EFFICACY OF MACHINE LEARNING AND LIDAR DATA FOR CROP TYPE MAPPING

By ADRIAAN JACOBUS PRINS

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

Supervisor: Prof A Van Niekerk

December 2019

DECLARATION

By submitting this report 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.

With regards to Chapter 3 and 4 the nature and scope of my contribution were as follows:

Chapter	Nature of contribution	Extent of contribution (%)
Chapter 3	This chapter was submitted as a journal article to International Journal of Applied Geospatial Research and co-authored by my supervisor who helped in the conceptualization and writing of the manuscript. I carried out the literature review, data collection and analysis components and produced the first draft of the manuscript. The manuscript was accepted and is currently in press.	AJ Prins 85% Prof A van Niekerk 15%
Chapter 4	This chapter was published as a journal article in Geo-spatial Information Science. It was co-authored by my supervisor who helped in the conceptualization and writing of the manuscript. I carried out the literature review, data collection and analysis components and produced the first draft of the manuscript.	AJ Prins 85% Prof A van Niekerk 15%

Date: 23 August 2019 (updated on 22 July 2020)

Signature:

SUMMARY

Accurate crop type maps are important for obtaining agricultural statistics such as water use or harvest estimations. The traditional approach to obtaining maps of cultivated fields is by manually digitising the fields from satellite or aerial imagery. However, manual digitising is time-consuming, expensive and subject to human error. Automated remote sensing methods have been a popular alternative for crop type map creation, with machine learning classification algorithms gaining popularity for classifying crop types from satellite imagery. However, using light detection and ranging (LiDAR) data for crop type mapping has not been widely researched. This study assessed the use of LiDAR data for crop type classification, by using it on its own and in combination with Sentinel-2 and aerial imagery.

The first experiment evaluated the use of LiDAR data and machine learning for classifying vineyards. The LiDAR data was obtained from a 2014 survey by the City of Cape Town. The normalised digital surface model (nDSM) and intensity raster derived from the LiDAR data were interpolated at four resolutions (1.5 m, 2 m, 2.5 m and 3 m) and then used for generating a range of texture measures. The textures measures were generated using two window sizes (3x3 and 5x5) per resolution scenario, which resulted in eight datasets. The resulting dataset was then used as input for 11 machine learning classification algorithms, which performed a binary classification of vineyards and non-vineyards. The results showed that LiDAR data are able to discriminate between vineyards and non-vineyards, with the random forest (RF) classifier obtaining the highest overall accuracy (OA) of 80.9%. Furthermore, the results showed that a significant difference in accuracy can be achieved with neural networks and distance-based classifiers when the input data are standardised.

The second experiment used the methods developed for the first experiment to perform a five-class classification. The five classes consisted of maize, cotton, groundnuts, orchards and non-agriculture. Sentinel-2 and aerial imagery data were added to the analysis and were compared to LiDAR data. The LiDAR data was obtain from a 2016 survey of the Vaalharts irrigation scheme. Furthermore, the three datasets (Sentinel-2, aerial imagery and LiDAR data) were combined in order to evaluate which combination of datasets produces the highest OA. The results showed that the performance of LiDAR data was similar to that of Sentinel-2 imagery, with LiDAR data obtaining a mean OA of 84.3%, while Sentinel-2 obtained a mean OA of 83.6%. The difference between the OAs of LiDAR and Sentinel-2 were statistically insignificant. The highest OA (94.6%) was obtained with RF when the LiDAR, Sentinel-2 and aerial datasets were combined. However, a combination of LiDAR data and Sentinel-2 imagery obtained similar results to when

all three datasets were used in combination, with the difference in OA being statistically insignificant.

Generally, LiDAR data are suitable for classifying different crop types, with RF obtaining the highest OAs in both experiments. The combination of multispectral and LiDAR data produced the highest OA.

KEY WORDS

Per-pixel image analysis, LiDAR, Sentinel-2, aerial imagery, machine learning, supervised classification, crop type classification

OPSOMMING

Akkurate digitale gewaskaarte is belangrik vir die verkryging van landboustatistieke soos watergebruiks- of gewasopbrengsberaming. Die tradisionele benadering tot die verkryging van digitale gewaskaarte is om dit met die hand van satelliet- of lugfoto's te versyfer. Hand-versyfering is egter tydrowend, duur en vatbaar vir menslike foute. Outomatiese afstandswaarnemingsmetodes is 'n gewilde alternatief vir die skep van gewaskaarte, met masjienleeralgoritmes wat gewild raak vir die klassifisering van gewasse vanaf satellietbeelde. Die gebruik van slegs ligbespeuring-en-afstandsbepaling (LiBEA)-data vir gewasklassifikasie is egter nog nie wyd ondersoek nie. Hierdie studie het die gebruik van LiBEA-data vir gewasklassifikasie geassesseer deur hierdie data op sy eie, asook in kombinasie met Sentinel-2 beelde en lugfoto's, te gebruik.

Die eerste eksperiment het die gebruik van LiBEA-data en masjienleer vir die klassifikasie van wingerde geëvalueer. Die LiBEA-data is van 'n 2014-opname deur die Stad Kaapstad verkry. Die LiBEA-afgeleide genormaliseerde digitale oppervlakmodel (gDOM) en intensiteitsbeeld is by vier resolusies (1,5 m, 2 m, 2,5 m en 3 m) geïnterpoleer en toe vir tekstuurmetings gebruik. Twee venstergroottes (3x3 en 5x5) per resolusie is vir die generering van die tekstuurmetings gebruik, wat agt datastelle tot gevolg gehad het. Die resulterende datastel is as toevoer vir 11 masjienleer-klassifikasie-algoritmes gebruik, wat 'n binêre klassifikasie van wingerde en nie-wingerde uitgevoer het. Die resultate het getoon dat LiBEA-data tussen wingerde en nie-wingerde kan diskrimineer, met die ewekansige woud (EW) klassifiseerder wat die hoogste algehele akkuraatheid (AA) van 80,9% behaal het. Verder het die resultate getoon dat die standaardisering van die toevoerdata 'n beduidende verbetering aan die resultate van die neurale netwerke en afstandsgebaseerde klassifiseerders te wee gebring het.

Die tweede eksperiment het die metodes wat vir die eerste eksperiment ontwikkel is gebruik om 'n vyfklas-klassifikasie uit te voer. Die vyf klasse het bestaan uit mielies, katoen, grondbone, boorde en nie-landbou. Sentinel-2 en lugfoto-data is ook by die analise gevoeg en is met LiBEA-data vergelyk. Die LiBEA-data is verkry uit 'n 2016-opname van die Vaalharts-besproeiingskema Verder is die drie datastelle (Sentinel-2, lugfoto's en LiBEA-data) gekombineer om te bepaal watter kombinasie van datastelle die hoogste AA tot gevolg het. Die resultate het getoon dat die werksverrigting van LiBEA-data soortgelyk aan dié van Sentinel-2-beelde was, met LiBEA-data wat 'n gemiddelde AA van 84,3% behaal het, terwyl Sentinel-2 'n gemiddelde AA van 83,6% behaal het. Die verskil tussen die AAs van LiDAR en Sentinel-2 was statisties onbeduidend. Die hoogste behaalde AA (94,6%) is verkry deur die EW-klassifiseerders wat van die gekombineerde data van LiBEA, Sentinel-2 en lugfoto's gebruik gemaak het. Met die kombinasie van LiBEA-

data en Sentinel-2 is soortgelyke resultate egter verkry as wanneer al drie datastelle in kombinasie gebruik is, met ombeduidende verskille in AA.

Oor die algemeen was LiBEA-data geskik om verskillende gewastipes te klassifiseer, met EW wat die hoogste AA in beide eksperimente behaal het. Die kombinasie van multispektrale data en LiBEA het die hoogste AA tot gevolg gehad.

TREFWOORDE

Per-piksel-beeldeanalise, LiBEA, Sentinel-2, lugfoto’s, masjienleer, gekontroleerde klassifikasie, gewasklassifikasie

ACKNOWLEDGEMENTS

I sincerely thank:

- My supervisor Prof Adriaan Van Niekerk;
- My family for their support;
- Centre for Geographical Analysis for providing me with LiDAR data;
- Northern Cape Department of Agriculture, Land Reform and Rural development for providing me with LiDAR data and aerial imagery;
- Helene van Niekerk of Linguafix (www.linguafix.net) for language editing;
- The National Research Foundation (grant number 112300) for their funding.

This work forms part of a larger project titled “Salt Accumulation and Waterlogging Monitoring System (SAWMS) Development” which was initiated and funded by the Water Research Commission (WRC) of South Africa (contract number K5/2558//4). More information about this project is available in the WRC Report TT 782/18 (ISBN 978-0-6392-0084-2) available at www.wrc.org.za.

CONTENTS

TABLES

FIGURES

APPENDICES

ACRONYMS AND ABBREVIATIONS

ALSM	Airborne laser swath mapping
ANOVA	Analysis of variance
AUC	Area under the curve
CART	Classification and regression tree
CHM	Canopy height model
DEM	Digital elevation model
d-NN	Deep neural network
DSM	Digital surface model
DT	Decision tree
DTM	Digital terrain model
EVI2	Enhanced vegetation index
FMIS	Farm management information systems
FP	False positives
GDP	Gross domestic product
GEOBIA	Geographic object-based image analysis
GIS	Geographic information systems
GLCM	Grey-level co-occurrence matrix
GSD	Ground sampling distance
HISTEX	Histogram-based texture measures
IDW	Inverse distance weighted
IMU	Inertial measurement unit
K-NN	K-nearest neighbour
LiDAR	Light detection and ranging
LR	Logistic regression
MNF	Minimum noise fraction transformed
NB	Naïve bayes
nDSM	Normalised digital surface model
NDVI	Normalized difference vegetation index
NDWI	Normalized difference water index
NIR	Near-infrared
NN	Neural network
OA	Overall accuracy
OLI	Operational land imager
OOB	Out-of-bag
PCA	Principal component analysis
RBF	Radial basis function
RF	Random forest
RGB	Red, green and blue

ROC	Receiver operating characteristic
SAR	Synthetic aperture radar
SD	Standard deviation
SVM	Support vector machine
SVM RBF	Svm with a radial basis function kernel
SVM-L	Svm with a linear kernel
TEX	Texture analysis
TIRS	Thermal infrared sensor
UAV	Unmanned aerial vehicle
UHR	Ultra-high resolution
VHR	Very high resolution
VIP	Vegetation interface process
XGBoost	Extreme gradient boosting

CHAPTER 1: INTRODUCTION

The growing world population, along with variations in annual crop yields, has caused short-term fluctuations in international food prices, as well as a long-term increase in global food demand (Sakamoto, Gitelson & Arkebauer 2014). An increase in agriculture is needed to ensure food security, which will require advances in technology to optimally manage the land and water usage (Yalcin & Günay 2016). Advances in agricultural technology have led to a 12% increase in agricultural fields globally, and production has almost doubled in the last fifty years, mainly due to the use of fertiliser, more efficient cultivars and increased water productivity (Acevedo 2017; Yalcin & Günay 2016). In South Africa, agriculture contributes to about 12% of gross domestic product (GDP). Twelve per cent of South Africa's surface area is suitable for crop production, and only 2.6% is considered high potential arable land. Only around 1 330 000 ha is under irrigation (Van Niekerk et al. 2018). Clearly, South Africa's limited agricultural resources should be managed as efficiently as possible to ensure food security.

1.1 IMPORTANCE OF AGRICULTURAL DATABASES

Agriculture relies on timely information for decision-making (Fountas et al. 2015). Agricultural expert systems have evolved from simple record keeping to large, comprehensive farm management information systems (FMISs) used for crop prediction estimates, crop disease diagnostics, farm planning and irrigation monitoring (Doluschitz & Schmisseur 1988; Fountas et al. 2015). FMISs generally have eleven generic functions, namely field operations management, best practice (including yield estimation), finance, inventory, traceability, reporting, sales, machinery management, human resource management and quality assurance (Fountas et al. 2015).

1.2 REMOTE SENSING FOR AGRICULTURE

1.2.1 Remote sensing

Remote sensing is the use of data acquired from a distance to derive information about the earth's land and water surfaces. Remotely sensed data are acquired with the use of electromagnetic radiation in one or more regions of the electromagnetic spectrum. Radiation is either reflected or emitted from the earth's surface.

