakizi C, Oikonomou M & Karantzalos K 2016. Vineyard detection and vine variety discrimination from very high resolution satellite data. *Remote Sensing* 8, 3: 1-25.

Kawaguchi S & Nishii R 2007. Hyperspectral image classification by bootstrap adaboost with random decision stumps. *IEEE Transactions on Geoscience and Remote Sensing* 45, 11: 3845-3851.

Kejela G & Rong C 2016. *Cross-device consumer identification*. Proceedings of the 15th IEEE International Conference on Data Mining Workshop (ICDMW 2015). Atlantic City, NJ, USA. IEEE. 1687-1689.

Kim Y, Glenn DM, Park J, Ngugi HK & Lehman BL 2011. Hyperspectral image analysis for water stress detection of apple trees. *Computers and Electronics in Agriculture* 77, 2: 155-160.

Knauer U, Matros A, Petrovic T, Zanker T, Scott ES & Seiffert U 2017. Improved classification accuracy of powdery mildew infection levels of wine grapes by spatial-spectral analysis of hyperspectral images. *Plant Methods*: 1-15.

Kohavi R & Provost F 1998. Glossary of terms. *Machine Learning* 30: 271-274.

Kruskal WH & Wallis WA 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47, 260: 583-621.

Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Candan C, Tang Y & Hunt T 2017. caret: Classification and regression training.

Lagrange A, Fauvel M & Grizonnet M 2017. Large-scale feature selection with gaussian mixture models for the classification of high dimensional remote sensing images. *IEEE Transactions on Computational Imaging* 3, 2: 230-242.

Lawrence R, Bunn A, Powell S & Zambon M 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment* 90, 3: 331-336.

Lawrence RL, Wood SD & Sheley RL 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment* 100, 3: 356-362.

Lewis DD 1992. *Feature selection and feature extraction for text categorization*. Proceedings of the Workshop on Speech and Natural Language. Harriman, NY, USA. Association for Computational Linguistics. 212-217.

Li S, Wu H, Wan D & Zhu J 2011. An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems* 24, 1: 40-48.

Li W, Feng F, Li H & Du Q 2018. Discriminant Analysis-based dimension reduction for hyperspectral image classification: A survey of the most recent advances and an experimental comparison of different techniques. *IEEE Geoscience and Remote Sensing Magazine* 6, 1: 15-34.

Liaw A & Wiener M 2002. Classification and regression by randomForest. *R news* 2, December: 18-22.

Ligges U, Short T & Kienzle P 2015. signal: Signal processing.

Liu H & Setiono R 1995. *Chi2: Feature selection and discretization of numeric attributes*. Proceedings of the 7th IEEE International Conference on Tools with Artificial Intelligence. Herndon, VA, USA. IEEE. 388-391.

Liu L, Ji M, Dong Y, Zhang R & Buchroithner M 2016. Quantitative retrieval of organic soil properties from visible near-infrared shortwave infrared feature extraction. *Remote Sensing* 8, 12: 1035.

Loggenberg K, Strever A, Greyling B & Poona N 2018. Modelling water stress in a Shiraz vineyard using hyperspectral imaging and machine learning. *Remote Sensing* 10, 2: 1-14.

Lopo M, Teixeira dos Santos CA, Páscoa RNMJ, Graça AR & Lopes JA 2018. Near infrared spectroscopy as a tool for intensive mapping of vineyards soil. *Precision Agriculture* 19, 3: 445-462.

Luo Y, Zou J, Yao C, Li T & Bai G 2018. HSI-CNN: A novel convolution neural network for hyperspectral image. *arXiv:1802.10478*.

Maclaurin D, Duvenaud D & Adams RP 2015. *Gradient-based hyperparameter optimization through reversible learning*. Proceedings of the 32nd International Conference on Machine Learning. Lille, France. JMLR, W&CP.

Maimaitiyiming M, Ghulam A, Bozzolo A, Wilkins JL & Kwasniewski MT 2017. Early detection of plant physiological responses to different levels of water stress using reflectance spectroscopy. *Remote Sensing* 9, 7: 745.