Remote sensing sensors can be either passive or active. Passive sensors are used to record solar radiation after it has been reflected from objects of interest and typically function within the visible and near-infrared spectrums. Optical sensors, such as those mounted on the Landsat and SPOT satellites, are good examples of passive sensors. Unlike passive sensors, active sensors generate

their own radiation, which is then projected towards a target object. The portion of the radiation that is reflected off the target object is recorded. Active sensors do thus not rely on solar radiation. Typical active sensors include light detection and ranging (LiDAR) and synthetic aperture radar (SAR) (Campbell & Wynne 2011).

1.2.2 Applications of remote sensing in agriculture

1.2.2.1 Precision agriculture

Precision agriculture uses a range of technologies and data obtained from multiple sources to support decisions about land management, water use and planting systems (Ishida et al. 2004; Turker & Kok 2013). It involves the collection and analysis of spatial-temporal information obtained through remote sensing, in situ yield monitoring, satellite-based field positioning and computer processing (Mulla 2013). The application of remote sensing for agriculture requires information on crop yield, crop canopy, biomass, weed infestation, crop nutrient and water stress, as well as soil properties, nutrients, pH and salinity (Lee et al. 2010; Mulla 2013). Crop yield is possibly the most important variable in crop management as it integrates the effects of various spatial factors such as soil properties, topography, crop nutrients, plant population and irrigation (Lee et al. 2010). Unlike in situ crop yield data that can only be collected during harvest, remote sensing methods can be used to estimate yields during the growing season in near real-time, which allows for dynamic crop management (Lee et al. 2010).

1.2.2.2 Crop yield estimations

With a growing population, improved food security is required to ensure an adequate food supply. Accurate crop yield estimations are important in establishing food security but is often difficult to establish (You et al. 2017). Traditional methods of crop yield estimation and forecasting is to sample ground plots during harvest or to use regression models using rainfall and past yield data as input to model expected yields. However, these approaches are time-consuming and labour intensive and do not truly capture the spatial variability of crops (Liaqat et al. 2017).

Remote sensing is widely used for crop yield estimations owing to its ability to produce timely and accurate spatial data on crop growth. Many developed countries have well-established methods of crop yield forecasting that use geographic information systems (GIS) and satellite imagery, with the normalized difference vegetation index (NDVI), soil properties, two-band enhanced vegetation index (EVI2) and normalized difference water index (NDWI) commonly being used as input data (Liaqat et al. 2017; You et al. 2017). The data are then used as input for regression models or machine learning algorithms (Bolton & Friedl 2013; Pandey & Mishra 2017; Whetton et al. 2017; You et al. 2017).

1.3 CROP TYPE MAPS

Crop type maps are fundamental datasets in agricultural analysis (agricultural statistics, soil erosion models, yield-prediction models, water use estimation and precision agriculture) as they represent minimum mapping units for extracting statistics per production unit (e.g. field, orchards and vineyard). Obtaining accurate agricultural statistics requires up-to-date crop type geodatabases (Gilbertson, Kemp & Van Niekerk 2017). The traditional approach to producing crop type maps is to manually digitise fields from aerial or satellite imagery and then assigning crop type to each field from crop information collected from aerial and ground surveys. This can, however, be very time-consuming, labour intensive and costly (Yalcin & Günay 2016).

Once crop type maps have been created, they can be incorporated in different analyses (Turker & Kok 2013) including agricultural statistics, soil erosion models, yield-prediction models, water use estimation and precision agriculture (Ghariani et al. 2014; Mo et al. 2009; Pedroso et al. 2010; Rydberg & Borgefors 2001; Senturk, Bagis & Berk Ustundag 2014; Turker & Kok 2013).

Semi-automated and automated methods for crop type map creation have been employed on remotely sensed imagery using different techniques, including per-pixel analysis (Bargiel 2017; Immitzer, Vuolo & Atzberger 2016; Tatsumi et al. 2015; Wu et al. 2017), object-based image analysis (Belgiu & Csillik 2018; Gilbertson, Kemp & Van Niekerk 2017; Li et al. 2015; Liu & Bo 2015; Peña-Barragán et al. 2011), unsupervised classification (Hoekman, Vissers & Tran 2011; Mathews & Jensen 2012; De Rainville et al. 2014) and supervised classification (Gilbertson & Van Niekerk 2017; Kussul et al. 2017; Valero et al. 2016; Vuolo et al. 2018). Per-pixel analysis and object-based image analysis determine if the pixel values are evaluated individually or in groups (objects), while unsupervised and supervised classification are approaches to evaluating the pixel or object values.

Segmentation is a geographic object-based image analysis (GEOBIA) method that generates image objects (contiguous groups of pixels) by assessing the spatial, spectral and temporal characteristics of images (Arvor et al. 2013; Pedroso et al. 2010). GEOBIA is an alternative to per-pixel image processing and tends to be more reliable along feature boundaries where mixed pixels often occur (Evans et al. 2002; Turker & Ozdarici 2011; Yalcin & Günay 2016). Spectral statistics (e.g. mean, median, minimum, maximum, variance and range) can also be derived from the objects and used in image analysis as additional variables (Blaschke 2010). GEOBIA has been shown to result in higher classification accuracies compared to traditional per-pixel image analyses (Arvor et al. 2013; Blaschke et al. 2014; Gilbertson, Kemp & Van Niekerk 2017).

Despite the recent advances made in crop classification, a number of challenges still exist. The main limitations of GEOBIA approaches are over- and under-estimation of the boundary location

(Liu & Bo 2015; Pedroso et al. 2010; Turker & Kok 2013; Yalcin & Günay 2016). One approach to reducing over- and under-estimation of the boundary location is to make use of higher resolution imagery as this reduces mixed pixels, resulting in boundaries between crops to become more clearly visible (Turker & Kok 2013). However, even with high resolution imagery, the small spectral differences among certain crops negatively affect crop classification accuracies (Pedroso et al. 2010; Turker & Kok 2013). This reduction in overall accuracy caused by the over- and under-segmentation could outweigh the performance increase GEOBIA provides over per-pixel image analysis (Gilbertson, Kemp & Van Niekerk 2017). Data fusion – the combination of different data sources – has been proposed as a possible solution to improve crop classification (Zhang 2010). For instance, it has been shown that by combining LiDAR data with optical imagery, crop classification accuracies were improved (Liu & Bo 2015; Mathews & Jensen 2012).

1.4 MACHINE LEARNING

Machine learning is a highly versatile classification method and has been used for a range of applications in remote sensing, with land cover classification being one of the most popular (Al-doski et al. 2013; Lary et al. 2016; Loggenberg et al. 2018; Qian et al. 2015). Machine learning algorithms provide automated ways for classifying datasets and can be grouped into two main types, namely supervised and unsupervised (Eastman 2006; Möller et al. 2016). In supervised machine learning, the user provides training data (known labels) from which the algorithm develops a statistical characterisation of each class. Once created, the characterisation is employed to label unknown data (Al-doski et al. 2013; Eastman 2006). The algorithm classifies the dataset by assigning every unknown record to a class (statistical characterisation) that it resembles the most (Eastman 2006). Unsupervised machine learning requires no training data and splits the input data into the most prevalent spectral clusters or classes (Eastman 2006). The analyst is then required to assign an informational label to each cluster (Eastman 2006; Kotsiantis 2007). Machine learning algorithms can be further categorised into sub-groups, namely: parametric and non-parametric algorithms; or hard and soft (fuzzy) classifiers (Al-doski et al. 2013). Non-parametric machine learning algorithms have gained popularity in remote sensing as they can classify different types of data, have the capacity to deal with non-normal distributed data (which is often the case in remotely sensed data) and are robust in conditions of high dimensionality. Popular non-parametric algorithms include decision tree (DT), neural network (NN), random forest (RF), k-nearest neighbour (K-NN) and support vector machine (SVM) (Al-doski et al. 2013; Gilbertson, Kemp & Van Niekerk 2017).

1.5 LiDAR

LiDAR is an active sensor that transmits and receives energy pulses in a narrow range of frequencies. The brightness, angular position, change in frequency and the timing of the reflected pulses can be analysed to describe the structure of the terrain and vegetation feature-information not obtained from conventional optical sensors. LiDAR provides detailed spatial data of high accuracy and precision with horizontal accuracy in the range of 20–30 cm and vertical accuracies in the range of 15–20 cm (Campbell & Wynne 2011). The main advantages of LiDAR data over optical imagery is its neutrality to relief displacements, penetrative capability through vegetation canopies and insensitivity to lighting conditions (Yan, Shaker & El-Ashmawy 2015).

LiDAR height data have been successfully used to separate crops with similar spectral characteristics (Liu & Bo 2015). Antonarakis, Richards & Brasington (2008) showed that land cover classification and field boundary delineation can be performed using LiDAR data on its own by performing segmentation on LiDAR derivatives. LiDAR data can be used to create digital surface models (DSMs), digital terrain models (DTMs) and a normalised DSMs (nDSMs) that can be used to measure vertical structural information of vegetation, which is invaluable in land cover classifications (Bietresato et al. 2016; Liu & Bo 2015). nDSM and z-deviation (variability in the point cloud) values remain relatively stable across heterogeneous landscapes and could potentially be used for accurate large-scale analysis (O'Neil-Dunne et al. 2012; Zhou 2013). Multi-return LiDAR has also been used to provide physical measures of vegetation structure (McCarley et al. 2017), while the intensity of returns is often used to discriminate between non-metallic or biological objects (Bietresato et al. 2016; Zhou 2013), with vegetation having the highest reflectance and water having the lowest reflectance (Antonarakis, Richards & Brasington 2008).

Mathews & Jensen (2012) successfully used a generalised LiDAR-derived nDSM to differentiate between vineyards and other land covers, with the principle aim of delineating vineyard boundaries. They tested different window sizes for the generalisation (focal statistics) of the nDSM and applied an unsupervised classifier in three areas covering 7.8 km^2 of vineyards. The LiDAR data, which had a point cloud density of 0.33 points/m^2 and an average point spacing of 1.74 m, were used to derive a 0.6 m nDSM. The nDSM that was generalised by using focal statistics with a window size of 12x12 pixels was found to most suited for filling the gaps between the vineyard rows. This nDSM was used as input for an ISODATA classification algorithm for generating six clusters, which were then manually assigned to vineyard and non-vineyard classes. The study obtained overall accuracy (OA) ranging from 97% to 98.2% and showed that LiDAR data hold potential for vineyard mapping at a local scale. However, it would be worth extending this work

by investigating whether supervised machine learning approaches can achieve similar or better results and whether LiDAR data can be used to map vineyards over large areas.

Other examples in which LiDAR derivatives were used for land cover classification includes Brennan & Webster (2006), O'Neil-Dunne et al. (2012) and Zhou (2013), who achieved overall accuracies exceeding 90% in some experiments. The latter two studies derived image texture from the LiDAR derivatives, while Brennan & Webster (2006) indicated that texture might help with one of their studies limitations, as the texture can provide supplementary information relating to the land cover patterns and thus be useful for discriminating between heterogeneous crop fields (Peña-Barragán et al. 2011). Studies that combined LiDAR with optical data (Bork & Su 2007; Bujan et al. 2012; Chen et al. 2009; Geerling et al. 2007; Liu & Bo 2015; Sasaki et al. 2012) showed an overall increase in classification accuracies.

1.6 PROBLEM STATEMENT

Crop type maps are valuable assets that can be used for regional crop analyses such as crop yield forecasting and water use estimation (Hämmerle & Höfle 2014). Crop type maps have traditionally been created by visual interpretation of aerial or satellite imagery, manual digitising and field surveys. This process is labour intensive, time-consuming and costly and subject to human error and bias (Yalcin & Günay 2016). The dynamic nature of cultivation requires crop type maps to be updated every season (Gilbertson, Kemp & Van Niekerk 2017), but this would be prohibitively expensive using traditional methods, especially at regional (national) scales. The only viable solution is to produce crop type maps automatically using remote sensing techniques. Although crop type mapping has been carried out with some success using optical Landsat (Gilbertson, Kemp & Van Niekerk 2017; Sonobe, Tani & Wang 2017), SPOT (Waldhoff, Lussem & Bareth 2017; Yang et al. 2013), GeoEye (Etoughe Kongo 2015), IKONOS (Bannari et al. 2006; Turker & Ozdarici 2011) and aerial imagery (Fiorillo et al. 2012; Yalcin & Günay 2016), several challenges need to be solved before this approach will become operational. For instance, it is challenging to classify perennial crops (e.g. vines and fruit trees) owing to the similar spatial patterns (e.g. rows) in which they are planted and canopy structure (Mathews & Jensen 2012; Peña-Barragán et al. 2011). The principle challenge is the mixed-pixel effect (Chen et al. 2018) that results from inter-row bare soil, shadow, cover crops and weeds (Hall, Louis & Lamb 2003; Liu & Bo 2015; Mathews & Jensen 2012), particularly when imagery with resolutions lower than the row spacing is used. Another challenge is that reflection values and shadows in optical imagery can vary substantially across heterogeneous landscapes due to varying lighting and atmospheric conditions (O'Neil-Dunne et al. 2012). Given that LiDAR is an active sensor and thus unaffected by lighting or weather conditions, LiDAR-derived nDSM and z-deviation (variability in the point

cloud) values are relatively stable and have the added benefit of not including effects of relief displacement (O'Neil-Dunne et al. 2012; Yan, Shaker & El-Ashmawy 2015).

LiDAR data are becoming increasingly available and have successfully been used for land cover classification in which either LiDAR data alone (Antonarakis, Richards & Brasington 2008; Brennan & Webster 2006; Mathews & Jensen 2012; Zhou 2013) or a combination of LiDAR data and optical imagery was used (Bujan et al. 2012; Chen et al. 2009; Liu & Bo 2015; O'Neil-Dunne et al. 2012). However, apart from Mathews & Jensen (2012) who used LiDAR data for delineating vineyards, no published research comparing classification techniques on LiDAR data (and its derivatives) for crop type classification are available. The pioneering work of Mathews & Jensen (2012) can be extended by also incorporating intensity data in the classifications. Furthermore, the effect of image textures and the efficacy of different machine learning classifiers should be investigated. It is also not clear how LiDAR data compare to multispectral aerial and satellite imagery. These gaps in the current knowledge can be addressed by answering the following research questions:

1. What spatial resolution of LiDAR derivatives is most suited for classifying crops?
2. Which LiDAR derivatives are most effective for differentiating among crop types?
3. Which machine learning algorithm are most effective for differentiating among crop types?
4. Compared to multispectral imagery, how effective is LiDAR data for crop type mapping at regional scales?

1.7 AIM AND OBJECTIVES

The aim of this study is to develop and assess a method whereby crops can be automatically classified using LiDAR data and machine learning. LiDAR data – on its own and in combination with multispectral imagery – are used as input to various machine learning algorithms. The resulting maps are quantitatively and qualitatively compared to evaluate the value of LiDAR data for crop type mapping at regional scales.

The following objectives have been set:

1. Quantify the effect of different spatial resolutions and window sizes for generating LiDAR derivatives used as input to machine learning algorithms.
2. Determine the most successful classification methods for discriminating between crop types when LiDAR derivatives are used as predictor variables.
3. Assess the value of using LiDAR data in combination with Sentinel-2 and aerial imagery.

4. Critically evaluate the value of LiDAR data for operational mapping of crop types at regional scales.

1.8 RESEARCH METHODOLOGY

The research is empirical in nature as it involves observations and experimentation on datasets obtained from agricultural databases (of in situ observations) and remotely sensed data. The methods that were assessed included several machine learning methods (supervised classification), the results of which were compared to statistical techniques (regression analysis), and is as such quantitative in nature. Furthermore, the results were assessed using qualitative methods by visually comparing the results.

Two experiments were carried out in this study. The first experiment (Chapter 3) involved the classification and delineation of vineyards and thus contributed towards research objectives three and four. Quantifying the effect of the different spatial resolutions and window sizes (Objective 3) is presented in Chapter 3, while Chapter 3 and 4 outline how the most successful classification method (Objective 4) was determined. The experiment used empirical vineyard field observations (for model building and assessment) and LiDAR derivatives (as predictor variables). The City of Cape Town was used as the study area due to the availability of LiDAR data. The data was quantitatively analysed using regression analysis and machine learning algorithms.

The second experiment (Chapter 4) made use of the results of the first experiment and introduced optical remotely sensed imagery (aerial and Sentinel-2 satellite images) to assess how it compares to the LiDAR-based classification accuracies (Objective 5). The Vaalharts irrigation scheme was used in this experiment, due to the availability of LiDAR data. Unlike the first experiment, this experiment was not performed on vineyards, but rather on three annual crop types, namely cotton, maize and groundnuts. Tree-based perennial crops were grouped into a fourth category, namely orchards.

Qualitative methods (e.g. visual interpretation of imagery and thematic maps) were used in both experiments to assess the overall ability of the methods evaluated.

Figure 1.1 illustrates the research design and thesis structure. This chapter (Chapter 1) introduced the background, problem statement and aims and objectives of the study. Chapter 2 provides an overview of remote sensing techniques considered in this research, followed by a review of previous studies on crop type classification. Chapter 2 concludes with a motivation for the methods used in this research.

Chapter 3 and 4 present the two experiments described above, while Chapter 5 reflects on the research questions and aims and objectives, discusses the key findings of the research and makes

recommendations for operational crop type classifications using LiDAR data and optical imagery. The evaluation of LiDAR data for operational mapping of crop types at regional scale (Objective 6) is presented in Chapter 5.

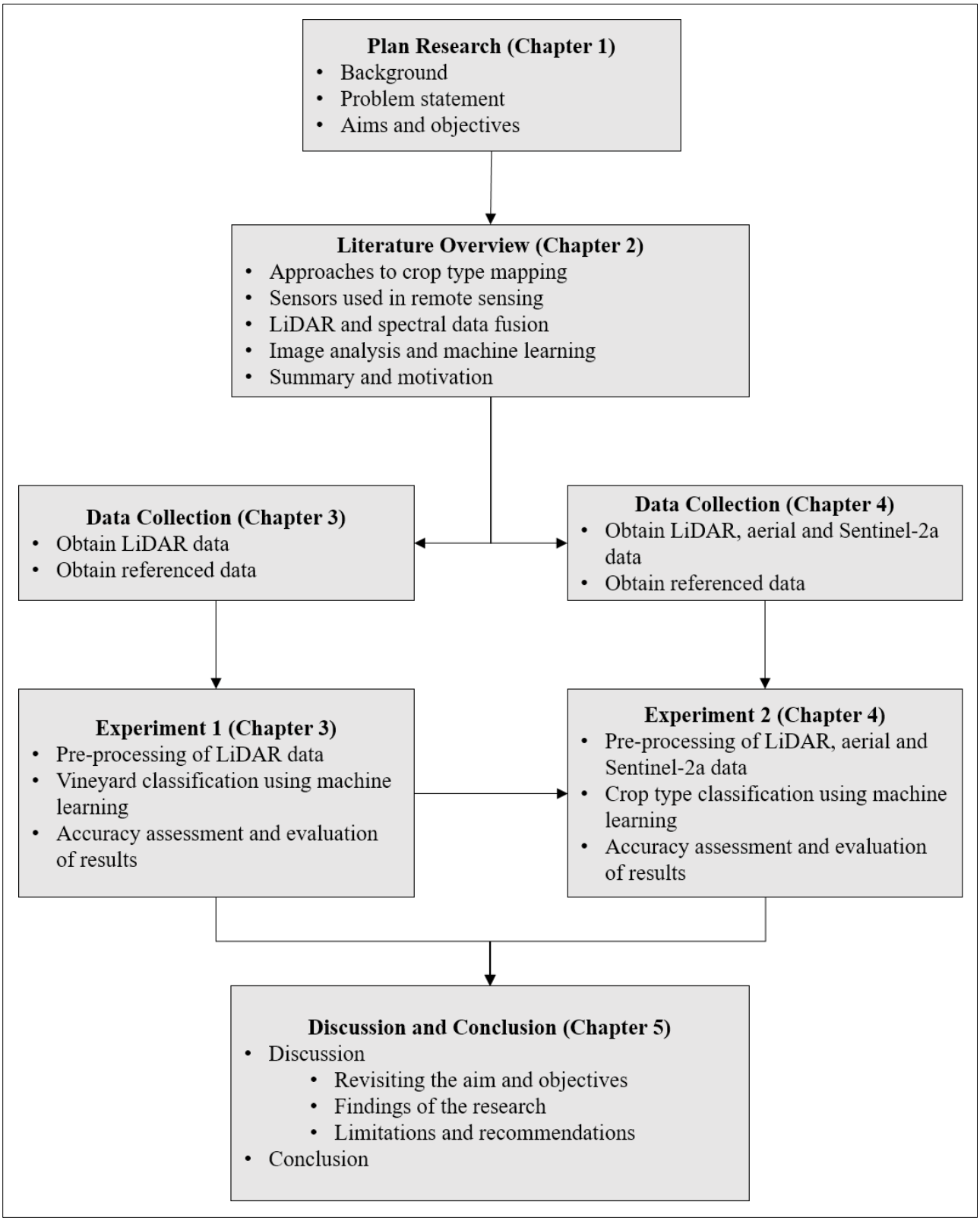


Figure 1.1: Research design

CHAPTER 2: LITERATURE OVERVIEW

This chapter provides an overview of methods and data commonly used for crop type mapping. A background of the sensors used for this purpose is also given, followed by a review of data fusion, image analysis and machine learning classification methods. The chapter concludes with a summary of the main findings, a motivation for the methods used in this study and a brief overview of the chapters to follow.

2.1 EARTH OBSERVATION DATA USED IN AGRICULTURE

The earth observation data that is available usually determine the type of applications it is used for, particularly in agriculture (Mulla 2013). Remotely sensed data can be categorised according to the type of platform (e.g. ground, airborne or satellite), type of sensor (i.e. passive or active), region of the electromagnetic spectrum (e.g. visible, infrared and microwave), spectral resolution (e.g. panchromatic, multispectral or hyperspectral), radiometric resolution (e.g. 8, 12 or 16 bits), temporal resolution (e.g. low or high revisit time) and spatial resolution (e.g. low, medium, high or very high) (Khanal, Fulton & Shearer 2017). The earth observation data most commonly used for remote sensing applications in agriculture primarily covers the visible, near-infrared and shortwave-infrared electromagnetic spectrum with sensors mounted on either a satellite, aircraft or unmanned aerial vehicle (UAV) (Khanal, Fulton & Shearer 2017).

Remote sensing sensors can be categorised into two types, namely active and passive. Passive sensors measure radiation reflected or emitted from the earth's surface, for instance measuring solar radiation reflected by the earth. Passive sensors require an external source of radiation as they do not produce their own. A typical example of a passive remote sensing sensor is a multispectral satellite sensor, i.e. Landsat 8, SPOT5 or Sentinel-2. Active sensors produce their own energy source and are not dependent on solar and terrestrial radiation. SAR or LiDAR are forms of active sensors as they transmit radiation towards the earth's surface and then measure the reflected radiation.

2.1.1.1 Aerial imagery

Aerial imagery can be collected by sensors mounted on different platforms, such as aircraft, UAV, blimps or parachutes (Matese et al. 2015; Sankaran et al. 2015). The sensors are capable of collecting very high resolution (VHR) imagery; however, they are limited by their respective flight times for remotely sensing local scale areas. The sensors mounted on aircrafts are used for sensing larger areas, while sensors on UAV are used for smaller areas. The process of collecting aerial

imagery is more flexible than that of collecting satellite imagery; however, the imagery survey can be expensive (Matese et al. 2015).

2.1.1.2 Spaceborne multispectral sensors

Satellites such as Landsat 8, SPOT5, Quickbird, WorldView and Sentinel-2 collect satellite imagery that maps large areas at the same time but at a coarser resolution and at more fixed times compared to aerial imagery, which collects VHR imagery for smaller areas.

The Landsat programme consists of multiple satellites that have been capturing multispectral imagery for over 40 years, making this continuously acquired data the space-based moderate resolution remote sensing data that has been collected for longer than any other data. The first Landsat satellite was launched on 23 July 1972, while the newest Landsat satellite, Landsat 8, was launched on 11 February 2013. Landsat 8 carries two sensors – an operational land imager (OLI) and a thermal infrared sensor (TIRS) – with the OLI containing nine spectral bands that all have a resolution of 30 m, with the exception of the panchromatic band that has a resolution of 15 m. The Landsat 7 satellite is still active and has the same spatial resolution as Landsat 8, but it only has eight bands. Both Landsat 7 and 8 have a revisit time of 16 days.

The SPOT satellite system is a commercial earth observation satellite system that has been providing high resolution imagery since the launch of the SPOT-1 satellite on 22 January 1986. Since then, several SPOT satellites have been launched and decommissioned, with the SPOT-7 (the newest in the system) launched on 30 June 2014. Currently, only two identical satellites are still active, namely SPOT-6 and 7. These satellites carry a five-band multispectral sensor. The blue, green, red and near-infrared bands have a resolution of 6 m, while the panchromatic band has a resolution of 1.5 m. When both are used, the SPOT-6 and 7 provide a daily revisit time.