Martinez-de-Pison FJ, Gonzalez-Sendino R, Aldama A, Ferreiro J & Fraile E 2017. *Hybrid methodology based on bayesian optimization and ga-parsimony for searching parsimony models by combining hyperparameter optimization and feature selection*.de Pisón FJ Urraca R Quintián H & Corchado E (eds). In Lecture Notes in Computer Science, Proceedings of the 12th International Conference on Hybrid Artificial Intelligent Systems (HAIS 2017). La Rioja, Spain. Springer, Cham. 52-62.

Maschler J, Atzberger C & Immitzer M 2018. Individual tree crown segmentation and classification of 13 tree species using Airborne hyperspectral data. *Remote Sensing* 10, 8: 1-29.

Matese A, Baraldi R, Berton A, Cesaraccio C, Di Gennaro SF, Duce P, Facini O, Mameli MG, Piga A & Zaldei A 2018. Estimation of water stress in grapevines using proximal and remote sensing methods. *Remote Sensing* 10, 1: 1-16.

Matese A & Di Gennaro SF 2015. Technology in precision viticulture: A state of the art review. *International Journal of Wine Research* 7, 1: 69-81.

Matese A, Di Gennaro SF & Berton A 2016. Assessment of a canopy height model (CHM) in a vineyard using UAV-based multispectral imaging. *International Journal of Remote Sensing* 38, 8: 2150-2160.

Matese A, Toscano P, Di Gennaro SF, Genesio L, Vaccari FP, Primicerio J, Belli C, Zaldei A, Bianconi R & Gioli B 2015. Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sensing* 7, 3: 2971-2990.

Mathews AJ 2013. Applying geospatial tools and techniques to viticulture. *Geography Compass* 1, 7: 22-34.

Mathews AJ & Jensen JLR 2013. Visualizing and quantifying vineyard canopy LAI using an unmanned aerial vehicle (UAV) collected high density structure from motion point cloud. *Remote Sensing* 5, 5: 2164-2183.

Meadows ME 2003. Soil erosion in the Swartland, Western Cape Province, South Africa: Implications of past and present policy and practice. *Environmental Science and Policy* 6, 1: 17-28.

Medjahed SA, Ait Saadi T, Benyettou A & Ouali M 2016. Gray Wolf Optimizer for hyperspectral band selection. *Applied Soft Computing Journal* 40: 178-186.

Miao X, Heaton JS, Zheng S, Charlet DA & Liu H 2012. Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data. *International Journal of Remote Sensing* 33, 6: 1823-1849.

Mohite J, Karale Y, Pappula S, Shabeer T. P. A, Sawant SD & Hingmire S 2017. *Detection of pesticide (Cyantraniliprole) residue on grapes using hyperspectral sensing*. Kim MS Chao K l Chin BA & Cho BK (eds). In Sensing for Agriculture and Food Quality and Safety IX, Proceedings of the SPIE Commercial+ Scientific Sensing and Imaging Conference. Anaheim, CA, USA. International Society for Optics and Photonics: Bellingham,WA, USA.

Möller A, Ruhlmann-Kleider V, Leloup C, Neveu J, Palanque-Delabrouille N, Rich J, Carlberg R, Lidman C & Pritchet C 2016. Photometric classification of type Ia supernovae in the supernova legacy survey with supervised learning. *Journal of Cosmology and Astroparticle Physics* 2016, 12: 1-26.

Monteiro ST, Murphy RJ, Ramos F & Nieto J 2009. *Applying boosting for hyperspectral classification of ore-bearing rocks*. IEEE International Workshop on Machine Learning for Signal Processing. Grenoble, France. IEEE. 1-6.

Mulla DJ 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering* 114, 4: 358-371.

Myburgh P, Cornelissen R & Southey T 2016. Interpretation of stem water potential measurements. *WineLand Magazine South Africa*: 78-80.