Quickbird-2 is a commercial imaging satellite from DigitalGlobe Inc. that was launched on 18 October 2001. The Quickbird-2 satellite captures VHR imagery with a four-band multispectral sensor with a resolution of 2.4 m and a panchromatic sensor with a resolution of 0.61 m. The satellite revisit time can vary from 1 up to 3.5 days, depending on latitude. Quickbird-2 is no longer active and entered the earth's atmosphere on 27 January 2015.

The WorldView satellites, the successors to Quickbird-2, are commercial imaging satellites from DigitalGlobe Inc. WorldView consists of four satellites, namely, WorldView-1, 2, 3 and 4, all of which capture VHR imagery with resolutions less than 2 m for the multispectral bands and resolutions equal to or less than 0.5 m for the panchromatic band. WorldView-1 is the only WorldView satellite that only contains a panchromatic sensor and not a multispectral sensor.

WorldView-2 contains an eight-band multispectral sensor along with a panchromatic band. Worldview-3 contains an eight-band multispectral sensor, an eight-band shortwave-infrared sensor and a panchromatic sensor, while WorldView-4 comprises a four-band multispectral sensor and a panchromatic sensor. WorldView-1 has the longest revisit time of 1.7 days, while WorldView-3 and 4 have revisit times of less than a day.

The Sentinel-2 multispectral satellites are part of the European Copernicus program and consist of two near-identical satellites, namely Sentinel-2a and 2b. Sentinel-2a was launched on 23 June 2015 and Sentinel-2b on 7 March 2017 and both satellites have a polar, sun-synchronous orbit at an altitude of 786 km. The dual-satellite constellation has a five-day revisit time. The two Sentinel-2 satellites carry a 13-band multispectral sensor with a swath width of 290 km and resolutions of 10 m, 20 m and 60 m (depending on the band), see Table 2.1 below.

Table 2.1: Sentinel-2a and 2b band central wavelength, bandwidth and resolution

BAND NUMBER	SENTINEL-2A		SENTINEL-2B		
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Resolution (m)
1	443.9	27	442.3	45	60
2	496.6	98	492.1	98	10
3	560	45	559	46	10
4	664.5	38	665	39	10
5	703.9	19	703.8	20	20
6	740.2	18	739.1	18	20
7	782.5	28	779.7	28	20
8	835.1	145	833	133	10
8A	864.8	33	864	32	20
9	945	26	943.2	27	60
10	1373.5	75	1376.9	76	60
11	1613.7	143	1610.4	141	20
12	2202.4	242	2185.7	238	20

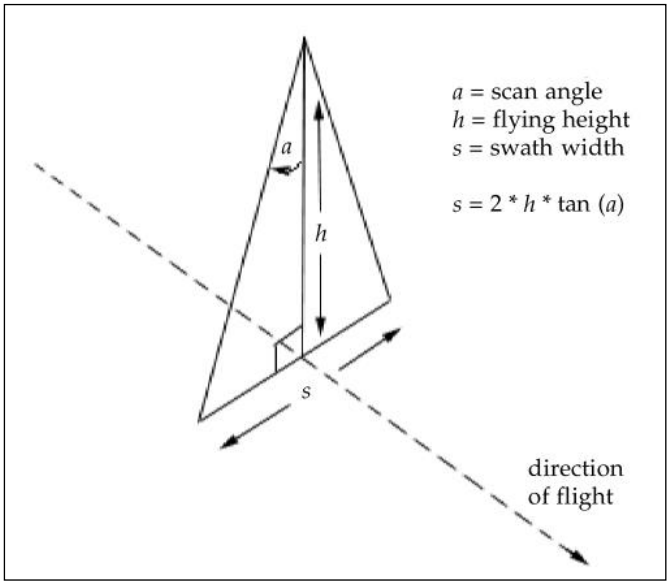
2.1.1.3 Active sensors

SAR belongs to the category of active microwave sensors that transmit electromagnetic radiation with wavelengths of 1 mm to 1 m and then receive portions of the backscatter reflected off the earth's surface (Campbell & Wynne 2011). The reflected backscatter is received by the sensor and is used as the basis for forming the images. SAR has several spaceborne sensors in operation that can provide high resolution imagery that is not affected by daylight, cloud coverage and weather conditions, unlike optical sensors (Cheney 2001). However, SAR is sensitive to surface properties such as soil moisture, small-scale surface roughness and slope. Spaceborne SAR sensors commonly operate at L-, C- and X-band wavelengths, with the smallest wavelength being that of the X-band and the largest belonging to the L-band. Furthermore, SAR sensors can transmit and

receive wavelengths at two polarisations, namely horizontally polarised or vertically polarised. The different wavelengths and polarisations of the sensor determine how the transmitted energy interacts with the earth's surface and can be used for mapping different features of interest (Campbell & Wynne 2011). One of the most common uses of SAR imagery is to create a digital elevation model (DEM); however, SAR does not provide discrete returns (like LiDAR) that can be used to create detailed digital surface models (DSM) and digital terrain models (DTM).

LiDAR, also known as airborne laser swath mapping (ALSM), is an active remote sensing sensor that transmits and receives a narrow range of the electromagnetic spectrum, which is scattered back to the sensor by objects on the earth's surface. Depending on the application, LiDAR transmits and receives electromagnetic radiation in the ultraviolet, visible or infrared region (Longley et al. 2005). Ultraviolet LiDAR systems are used to monitor the earth's atmosphere, visible LiDAR systems are used for bathometry (as green light can penetrate water bodies) and infrared LiDAR systems are used to map the earth's surface (infrared is sensitive to vegetation and is free from atmospheric scattering) (Campbell & Wynne 2011).

The designs of airborne LiDAR systems vary but mainly consist of four components: the aircraft to which the sensor is mounted, a differential GPS for precise geolocation, an inertial measurement unit (IMU) for precise orientation measurements of the airplane and the laser scanner (Lim et al. 2003). The laser scanner can transmit up to 300 000 laser pulses per second (depending on the sensor), which are directed back and forth across the scanning swath by a rotating scanning mirror. The scan angle and flying height (Figure 2.1) determine the swath width (Lim et al. 2003). The receiver on the laser scanner records the emitted pulses (after they were reflected off a surface) and measures the time delay between the emitted and received pulses. Since light travels at a constant speed, the measured time delay directly translates to the distance between the sensor and the object that reflected the pulse (Campbell & Wynne 2011).

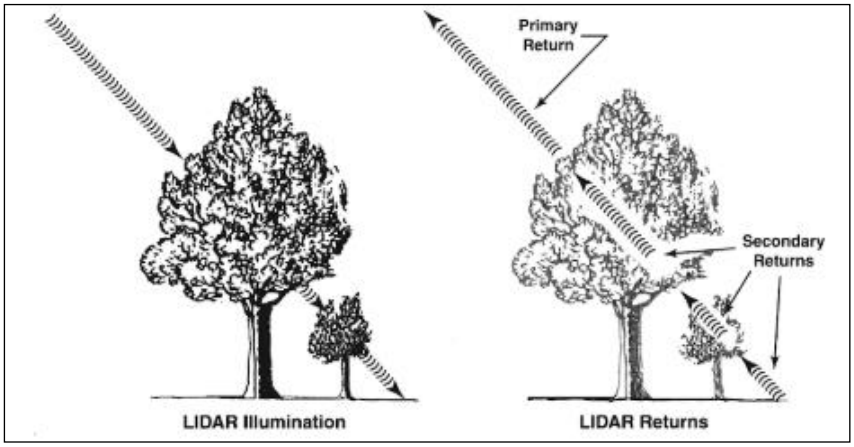


Source: Lim et al. (2003)

Figure 2.1: LiDAR sensor swath width as determined by the scan angle and flying height

LiDAR sensors can be categorised as either discrete-return or waveform LiDAR. Discrete-return LiDAR records four or five returns from each pulse, and for each return it records the time and intensity of the return pulse. Small-footprint LiDAR is typically discrete-return. Waveform LiDAR records the amount of energy returned at a series of equal time intervals, thus resulting in an amplitude-against-time waveform. Waveform LiDAR also gives more information about vertical distribution of vegetation canopy, whereas discrete-return LiDAR only provides a portion of the actual vegetation canopy (Figure 2.2) (Campbell & Wynne 2011; Lim et al. 2003).

LiDAR data have been used for various environmental applications such as monitoring coastal changes, mapping geological faults under forest canopies, assessing landslide hazards, monitoring ice sheets, accessing vegetation structures and mapping topography (Popescu 2011). The latter two applications have been the two most prominent uses of LiDAR data. Most environmental applications using LiDAR need topographic information, and LiDAR has proven to provide highly accurate and detailed elevation information. Most often, discrete-return LiDAR is used for deriving topographic information; moreover, a substantial number of the returns are disregarded, mainly those representing vegetation. On the other hand, when assessing vegetation structures, it is the LiDAR returns that represent vegetation that is of great interest (Popescu 2011). This is because the vegetation structure information can be used to calculate biomass and volume of a forest or characterise vegetation structures of wildlife habitat (Lim et al. 2003; Popescu 2011).



Source: Campbell & Wynne (2011)

Figure 2.2: LiDAR return interactions with vegetation. Primary return represents the first return and the secondary returns represent the second, third and last returns

The topographic and vegetation structure information derived from LiDAR data have been used as ancillary data during classifications, thereby enhancing the classification. LiDAR data in particular is commonly used as ancillary data and is derived from active sensors alongside SAR data (Khatami, Mountrakis & Stehman 2016).

2.2 CROP TYPE MAPPING

In agriculture, remotely sensed data are often used to create crop type maps. These maps are in turn used for further analysis, such as crop management and crop yield estimation (Tatsumi et al. 2015). Numerous studies have used the data from both the Landsat and SPOT series satellites (active since 1972 and 1986 respectively) to classify crops and create crop type maps. The Landsat satellites provide medium-to-high resolution data and have been used for crop classification in many studies (Bauer et al. 1979; Gilbertson & Van Niekerk 2017; Niel & Mcvicar 2004; Ortiz, Formaggio & Epiphanio 2010; Peña-Barragán et al. 2011; Sonobe, Tani & Wang 2017; Tatsumi et al. 2015; Ulaby, Li & Shanmugan 1982). The SPOT satellites (used in studies by Conrad et al. 2010; Duro, Franklin & Dubé 2012; Hubert-Moy et al. 2001; Myint et al. 2011; Simonneaux et al. 2010; Waldhoff, Lussem & Bareth 2017; Yang et al. 2013) provide higher spatial resolutions than that of the Landsat programme but has lower spectral resolutions. The Sentinel-2 satellite constellation, launched between 2015 and 2017, provides high spatial and spectral resolution data, which is popular for crop classification (Belgiu & Csillik 2018; Estrada et al. 2017; Immitzer, Vuolo & Atzberger 2016; Vuolo et al. 2018).

Active sensors have also been used for creating crop type maps (used either as additional features for classification or used on its own), with synthetic aperture radar sensors being used more often than LiDAR (Bargiel 2017; Dadhwal et al. 2002; Mcnairn et al. 2009; Mcnairn & Brisco 2004; Melgani & Blanzieri 2008). SAR has also been used as an additional feature in combination with

optical data (Blaes, Vanhalle & Defourny 2005; Ulaby, Li & Shanmugan 1982; Wu et al. 2014). LiDAR has been used in combination with optical imagery to create crop type maps (Antonarakis, Richards & Brasington 2008; Brennan & Webster 2006; Jahan & Awrangjeb 2017; Liu & Bo 2015), but is rarely used on its own for this purpose, with Mathews & Jensen (2012) being the only exception (they used LiDAR to map vineyards).

2.3 DATA FUSION

Data fusion is a technique that combines data from more than one source into one dataset, thereby reducing uncertainty associated with only obtaining data from one sensor (Solberg, Jain & Taxt 1994). This technique is commonly used in remote sensing to combine data that has different spatial, spectral and temporal resolutions. The data used in the fusion can be obtained from sensors mounted on satellites, aircraft and ground platforms (Zhang 2010).