Nelson GC, Valin H, Sands RD, Havlík P, Ahammad H, Deryng D & Elliott J 2014. Climate change effects on agriculture : Economic responses to biophysical shocks. *Proceedings of the National Academy of Sciences of the United States of America* 111, 9: 3274-3279.

Pappu V & Pardalos PM 2014. *Clusters, Orders, and Trees: Methods and Applications: In Honor of Boris Mirkin's 70th Birthday*. Aleskerov F Goldengorin B & Pardalos P (eds). New York, NY:

Springer.

Patteti S, Samanta B & Chakravarty D 2015. Design of a feature-tuned ANN model based on bulk rock-derived mineral spectra for endmember classification of a hyperspectral image from an iron ore deposit. *International Journal of Remote Sensing* 36, 8: 2037-2062.

Pedergnana M, Marpu PR & Mura MD 2013. A novel technique for optimal feature selection in attribute profiles based on genetic algorithms. *IEEE Transactions on Geoscience and Remote Sensing* 51, 6: 3514-3528.

Poblete-Echeverría C, Olmedo GF, Ingram B & Bardeen M 2017. Detection and segmentation of vine canopy in ultra-high spatial resolution RGB imagery obtained from unmanned aerial vehicle (UAV): A case study in a commercial vineyard. *Remote Sensing* 9, 3: 1-14.

Pôças I, Rodrigues A, Gonçalves S, Costa PM, Gonçalves I, Pereira LS & Cunha M 2015. Predicting grapevine water status based on hyperspectral reflectance vegetation indices. *Remote Sensing* 7, 12: 16460-16479.

Poona NK & Ismail R 2014. Using boruta-selected spectroscopic wavebands for the asymptomatic detection of fusarium circinatum stress. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 9: 3764-3772.

Poona NK, Van Niekerk A, Nadel RL & Ismail R 2016. Random forest (RF) wrappers for waveband selection and classification of hyperspectral data. *Applied Spectroscopy* 70, 2: 322-333.

Poona N, van Niekerk A & Ismail R 2016. Investigating the utility of oblique tree-based ensembles for the classification of hyperspectral data. *Sensors* 16, 11: 1-16.

Prasad KA, Gnanappazham L, Selvam V, Ramasubramanian R & Kar CS 2015. Developing a spectral library of mangrove species of Indian east coast using field spectroscopy. *Geocarto International* 30, 5: 580-599.

Pudil P, Novovičová J & Kittler J 1994. Floating search methods in feature selection. *Pattern Recognition Letters* 15, 11: 1119-1125.

Qiao T, Yang Z, Ren J, Yuen P, Zhao H, Sun G, Marshall S & Benediktsson JA 2018. Joint bilateral filtering and spectral similarity-based sparse representation: A generic framework for effective feature extraction and data classification in hyperspectral imaging. *Pattern Recognition* 77: 316-328.

Raczko E & Zagajewski B 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing* 50, 1: 144-154.

R Development Core Team 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Radovic M, Ghalwash M, Filipovic N & Obradovic Z 2017. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics* 18, 1: 1-14.

Reis MJCS, Morais R, Peres E, Pereira C, Contente O, Soares S, Valente A, Baptista J, Ferreira PJSG & Bulas Cruz J 2012. Automatic detection of bunches of grapes in natural environment from color images. *Journal of Applied Logic* 10, 4: 285-290.

Ren X, Guo H, Li S & Wang S 2017. *A novel image classification method with cnn-xgboost model*. International Workshop on Digital Watermarking. Magdeburg, Germany. Springer, Cham. 378-390.

Ricci A, Lagel M, Parpinello GP, Pizzi A, Kilmartin PA & Versari A 2016. Spectroscopy analysis of phenolic and sugar patterns in a food grade chestnut tannin. *Food Chemistry* 203: 425-429.

Rivera-Caicedo JP, Verrelst J, Muñoz-Marí J, Camps-Valls G & Moreno J 2017. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing* 132: 88-101.

Robnik-Sikonja M & Kononenko I 2003. Theoretical and empirical analysis of relieff and rrieff. *Journal of Machine Learning Research* 53: 23-69.