Within remote sensing, data fusion techniques can be categorised into three different levels: pixel/data level, feature level and decision level (Pohl & Van Genderen 1998). Pixel-level data fusion is the combination of data at the lowest processing level (raw data) into a single dataset with one resolution. This technique requires images to be resampled and georeferenced to ensure that the information in the different data sources is not misaligned (Zhang 2010). Feature level data fusion uses objects recognised in the different data sources, obtained through segmentation. The features (extent, shape, texture, etc.) are extracted from the initial data sources and are then combined into one dataset (Pohl & Van Genderen 1998). Decision level data fusion does not combine the data but rather the outcome of different algorithms to create the final output. The outcomes for the different algorithms can be combined in two ways: soft fusion or hard fusion. Soft fusion scores the outputs before combining it for the final output, but when the different outputs are used as decisions, it is considered to be hard fusion (Zhang 2010).

LiDAR data have been used as ancillary data in combination with spectral data (multispectral or hyperspectral) to aid image classification. Fusing LiDAR data with spectral data are a preferred data enhancement method as it generally improves the OA by about 5% to 10% (Khatami, Mountrakis & Stehman 2016). The increase is usually attributed to height values (proved by the LiDAR data), making it easier to differentiate between different land covers with similar spectral signatures (Chen et al. 2009).

Chen et al. (2009) combined LiDAR with Quickbird imagery at a feature level for land cover classification over an urban area. The LiDAR data was used to create an nDSM, which is created by subtracting a DTM from a DSM. The nDSM, along with the Quickbird imagery, was segmented

and then classified using a rule-based classification approach, which resulted in an OA of 89.4%. The authors also performed a classification on the Quickbird data only, i.e. without including the nDSM, and obtained an OA of 69.1%. An increase of 20.3% in OA was thus achieved when LiDAR data were combined with the Quickbird spectral data.

Hartfield, Landau & Van Leeuwen (2011) performed a pixel-level data fusion of LiDAR data and high resolution aerial imagery (1 m resolution). Similar to Chen et al. (2009), Hartfield, Landau & Van Leeuwen (2011) created a nDSM before combining the image with the aerial imagery. The fused dataset was then used for urban land cover classification (a classification and regression tree (CART) algorithm was used for the classification). The classification was performed on the fused LIDAR and aerial imagery and on the aerial imagery alone. When performing the classification on the fused data, an OA of 89.2% was obtained; and when only the aerial imagery was used, an OA of 84% was achieved. The addition of LiDAR data therefore resulted in an increase of 5.2% in OA.

LiDAR data have also been fused with hyperspectral data, with Liu & Bo (2015) fusing LiDAR and hyperspectral data for crop type classification, while Jahan & Awrangjeb (2017) fused this data for land cover classification. Both studies created an nDSM from the LiDAR data and performed texture measures on the nDSM. In addition, both studies performed the classification on different combinations of the LiDAR and hyperspectral data. Liu & Bo (2015) performed the data fusion at feature level as the study used an object-based classification. Their study obtained an increase of 9.2% for OA when the LiDAR data and hyperspectral were used in combination, compared to using the hyperspectral data only. Furthermore, when the texture measures were added, the OA increased by another 2%. Jahan & Awrangjeb (2017) fused the data at pixel level and used two different classification algorithms, namely support vector machines (SVM) and decision tree (DT). The study obtained an OA increase of 7.6% for the SVM classifier and an increase of 3% for the DT classifier. When the study added texture measures to the LiDAR and hyperspectral combination, the SVM classifier obtained a further increase of 0.5%, while the DT classifier obtained a further increase of 1.3%.

LiDAR has been combined with Sentinel-2 data at decision level by Estrada et al. (2017), who first classified crops using the Sentinel-2 data and then classified trees and hedges using a LiDAR point cloud. The two classifications were then combined with a protected sites dataset in order to create a final map used for ecological value assessment.

2.4 IMAGE ANALYSIS

2.4.1 Image transformation

Image transformation comprises methods used for either enhancing or deriving information from the spectral information captured in an image. The image transformation methods are usually performed by a local or neighbourhood raster operation. The local raster operations consist of band combinations (ratios between different bands) and statistical analysis (standardisation, principle component analysis), while the neighbourhood raster operations consist of texture measures or filters. Band combinations are typically used to emphasise variations between specific features, such as the variations between vegetation and non-vegetation. Statistical analysis can be used for reducing the data's dimensionality, whereas texture measure adds new dimensions to the data. Filters are commonly used to remove unwanted information or noise.

2.4.1.1 Neighbourhood transformations

Texture can be described as the spatial variation in grey level or a measurement of the spatial and spectral relationship between neighbouring pixels within an image (Gong et al. 2003; Pacifici, Chini & Emery 2009). The texture measure algorithms can be separated into four categories, namely signal-processing, geometrical, model-based and statistical algorithms (grey-level co-occurrence matrix (GLCM), semi-variance analysis) (Pacifici, Chini & Emery 2009). Signal-processing algorithms comprise the transformation of original images by using a filter and then calculating the energy of the transformed images using Gabor filters, Fourier transformation and wavelet packet transformation. Geometrical methods create textures that are made up of texture primitives; this method is only appropriate for areas with regular periodic texture. Model-based texture measures use mathematical models to generate the texture measure. Statistical methods generate the texture measures by using moving windows that cover every pixel in an image, with GLCM and histogram measures being the most popular texture measures within remote sensing (Dekker 2003; Yue et al. 2013).

GLCM texture measures are based on second-order statistics generated from co-occurrence probabilities, which represent the conditional joint probabilities of all pair-wise combinations of grey levels within the moving window according to two parameters, namely interpixel distance and orientation. The co-occurrence probabilities are then stored in a sparse matrix referred to as GLCM (Clausi 2002). Histogram texture measures are based on histogram statistics within the moving window, with mean, mean Euclidean distance, variance, skew, kurtosis, entropy and energy being the most well-known statistics (Dekker 2003). For both the histogram and GLCM

texture measures, a window size must be selected which should be large enough to capture the repeating feature of interest (Warner & Steinmaus 2005).

Yue et al. (2013) derived texture measures from VHR Quickbird imagery, using a GEOBIA approach for classifying land covers. The analysis classified the spectral data on its own and in combination with the texture measures. The spectral data obtained a mean OA of 81%, while the spectral data in combination with the texture measure obtained a mean OA of 86.5%. The addition of texture measures to the classification resulted in an increase in OA for all the different study areas.

2.4.1.2 Per-pixel transformation

Standardisation is a common practice and sometimes a requirement for many machine learning algorithms. Transforming the data using standardisation can improve the performance of machine learning algorithms such as NN, nearest neighbour and clustering classifiers. The performance is increased because the standardisation prevents features with large ranges to have a greater influence than features with smaller ranges (Shalabi, Shaaban & Kasasbeh 2006). A common standardisation method is the zero mean and unit variance (Equation 2.1) method, which centres the data by removing the mean value (of the dataset) from each feature, and then scaling it by dividing the features by the standard deviation of the dataset. This method of standardisation does not affect the distribution of the data (Pedregosa et al. 2012). Zero mean unit variance standardisation is defined as:

$$x' = \frac{x - \bar{x}}{\sigma} \qquad \text{Equation 2.1}$$

where x is the original value;

$\bar{x}$ is the mean of the feature; and

σ is the standard deviation of the feature.

2.4.1.3 Statistical transformations

Principal component analysis (PCA) is a statistical feature extraction method that identifies the optimum linear combinations (Equation 2.2) of a set of bands (that could possibly be related) from the input image. The linear combinations can account for the variation of pixel values within an image and result in a set of values of linearly uncorrelated variables known as the principal components. The number of principal components are always equal to or less than the number of

input bands. The first principal component contains the largest percentage of variance, while the percentage of variance decreases with every consecutive principal component (Campbell & Wynne 2011).

$$A = C_1X_1 + C_2X_2 + C_3X_3 \qquad \text{Equation 2.2}$$

where X_n pixel values for the different bands; and

C_n coefficients (eigenvectors) applied to respective bands.

Gilbertson & Van Niekerk (2017) investigated the value of dimensionality reduction for crop classification with multi-temporal imagery and machine learning. They showed that PCA was more effective than feature selection when reducing the dimensionality and was able to increase the OA obtained by a SVM classifier.

2.4.2 Per-pixel vs object-based paradigms

In remote sensing, the classification scheme determines the basic units of an image to be used when performing a classification (Tehrany, Pradhan & Jebuv 2014). Generally, the classification scheme can be separated into two categories, namely per-pixel image analysis or geographic object-based image analysis (GEOBIA) (Duro, Franklin & Dubé 2012). Per-pixel image analysis is the traditional approach to classification that uses the spectral information from each pixel in an image as a feature. Since each pixel value is used on its own, no spatial or contextual information is taken into account (Tehrany, Pradhan & Jebuv 2014). However, mixed pixels become more prevalent with the decrease in spatial resolution, which results in decreased spectral variability within the classes. Increasing the spatial resolution also causes the ‘salt-and-pepper effect’ to become more abundant in the classification. This can be minimised with a GEOBIA approach that groups pixels together (called image segmentation) and provide spatial and contextual information (Whiteside, Boggs & Maier 2011). However, with image segmentation the user has to select the parameters that will minimize over- or under-segmentation, as both have a negative effect on the overall performance of the classification. Gilbertson et al. (2017) stated that the negative effects of under- or over-segmentation could outweigh the performance increase GEOBIA provides. Furthermore, Duro, Franklin & Dubé (2012) stated that (based on their results) there appeared to be no advantage of using a GEOBIA or per-pixel image analysis classification approach.

2.4.3 Machine learning algorithms

As explained in Section 1.5, machine learning algorithms are highly versatile and have been used in remote sensing to provide automated methods for classifying data (Al-doski et al. 2013; Möller et al. 2016). Two general types of machine learning methods exist: unsupervised and supervised. The unsupervised method requires no training data input from the user and classifies the data based on the most prevalent spectral clusters, whereas the algorithms of the supervised method use training data provided by the user to create a model, which is then used to classify data (Eastman 2006). Of the sub-categories of machine learning algorithms, non-parametric algorithms have gained popularity as they can deal with non-normal distributed data and are robust under high dimensionality. Some of the more popular non-parametric algorithms include DT, NN, RF, K-NN and SVM (Al-doski et al. 2013; Gilbertson & Van Niekerk 2017).

2.4.3.1 DT

The DT classification algorithm recursively separates a dataset into smaller subdivisions according to defined tests at each branch (node) in the tree (Friedl & Brodley 1997). The DT consists of a start node, a set of internal nodes and a set of end nodes (leaves). The starting node is created using the whole dataset and splits (based on the value of one variable) into internal nodes, each representing a class. Each internal node has only one parent node and each parent node (including the starting node) has two children nodes, but the parent node can have multiple descendent nodes. An internal node only processes a subset of the dataset obtained from its parent node and creates two new internal nodes; however, if a node results in all the data in the subset being classified as a single class, then that node is turned into a leaf (end node) (Rutkowski et al. 2014).

Sasaki et al. (2012) tested the DT algorithm for classifying land covers using aerial imagery on its own and then in combination with LiDAR data, and also tested a GEOBIA and per-pixel image analysis approach. They obtained accuracies of 97.5% (with LiDAR) and 95% (without LiDAR) for the GEOBIA classification and 91.7% (with LiDAR) and 62.6% (without LiDAR) for the per-pixel classification. Li et al. (Li et al. 2015) used DT along with GEOBIA to classify crop types using a high temporal resolution Landsat-MODIS enhanced NDVI time series as input. They showed that DT is capable of obtaining an OA of 90.9% and noted that the classification can be further improved by using more features.

2.4.3.2 RF

The RF classification algorithm is an ensemble classifier that consists of multiple DTs. The DTs are generated on a subset of training samples through replacement (bootstrap aggregation, i.e. bagging). The final classification is an average of all the probable classifications in the different DTs (Möller et al. 2007). The RF algorithm divides the samples into in-bag samples (that usually consist of about two-thirds of the samples) and out-of-the bag samples (remaining samples). The in-bag samples are used to train the RF model, while the out-of-bag samples are used in an internal cross-validation technique for assessing the performance of the trained RF model, resulting in an error-estimate known as the out-of-bag (OOB) error (Belgiu & Drăgut 2016). The user defines the number of DTs that is used to grow the 'forest'. Each DT is generated independently without pruning and splits each node at a user-defined number of features selected randomly, creating DTs that have high variance and low bias. When a new unlabelled dataset is used as input for the model, it is evaluated against all the DTs in the RF ensemble, and each DT allocates a sample to a class. The class with the majority of votes is then assigned to the sample (Belgiu & Drăgut 2016).