Rodríguez-Pérez JR, Riaño D, Carlisle E, Ustin S & Smart DR 2007. Evaluation of hyperspectral reflectance indexes to detect grapevine water status in vineyards. *American Journal of Enology and Viticulture* 58, 3: 302-317.

Rodrìguez JJ, Kuncheva LI & Alonso CJ 2006. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 10: 1619-1630.

Rogiers SY, Greer DH, Hatfield JM, Hutton RJ, Clarke SJ, Hutchinson PA & Somers A 2012. Stomatal response of an anisohydric grapevine cultivar to evaporative demand, available soil moisture and abscisic acid. *Tree Physiology* 32, 3: 249-261.

Rojas-Moraleda R, Valous NA, Gowen A, Esquerre C, Härtel S, Salinas L & O'Donnell C 2017. A frame-based ANN for classification of hyperspectral images: assessment of mechanical damage in mushrooms. *Neural Computing and Applications* 28: 969-981.

SA Wine Industry Information & Systems (SAWIS) 2016. *SA Wine Industry 2016 Statistics NR 4. South African Wine Industry Statistics.*

Sai-Sesha MVR, Murthy CS, Chandrasekar K, Jeyaseelan AT, Diwakar PG & Dadhwal VK 2016. Agricultural drought : Assessment & monitoring. *Mausam* 67, 1: 131-142.

Sandika B, Avil S, Sanat S & Srinivasu P 2016. *Random forest based classification of diseases in grapes from images captured in uncontrolled environments*. Proceedings of the 13th IEEE International Conference on Signal Processing. Chengdu, China. IEEE. 1775-1780.

Santara A, Mani K, Hatwar P, Singh A, Garg A, Padia K & Mitra P 2017. BASS Net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 55, 9: 5293-5301.

Sanz R, Rosell JR, Llorens J, Gil E & Planas S 2013. Relationship between tree row LIDAR-volume and leaf area density for fruit orchards and vineyards obtained with a LIDAR 3D dynamic measurement system. *Agricultural and Forest Meteorology* 171-172: 153-162.

Savitzky A & Golay MJ 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36, 8: 1627-1639.

Schapire RE 1990. The strength of weak learnability. *Machine Learning* 5, 2: 197-227.

Schmidt KS & Skidmore AK 2004. Smoothing vegetation spectra with wavelets. *International Journal of Remote Sensing* 25, 6: 1167-1184.

Semmens KA, Anderson MC, Kustas WP, Gao F, Alfieri JG, McKee L, Prueger JH, Hain CR, Cammalleri C, Yang Y, Xia T, Sanchez L, Mar Alsina M & Vélez M 2016. Monitoring daily evapotranspiration over two California vineyards using Landsat 8 in a multi-sensor data fusion approach. *Remote Sensing of Environment* 185: 155-170.

Shafri HZM, Suhaili A & Mansor S 2007. The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis. *Journal of Computer Science* 3, 6: 419-423.

Shimada S, Funatsuka E, Ooda M, Takyu M, Fujikawa T & Toyoda H 2012. Developing the monitoring method for plant water stress using spectral reflectance measurement. *Journal of Arid Land Studies* 22, 1: 251-254.

Shuaibu M, Lee WS, Schueller J, Gader P, Hong YK & Kim S 2018. Unsupervised hyperspectral band selection for apple *Marssonina* blotch detection. *Computers and Electronics in Agriculture* 148, June 2016: 45-53.

Singh A, Ganapathysubramanian B, Singh AK & Sarkar S 2016. Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science* 21, 2: 110-124.

Smit JL, Sithole G & Strever AE 2016. Vine signal extraction: An application of remote sensing in precision viticulture. *South African Journal of Enology and Viticulture* 31, 2: 65-74.

Tang J, Woods M, Cossell S, Liu S & Whitty M 2016. Non-productive vine canopy estimation through proximal and remote sensing. *IFAC-PapersOnLine* 49, 16: 398-403.