Pelletier et al. (2016) used RF and SVM to classify 18 land covers from Landsat 8 and SPOT-4 imagery. RF obtained an average OA 82.2%, while the SVM obtained an average OA of 77.4%. They noted that RF was less sensitive to different input feature sets and had better trade-offs between classification performance and computational times compared to SVM. Inglada et al. (2015) also compared RF and SVM for crop type classification. They classified 12 sites, and for eight of the sites RF obtained higher OAs compared to that of SVM. The difference was statistically significant. Matikainen et al. (2017) classified multispectral LiDAR using RF. They classified six land covers in an urban setting and obtained an average OA of 94.4%.

2.4.3.3 Extreme gradient boosting (XGBoost)

XGBoost is an extension of traditional boosting ensemble techniques that are part of the DT family. Boosting generates models sequentially and then combines the weaker performing classifiers into one strong model that constantly improves on the previous classifier errors. The weaker performing classifiers are models that perform only slightly better than a random classification. XGBoost constructs an additive classifier model while optimising a loss function, which accounts for inaccuracies present in the classification. The loss function is a measurement of how well the model fits the training dataset, and when the loss function reduction becomes limited, the boosting will stop (Xia et al. 2017).

Möller et al. (2016) compared XGBoost, RF and DT boosted using AdaBoost by classifying Supernova Legacy survey data. They found that XGBoost outperformed both RF and DT boosted using AdaBoost. Xia et al. (2017) found similar results to Möller et al. (2016) when comparing XGBoost to other machine learning classifiers (RF, SVM, NN, LR, DT, Adaboost, Adaboost – NN, bagging – NN, bagging – DT, gradient boosted DT) used to classify credit datasets. XGBoost was also the best performer in their study, and they noted that XGBoost performed well with imbalanced datasets. Georganos et al. (2018) compared XGBoost, RF and SVM for classifying land covers in an urban setting. They used three VHR datasets: WorldView-3 (0.5 m resolution), Pleiades tristereo imagery (0.5 m resolution) and aerial imagery (0.09 m resolution). XGBoost obtained the highest accuracies (above 80%) for all three datasets.

2.4.3.4 k-NN

k-NN, one of the oldest and simplest non-parametric supervised learning algorithms, is a distance-based classification algorithm that generates classification rules from the training data without any additional data (Weinberger, Blitzer & Saul 2006). The algorithm predicts the class of samples in a dataset according to closest nearest neighbour (training data) in feature space and assigns the sample to the class with the largest category probability (Adejuwon & Mosavi 2010). k-NN requires three variables to perform the classification, namely the distance metric (to assess similarity), the number of nearest neighbours and a scheme to weigh individual neighbours, with the last two used for calculating the classification (Chirici et al. 2016).

Blanzieri et al. (Melgani & Blanzieri 2008) compared k-NN and SVM by classifying land covers from aerial imagery (two datasets) and SAR data (one dataset). For the first aerial imagery dataset and the SAR dataset, k-NN performed better than SVM with OAs of 86.7% for the aerial imagery dataset and 93% for the SAR dataset. However, for the second aerial imagery dataset, SVM obtained the highest OA of 88.8%, whereas k-NN obtained an OA of 78.1%. Li & Cheng (2009) classified land covers from SPOT-5 imagery using k-NN and showed that the classification algorithm is capable of obtaining an average accuracy of 85.6%.

2.4.3.5 Logistic regression (LR)

LR has a misleading name, as it is not used for regression but is rather a linear model used for classification. LR is a multivariate analysis model that allows a multivariate regression relationship between a dependent variable and several independent variables (Pradhan 2010). In comparison with ordinary linear regression, the advantage of LR is that the variables in the model can either

be continuous or categorical or any combination of the two types, and it also does not have to have a normal distribution (Lee 2005).

Pradhan & Lee (2010) delineated landslide hazard areas using three methods, namely frequency ratio, LR and NN. Out of the three methods, LR obtained the highest prediction accuracy (90%), frequency ratio obtained the second highest prediction accuracy (86.4%) and NN obtained the lowest prediction accuracy (83.6%). From the results the authors concluded that LR would produce more accurate landslide hazard maps.

2.4.3.6 Naïve Bayes (NB)

NB, a probabilistic classifier based on Bayes' theorem from Bayesian statistics, is an effective classifier that can handle multiple classes. Furthermore, NB is the simplest form of a Bayesian network. NB is 'naïve' since it assumes that all the features in a dataset are independent from each other; however, in practice features are seldom independent (Zelinsky 2009). As opposed to other classifiers, NB does not need to set any tuning parameters, thereby saving time and avoiding subjectivity (Qian et al. 2015).

Solares, Mar & Sanz (2005) performed land cover classification on Landsat TM data using four Bayes classification methods, of which NB was one. The NB classifier obtained a training accuracy of 90.1% and a test accuracy of 89.3%. Han, Zhu & Yao (2012) compared NB to variations of the RF algorithm and NN for land cover classification. Even though the NB classifier obtained the lowest accuracy of 52.3%, they noted that NB produced the most distinct image classification result.

2.4.3.7 SVM

SVM is a non-parametric supervised classification algorithm that was developed in the late 1970s. The algorithm builds a model by mapping the training dataset into higher dimensional space and attempts to separate the different classes using hyperplanes with minimum classification errors. The training samples that lie on the edge of the class distribution in feature space are used to define the optimal hyperplane; these 'edge'-training samples are referred to as support vectors. The training data that is not support vectors do not contribute to defining the hyperplane positions and can be ignored, thus it is possible for SVM to create a high performance classification model using a small training dataset. As is the case with other non-parametric classifiers, SVM does not require the training data to have a normal distribution (Zheng et al. 2015). For training data that has nonlinear boundaries between the class distribution, SVM can use different kernel functions such

as linear, polynomial, radial basis function (RBF) and sigmoid kernels in order to define the hyperplanes (Qian et al. 2015).

Qian et al. (2015) compared SVM to k-NN, DT and NB for land cover classification. The SVM classifier obtained the highest OAs with accuracies ranging between 96.2% to 97.6%, while the NB obtained the second highest OAs with accuracies ranging between 95% to 96.4%. The k-NN and DT classifiers both obtained OAs under 90%. Qian et al. (2015) noted that SVM was able to obtain high accuracy with relative small numbers of training samples compared to the other three classifiers. Gilbertson, Kemp & Van Niekerk (2017) compared SVM, NN, RF and DT for classifying crops using Landsat 8 data. Out of the four classifiers, SVM obtained the highest OA with an average of 82.9%, whereas RF obtained 77.5%, k-NN obtained 61.8% and DT obtained 72.5%. The high accuracies obtained by the SVM classifier were attributed to the classifiers performing well with few training samples, being effective at separating classes with spectral similarities and being robust under conditions of high dimensionality.

2.4.3.8 NN

NN is a type of machine learning that is modelled after the constructs of the human brain. In a human brain, the intelligence is stored in neural pathways as well as in memory; for NN the knowledge is stored in weights applied to each node (neurons) (Miller, Kaminsky & Rana 1995). An NN consists of three different layers, namely an input, hidden and an output layer. The neurons are interconnected processing units that make up a network and work in agreement to solve specific problems such as pattern recognition for data classification. Each neuron has an activation function that consists of weight vectors, input variables and a threshold or bias. The activation function defines the output of a neuron and is determined by a supervised learning process in which training data are used with corresponding correct outputs in order to train the network. When new data are used, the input layer will transfer the data to the hidden layers where the neurons will apply the activation function and classify the new data, and then the output layer will present the final classification (Miller, Kaminsky & Rana 1995; Nogueira, Penatti & Dos Santos 2017).

Another form of an NN is a deep neural network (d-NN), which refers to multi-layered NN. A d-NN has a similar structure to that of NN (an input layer and output layer); however, a d-NN differs from an NN when it comes to the hidden layer. Unlike an NN that has one hidden layer, a d-NN has multiple hidden layers where each hidden layer trains on the output of the previous hidden layer (Goodfellow, Bengio & Courville 2016).

Han, Zhu and Yao (2012) compared NN with a RBF kernel to NB and three variations of RF for classifying Landsat 7 data. Of the classifiers, the NN obtained the highest OA (67%). The best performing RF classifier obtained an OA of 66.7%, while the NB classifier obtained an OA of 52.3%. Han, Zhu and Yao (2012) stated that the NN classifier had the lowest over-learning phenomenon and obtained the highest OA, which shows that the classifier has good generalisation performance. Castelluccio et al. (2015) used d-NN for classifying land cover from two datasets: one consisting of aerial imagery and the second consisting of satellite imagery. They compared two d-NN algortihms, CaffeNet and GoogLeNet, with the latter obtaining the highest OA for both datasets.

2.4.4 Training data

Supervised machine learning algorithms require training data that represent the classes to be used in the classification. The training data provide information about each class that is then used by the classification algorithms to build a model that can classify new data. Ideally, machine learning needs large training datasets; however, it is not always possible to obtain them (Heydari & Mountrakis 2018). The amount of training data should be enough to accurately represent each class, with Campbell & Wynne (2011) recommending at least 100 samples per class. Furthermore, the training data should be as balanced as possible as imbalanced training data can significantly affect the performance of the classification algorithms (Heydari & Mountrakis 2018).

2.5 ACCURACY ASSESSMENT

Different metrics are used to assess the performance of classification algorithms and consist of an accuracy score or OA, kappa, f-score and area under the curve (AUC) score. OA is one of the most widely used metrics and is a measure of the overall proportion of correctly classified data in the dataset (Campbell & Wynne 2011). Kappa is a statistic that measures inter-annotator agreement, which compares the classification accuracy to an expected accuracy. The expected accuracy is a measurement of the contribution of chance agreement to the classification accuracy (Campbell & Wynne 2011; Pontius et al. 2011). F-score is the weighted average of the precision and recall. Precision is the number of true positives over the sum of the number of true positives and false positives, while the recall is the number of true positives over the sum of the number of true positives and false negatives (Li et al. 2012). AUC is the area under the receiver operating characteristic (ROC) curve and measures how well a classification model can classify a binary dataset (Bradley & Bradley 1997).

The McNemar's (Hartfield, Landau & Van Leeuwen 2011) and Friedman (Zimmerman & Zumbo 1993) test can be used to evaluate if there is a significant statistical difference between multiple classification results. Both of these tests are non-parametric. McNemar's test calculates a chi-squared value from a 2x2 matrix based on a binary difference between correctly and incorrectly classified data (Hartfield, Landau & Van Leeuwen 2011). The Friedman test is a generalisation of the Wilcoxon test and is a non-parametric alternative to repeated-measure analysis of variance (ANOVA) that can be used with ordinal, interval and ratio data (Sheldon, Fillyaw & Thompson 1996; Zimmerman & Zumbo 1993).

2.6 SUMMARY

This chapter provided a review of literature on the use of earth observation data for crop type classification. LiDAR data are commonly used as an additional feature in classifications, but are rarely used as the only source of data for classifying crop types. Mathews & Jensen (2012) is the only exception and thus the lack of research on the use of LiDAR data for crop type mapping warranted further investigation and motivated experiment 1 (Chapter 3). For experiment 2 (Chapter 4), LiDAR data are combined with spectral data for differentiating crop types. Sentinel-2 data was chosen for this purposes given that sensor has a high spatial resolution and 13 spectral bands. In addition, Sentinel-2 imagery can be downloaded automatically and at no cost and is as such ideal for operational crop type mapping implementations. From the literature it is clear that VHR aerial imagery is also frequently used for land cover and crop type classifications and it was thus also incorporated in experiment 2, particularly given the growing popularity of VHR imagery provided by drones.

The next chapter (Chapter 3) focuses on classifying vineyards using LiDAR data and machine learning algorithms. A binary classification (vineyards and non-vineyards) along with using the default parameters for the machine learning algorithms were performed in order to gain a better understanding of the LiDAR derivatives. The LiDAR derivatives that were assessed were interpolated at four resolutions and used two window sizes for the texture measures. However, even though the main focus is on the LiDAR derivatives, the machine learning and standardised and unstandardised data are also compared.

The findings of Chapter 3 is used to inform Chapter 4, which compares LiDAR data, Sentinel-2 and aerial imagery along with the same machine learning algorithms used in Chapter 3. The different datasets are individually compared – and then compared in combination with one another – by performing a five-class classification, with four of the classes representing crop types.