Taşkın G, Hüseyin K & Bruzzone L 2017. Feature selection based on high dimensional model representation for hyperspectral images. *IEEE Transactions on Image Processing* 26, 6: 2918-2928.

Tavares GM, Mastelini SM & Barbon Jr. S 2017. *User classification on online social networks by post frequency*. Proceedings of the 12th Brazilian Symposium on Information Systems on Brazilian Symposium on Information Systems. Florianopolis, Santa Catarina, Brazil. Brazilian Computer Society. 464-471.

Thorp KR, Wang G, Bronson KF, Badaruddin M & Mon J 2017. Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield. *Computers and Electronics in Agriculture* 136: 1-12.

Tong Q, Xue Y & Zhang L 2014. Progress in hyperspectral remote sensing science and technology in China over the past three decades. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 1: 70-91.

Torlay L, Perrone-Bertolotti M, Thomas E & Baciu M 2017. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. *Brain Informatics* 4, 3: 159-169.

Usha K & Singh B 2013. Potential applications of remote sensing in horticulture: A review. *Scientia Horticulturae* 153: 71-83.

Vélez Rivera N, Gómez-Sanchis J, Chanona-Pérez J, Carrasco JJ, Millán-Giraldo M, Lorente D, Cubero S & Blasco J 2014. Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning. *Biosystems Engineering* 122: 91-98.

Vora S & Yang H 2017. *A comprehensive study of eleven feature selection algorithms and their impact on text classification*. Proceedings of the 5th Computing Conference. Hilton Kensington, London, UK. IEEE. 440-449.

Waad B, Ghazi BM & Mohamed L 2013. *On the effect of search strategies on wrapper feature selection in credit scoring*. 2013 International Conference on Control, Decision and Information Technologies (CoDIT): 218-223.

Wendisch M & Brenguier JL 2013. *Airborne measurements for environmental research: Methods and instruments*. Wendisch M & Brenguier JL (eds). New Jersey, US: John Wiley & Sons.

Wu Y, Yang X, Plaza A, Qiao F, Gao L, Zhang B & Cui Y 2016. Approximate computing of remotely sensed data: SVM hyperspectral image classification as a case study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9, 12: 5806-5818.

Xia J, Du P, He X & Chanussot J 2014. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters* 11, 1: 239-243.

Xia Y, Liu C, Li YY & Liu N 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications* 78: 225-241.

Xu L, Li J & Brenning A 2014. A comparative study of different classification techniques for marine oil spill identification using radarsat-1 imagery. *Remote Sensing of Environment* 141: 14-23.

Xu Q-S & Liang Y-Z 2001. Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems* 56: 1-11.

Xue J & Su B 2017. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors* 2017.

Yelenik SG, Stock WD & Richardson DM 2004. Ecosystem level impacts of invasive *Acacia saligna* in the South African fynbos. *Restoration Ecology* 12, 1: 44-51.

Zarco-Tejada PJ, González-Dugo V & Berni JAJ 2012. Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sensing of Environment* 117: 322-337.

Zarco-Tejada PJ, González-Dugo V, Williams LE, Suárez L, Berni JAJ, Goldhamer D & Fereres E 2013. A PRI-based water stress index combining structural and chlorophyll effects: Assessment using diurnal narrow-band airborne imagery and the CWSI thermal index. *Remote Sensing of Environment* 138: 38-50.

Zhang T, Su J, Liu C & Chen W 2017. *Band selection in sentinel-2 satellite for agriculture applications band selection in sentinel-2 satellite for agriculture applications*. Proceedings of the 23rd International Conference on Automation & Computing. University of Huddersfield, Huddersfield. IEEE.: 7-8.

Zhao Z, Morstatter F, Sharma S, Alelyani S, Anand A & Liu H 2010. Advancing feature selection research. *ASU Feature Selection Repository*: 1-28.

Zygielbaum AI, Gitelson AA, Arkebauer TJ & Rundquist DC 2009. Non-destructive detection of water stress and estimation of relative water content in maize. *Geophysical Research Letters* 36, 12.