CHAPTER 3: REGIONAL MAPPING OF VINEYARDS USING MACHINE LEARNING AND LIDAR DATA[1]

3.1 ABSTRACT

This study evaluates the use of LiDAR data and machine learning algorithms for mapping vineyards at regional scale. Vineyards are planted in rows spaced at various distances, which creates spatial patterns that complicate classification. The varied spatial patterns result in interrows that contain cover crop or bare ground, which can cause spectral mixing within individual pixels. Four resolutions (1.5 m, 2 m, 2.5 m and 3 m) were used for generating normalized digital surface model (nDSM) and intensity derivatives from the LiDAR data. In addition, texture measures with window sizes of 3x3 and 5x5 were generated from the different LiDAR derivatives. The different combinations of the resolutions and window sizes resulted in eight datasets that were used as input to 11 machine learning algorithms, namely random forest (RF), decision tree (DT), XGBoost, k-nearest neighbour (k-NN), naïve Bayes (NB), logistic regression (LR), neural network (NN), deep neural network (d-NN), support vector machine (SVM) with linear kernel (SVM L), SVM with radial basis function kernel (SVM RBF) and SVM RBF with grid search (SVM GS). The classification results were analysed by comparing the overall accuracy (OA), kappa, f-score, area under curve (AUC) and standard deviation of each dataset combination. The Friedman and McNemar's tests were used to assess the statistical significance of difference in accuracy among the resolutions, window sizes and classification algorithms. A larger window size (5x5) was found to improve the OA for all the classifier-resolution combinations. The results showed that RF with texture measures generated at a 5x5 window size outperformed the other experiments, regardless of the resolution used. This result was significantly higher than the second best classifier, XGBoost. We conclude that the RF algorithm used on LiDAR derivatives with a resolution of 1.5 m and a window size of 5x5 is the recommended configuration for vineyard mapping using LiDAR data.

3.2 INTRODUCTION

Agriculture directly contributes about 2.5% towards the gross domestic product (GDP) of South Africa (Greyling 2015), with another 14% contributed through related manufacturing and processing (World Wide Fund for Nature 2018). Fruits and vegetables, including grapes, make up 50.8% of food production, with about 90% produced under irrigation (Tibane 2016). Being able to accurately assess the area covered by crops (by creating crop type maps) is vital to government and agricultural-related agencies (Myburgh 2015; Yalcin & Günay 2016). Digital crop type maps

[1] This chapter was submitted as a journal article to International Journal of Applied Geospatial Research.

are often used to obtain agriculture statistics such as crop yield, water stress and soil properties, and can be used in agricultural regions to aid decision-making (Delenne et al. 2010; Lee et al. 2010; Van Niekerk et al. 2018; Turker & Kok 2013).

The traditional approach to mapping field boundaries is to manually digitise them from aerial or satellite imagery. However, manual digitising is time-consuming, labour intensive, costly, subjective and open to human error (Yalcin & Günay 2016). A variety of semi-automated image classification techniques has consequently been attempted to improve efficiencies and reduce costs (Yan, Shaker & El-Ashmawy 2015). Machine learning algorithms are increasingly being used for differentiating crop types in satellite imagery (Gilbertson, Kemp & Van Niekerk 2017; Möller et al. 2016). These non-parametric algorithms are robust under high dimensionality (i.e. large number of input variables) and are able to deal with non-normal distributed data (Al-doski et al. 2013; Gilbertson & Van Niekerk 2017). Popular machine learning algorithms include DT, neural network (NN), RF, k-nearest neighbour (k-NN) and SVM (Al-doski et al. 2013).

One of the challenges with remote sensing methods for semi-automated crop type mapping is that some crops (e.g. fruit trees or vines) are planted in rows spaced at various distances (Mathews & Jensen 2012; Warner & Steinmaus 2005). This creates field-specific spatial patterns that substantially complicates classification. In South Africa, about 113 000 ha is planted with wine and table grapes (Mogala 2012; Wines of South Africa 2018) with different trellising systems. Groups of adjacent vineyards often represent several vine varietals of different ages, causing high spectral variability among parcels. Inter-row spaces are generally left bare for easy access during harvesting, but can be planted with a cover crop or can become overgrown by weeds (Liu & Bo 2015), which can cause spectral mixing within individual pixels, especially when the pixel size of the imagery is larger than the individual plants or rows being targeted (Chen et al. 2018; Zhang et al. 2017). Spectral mixing is less of a problem when very high resolution (VHR) imagery (with ground sampling distances of less than 1 m) is used as individual rows can be analysed separately (Mathews & Jensen 2012). Similarly, spectral mixing in imagery obtained from unmanned aerial vehicles (UAVs) – typically at resolutions of about 0.05 m – is less of a problem because individual plants can be targeted with such imagery. However, the collection of such ultra-high resolution (UHR) imagery is only cost-effective for relatively small areas at a time (Bendig, Bolten & Bareth 2013) and is thus not viable for regional crop or field boundary mapping operations.

Light detection and ranging (LiDAR) is an emerging source of remotely sensed data for agricultural applications. LiDAR is an active sensor that transmits and receives energy pulses in a narrow range of frequencies, which means that it can penetrate vegetation canopies (Campbell & Wynne 2011). Furthermore, LIDAR is not affected by relief displacement (Donoghue et al. 2007;

Yan, Shaker & El-Ashmawy 2015). LiDAR can be used to derive digital surface models (DSMs), digital terrain models (DTMs) and intensity images (Bietresato et al. 2016). DSMs and DTMs can be used to produce a normalized DSM (nDSM), which represents the heights of objects (e.g. trees) on the earth's surface. nDSMs, often called canopy height models (CHMs) in vegetation studies (Duan et al. 2015), are stable across heterogeneous landscapes and can be used to measure vertical structural information of vegetation (Bietresato et al. 2016; Liu & Bo 2015). Intensity images can be used to discriminate between non-metallic or biological objects, with vegetation having the highest and water having the lowest intensity (Antonarakis, Richards & Brasington 2008; Bietresato et al. 2016; Zhou 2013).

LiDAR data have successfully been combined with multispectral (Etoughe Kongo 2015; O'Neil-Dunne et al. 2012; Singh et al. 2012) and hyperspectral (Jahan & Awrangjeb 2017; Liu & Bo 2015) data to perform land cover classifications. O'Neil-Dunne et al. (2012) obtained an overall accuracy (OA) of 95% for seven land cover classes in an urban setting and attributed the high accuracies achieved to the stable nature of the nDSMs derived from LiDAR data. Liu & Bo (2015) combined hyperspectral and a LiDAR-derived CHM for object-based crop species classification. Several texture-based image object features extracted from the 1 m resolution CHM were used to describe the visual homogeneity of the image and to show important information about structural arrangements of spatial entities and their relationship to the environment (Chica-Olmo & Abarca-Hernández 2000; Zhang & Zhu 2011). Texture provided supplementary information related to land cover patterns and was useful for discriminating between heterogeneous crop fields (Chica-Olmo & Abarca-Hernández 2000; Peña-Barragán et al. 2011). In Liu & Bo (2015), the best OAs were achieved when the VHR hyperspectral imagery, CHM, textural and geometric features were combined. Jahan & Awrangjeb (2017) combined hyperspectral and LiDAR derivatives for per-pixel land cover classification and achieved similar results to Liu & Bo (2015), where the combination of hyperspectral and LiDAR data obtained the best OAs. Jahan & Awrangjeb (2017) used six LiDAR derivatives, namely DSM, DTM, nDSM, difference between the first and last returns, intensity and nDSM entropy. They found that a combination of hyperspectral and LiDAR data obtained the highest OA and that using entropy and the difference between the first and last returns improved the OA by 1.53% on average.

Examples where land cover classification were performed on LiDAR data only (i.e. without optical imagery) include Brennan & Webster (2006) and Antonarakis, Richards & Brasington (2008). Brennan & Webster (2006) applied a rule-based classification on five 1 m resolution features (DSM, DTM, nDSM, intensity image and a multiple return image) derived from LiDAR data to classify 11 land covers. OAs of 94% and 98% were achieved with LiDAR height values

contributing the most to discriminating between classes, whereas intensity values were useful for differentiating between sub-classes (Brennan & Webster 2006). Antonarakis, Richards & Brasington (2008) performed a nine-class land cover classification using six LiDAR derivatives (CHM, percentage canopy hits model, intensity image, skewness model and kurtosis model) at 5 m resolution. They evaluated three classification methods, namely Verdun, Monbequi and Chatel and obtained OAs ranging from 91.5% to 97.2%. Zhou (2013) compared the value of using LiDAR data on its own to when it is combined with optical imagery. Two LiDAR derivatives (nDSM and intensity image) at 1 m resolution were used to classify four land covers. An OA of 90.7% was achieved when only the nDSM and intensity image were used as input to the classification, while the OA was only marginally higher (92%) when optical imagery was combined with the nDSM. From these studies one can conclude that LiDAR data hold much potential for classifying land cover and crop types. Canopy point frequency and multiple returns have been found to be particularly effective for classifying tall, woody vegetation (e.g. trees) (Antonarakis, Richards & Brasington 2008; Brennan & Webster 2006), while an intensity image can be used to classify grass or short vegetation as it often has a high intensity value (50% reflectance) (Antonarakis, Richards & Brasington 2008).

No research on the use of LiDAR data for mapping vineyards was found in the literature. The only exception is Mathews & Jensen (2012), who performed vineyard delineation on a 7.8 km^2 area with LiDAR data by deriving a 0.6 m nDSM from LiDAR data with a point cloud density of 0.33 points/m^2 and a point spacing of 1.74 m. The main purpose of the nDSM was to differentiate between vegetation heights. Focal statistics were used to generalise the nDSM as it helped to estimate the mean height of vines planted in rows. A window size of 12x12 pixels was found to adequately fill the gaps between rows. ISODATA, a popular unsupervised classifier, was used to group the generalised nDSM pixels into six clusters, which were then manually categorised into vineyards and non-vineyards. The resulting OAs ranged from 96.96% to 98.15%, confirming that LiDAR data hold much potential for vineyard mapping, at least at local scale. However, it would be worth extending this work by investigating whether supervised machine learning approaches can achieve similar or better results and whether it can be used to map vineyards of different ages, varietals, vine spacing and trellising systems.

This study investigates the performance of LiDAR data and various machine learning algorithms for vineyard mapping in an area with diverse vineyards, surrounded by range of land uses (e.g. urban, agriculture, natural vegetation). Although LiDAR is often used in combination with spectral (multispectral or hyperspectral) data for vegetation mapping (Liu & Bo 2015; O'Neil-Dunne et al. 2012; Zhou 2013), it is rarely used as the only data source for crop type differentiation. An

improved understanding of the value of LiDAR data for crop type mapping will be of great value, particularly within the context of increasing availability of such data at regional scales and planned spaceborne LiDAR missions (e.g. GEDI). Specifically, the study aims to determine if LiDAR derivative resolution, texture (generated at different window sizes) and standardisation have a significant impact on overall classification accuracies. A comprehensive set of LiDAR-derived variables was used as input for eleven machine learning classifiers, namely RF, DT, k-NN, logistic regression (LR), naïve Bayes (NB), NN, deep neural network (d-NN), SVM with a radial basis function kernel (SVM RBF), SVM with linear kernel (SVM L), SVM RBF with grid search (SVM GS) and extreme gradient boosting (XGBoost). Lower resolutions (above 3 m) and window sizes (above 5x5) were excluded given that a spatial accuracy of 15 m or higher is generally required for farm level agricultural applications. The results from the study is intended to be used for operational vineyard monitoring in South Africa, which may involve combining LiDAR data with other sources of remotely sensed imagery.

3.3 MATERIALS AND METHODS

3.3.1 Study area

Several study sites located in the City of Cape Town (South Africa) were chosen for carrying out the experiments (Figure 3.1). The total area of the sites is 595 km^2, of which about 80 km^2 (13.4%) is planted with vineyards. The sites were chosen due to the availability of LiDAR data and the mixture of vineyards and land cover/uses present. Land cover/use in the area is dominated by urban, natural vegetation and agriculture, of which most comprise vineyards representing a large range of varieties, ages, vine spacing and trellising systems. The region has a Mediterranean climate characterised by dry warm summers and cool wet winters, with average summer temperatures of 24–35 °C and average winter temperatures of 7–15 °C. The annual rainfall is 515 mm (Cousins et al. 2013; Gaigher & Samways 2010).

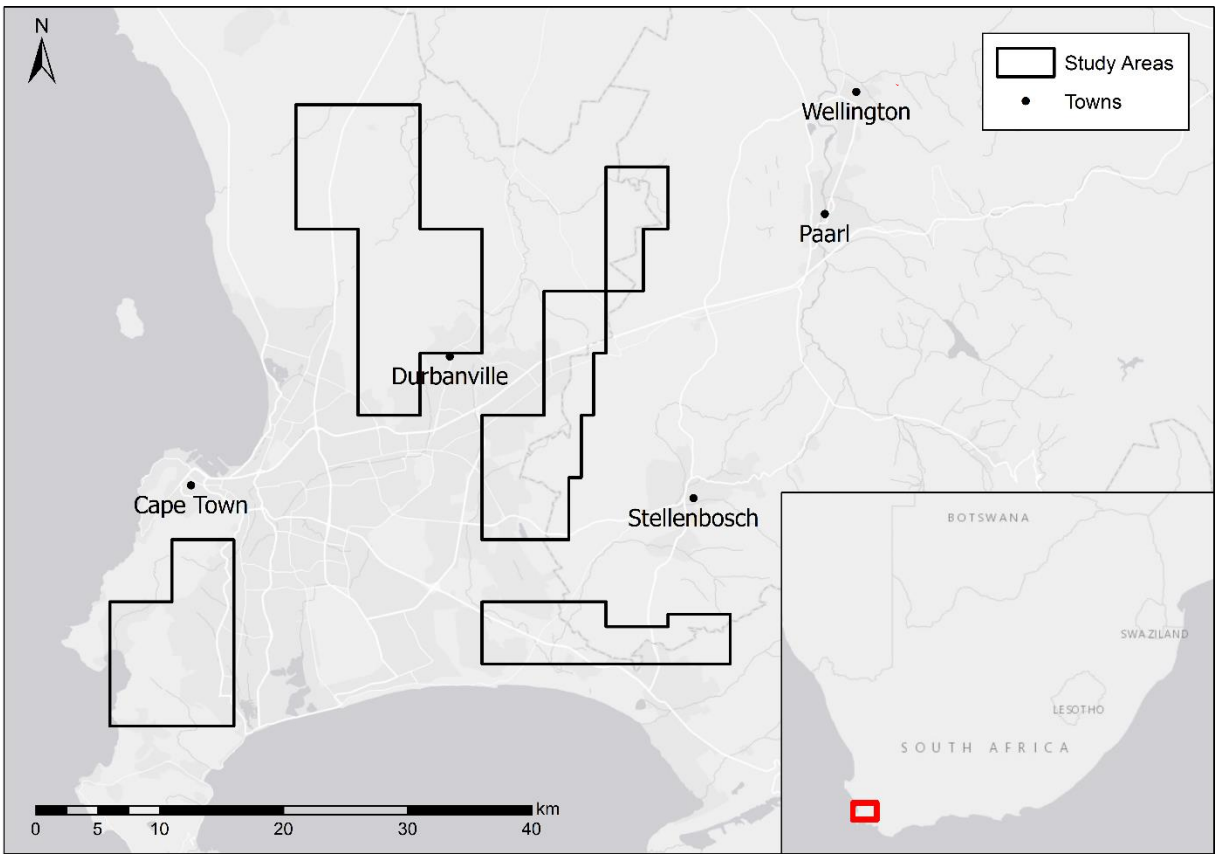


Figure 3.1: Study area in Cape Town, South Africa

3.3.2 LiDAR data acquisition and feature set preparation

The LiDAR survey was commissioned by the City of Cape Town and carried out in 2014. The data have an average point spacing of 0.74 m and an average point density of 1.83 points/m^2. Four nDSMs and intensity images were interpolated from the LiDAR data at 1.5 m, 2.0 m, 2.5 m and 3 m resolution respectively. The interpolations were performed in ArcGIS 10.4 using the inverse distance weighted (IDW) interpolation algorithm (Liu et al. 2007). Resolutions lower than the largest point spacing (i.e. less than 1.32 m) were not considered to prevent artefacts (Liu et al. 2007). The nDSMs were generalised using a focal (neighbourhood) statistic. The focal statistic was set to use the range (minimum value subtracted from the maximum value) of the neighbourhood (Mathews & Jensen 2012) and was performed with two different window sizes, namely 3x3 and 5x5. The generalisation was performed on all four resolutions of the nDSM. Owing to their positive impact on crop classification (Liu & Bo 2015), histogram-based texture measures (HISTEX) and texture analysis (TEX), as implemented in PCI Geomatica, were applied to the ungeneralised nDSMs and intensity images at 3x3 and 5x5 window sizes. LiDAR multi-return values were also included in the feature set, making up a total of 53 variables (Table 3.1).

Table 3.1: LiDAR features used as input to the classifiers

Type	Features	Number of features
LiDAR derivatives	nDSM	3
	Intensity	
	Focal nDSM	
Textural features	HISTEX: Mean, median, mean deviation from mean, mean deviation from median, mean Euclidean distance, variance, coefficient of variation, skewness, kurtosis, energy, entropy, weighted-rank fill ratio	24
	TEX: Homogeneity, contrast, dissimilarity, mean, variance, entropy, angular second moment, correlation, grey level difference vector (gldv) angular second moment, gldv entropy, gldv mean, gldv contract, inverse difference	26
	Total number of features:	53

Note: The resolutions of the features are not consistent (see Table 3.2)

The features were standardised using zero mean unit variance standardisation (Equation 3.1) (Jonsson et al. 2002) (also called z-score standardisation) as it was a requirement for the d-NN and can also improve the performance (accuracy and efficiency) of classification algorithms such as NN, k-NN and clustering classifiers (Shalabi & Shaaban 2006). The zero mean and unit variance standardisation is the most commonly used score standardisation technique (Mohabeer, Soyjaudah & Pavaday 2011) and was implemented using the Scikit-learn Python package, an open-source machine learning library developed by Pedregosa et al. (2012). Zero mean unit variance standardisation is defined as:

$$x' = \frac{x - \bar{x}}{\sigma} \qquad \text{Equation 3.1}$$

where x is the original value;

$\bar{x}$ is the mean of the feature; and

σ is the standard deviation of the feature.

3.3.3 Reference data

A reference database containing crop types was obtained from the Western Cape Department of Agriculture. The polygons database, representing agricultural fields and crop types, was created in 2013 following extensive field surveys. The time delay between the field surveys and the LiDAR campaign (2014) was considered insignificant given that the target crop was grapes, which is perennial and thus unlikely to have changed. This assumption was supported by a systematic visual assessment of high resolution aerial photographs acquired in 2014. Stratified random sampling

was used to extract 18 000 reference points, with each target class (vineyard and non-vineyard) being represented by 9000 samples. The points were used to spatially extract the values of the features in Table 3.1.

3.3.4 Classification and accuracy assessment

The classifications and accuracy assessments were performed with Scikit-learn. The library includes a wide range of classification algorithms, including RF, SVM, NN, k-NN, DT and NB. Xgboost, as implemented by Chen et al. (2017), was also applied. Tensorflow, developed by Abadi et al. (2016), was used as the d-NN. The default parameters were used for all classifiers except for the SVM GS and d-NN algorithms. The d-NN was configured to make use of three hidden layers, while the SVM GS algorithms iterated through different combinations of the parameters to optimise the classification (Hsu, Chang & Lin 2008).

The classification algorithms were iterated 100 times to assess model stability. For each iteration, the samples were randomly subdivided into subsets for model training (70% of samples) and accuracy assessment (30% of samples). The OA, kappa coefficient, f-score and AUC score were automatically calculated using Python code. The standard deviation (SD) of the OAs was used to assess stability.

The classification experiments were performed on 16 groupings of features, eight of which were standardised and the remaining eight were unstandardised (Table 3.2). The standardised and unstandardised datasets each consisted of two sub-groups, one in which a window size of 3x3 was used to generalise the features, and another in which a window size of 5x5 was used. Each of these groups considered features generated at four resolutions, namely 1.5 m, 2 m, 2.5 m and 3 m. All 11 classifiers were applied to each of these 16 datasets, which resulted in 176 experiments in total.

Table 3.2: Dataset configurations

Type	Window size	Resolution
Unstandardised	3x3	1.5 m, 2.0 m, 2.5 m, 3.0 m
	5x5	1.5 m, 2.0 m, 2.5 m, 3.0 m
Standardised	3x3	1.5 m, 2.0 m, 2.5 m, 3.0 m
	5x5	1.5 m, 2.0 m, 2.5 m, 3.0 m

The results were compared using the Friedman test (Zimmerman & Zumbo 1993) and the McNemar's test (Hartfield, Landau & van Leeuwen 2011) to evaluate if there was a significant

statistical difference between the classifiers, windows sizes and resolutions. The Friedman test is a generalisation of the Wilcoxon test and is a non-parametric alternative to repeated-measure ANOVA (Zimmerman & Zumbo 1993). The Friedman test was particularly suitable for this study as it can be used with ordinal, interval and ratio data (Sheldon, Fillyaw & Thompson 1996). The McNemar's test is a non-parametric test that calculates a chi-squared value from a 2x2 matrix based on a binary difference between correctly and incorrectly classified data (Hartfield, Landau & van Leeuwen 2011). P-values lower than 0.05 were considered significant.

3.4 RESULTS

The OAs of the experiments are summarised in Table 3.3. The full results containing the kappa, f-score, SD and AUC are provided in Appendix A. Table 3.3 shows the results for both the standardised data and the unstandardised dataset. Table 3.3 also shows the mean OAs for the different dataset configurations and classification algorithms used.

RF achieved the highest classification accuracy, irrespective of standardisation, with a mean OA of 79.4%. XGBoost (unstandardised) and NN (standardised) produced the second and third best results, with mean OAs of 78.2% and 77.5% respectively. McNemar tests revealed that there is a significant difference between the RF results and XGboost results, whereas there is not a significant difference between the XGboost and NN results. Overall, LR was the weakest classifier, achieving a mean OA of 49.9%.

When the Friedman test was applied to the different window sizes and resolutions, the resolutions to which a 3x3 window size were applied had OAs that showed statistically significant differences. However, when a 5x5-window size was applied, the difference in OAs was statistically insignificant. These results essentially show that, by applying a 5x5 window size, the resolutions (tested in this study) did not affect the results. However, when a 3x3-window size was applied, the resolutions did affect the results, but this configuration did not perform as well as when a 5x5-window size was used.

An increase in accuracy was observed for seven of the classifiers (NN, k-NN, LR, NB, SVM L, SVM RBF and SVM GS) when the input features were standarised, with LR and SVM RBF benefiting the most (mean OA increase of 22.3%). These increases in accuracy were all statistically significant ($p < 0.05$). The remaining three classifiers (RF, DT and XGBoost) recorded a slight decrease in OA when the input features were standardised, with RF being the most affected (OA decreased by 0.8%). However, this difference was not statistically significant. The d-NN classifier failed when unstandardised input features were used as input. The SD values (provided in

Appendix A) showed that the stability of most classifiers increased (i.e. SD decreased) with feature standardisation, whereas RF, XGBoost and NB showed a slight decrease in stability.

Table 3.3: Overall accuracy results for the standardised and unstandardised dataset. All the classifiers are shown, except for deep neural network for the unstandardised dataset

	Classifier	3x3 window				5x5 window				Classifier mean
		Resolution				Resolution				
		1.5 m	2.0 m	2.5 m	3.0 m	1.5 m	2.0 m	2.5 m	3.0 m	
Standardised	D-NN	73.8	75.2	75.6	75.7	77.7	78.1	77.8	77.9	76.5
	DT	67.3	68.4	68.7	69.2	70.3	70.8	71.1	71.3	69.6
	K-NN	69.3	70.3	70.4	70.5	75	75.3	75.1	74.6	72.6
	LR	69.7	70.4	71.5	71.7	73.4	73.1	73.7	73.6	72.1
	NB	56	56.3	56.2	60.4	58.3	57	56	58.8	57.4
	NN	74.8	76.2	76.5	76.8	78.8	79.2	78.9	79.2	77.5
	RF	76.4	77.5	77.8	78.1	79.9	80.2	80.2	80.4	78.8
	SVM GS	69.9	70.6	71.6	71.6	74.1	73.5	73.6	74.1	72.4
	SVM L	68.6	69.4	70.8	71	72.9	72.5	73.3	73.1	71.5
	SVM RBF	71.7	72.4	72.9	73	75.5	75.9	75.9	75.8	74.1
	XGBoost	75.4	76.4	76.8	77.6	78.7	79	79.2	79.4	77.8
	Mean	70.3	71.2	71.7	72.3	74	74	74.1	74.4	
Unstandardised	D-NN	-	-	-	-	-	-	-	-	-
	DT	67.7	68.5	68.9	69.7	70.5	71	71.3	71.7	69.9
	K-NN	62.6	63.4	63.6	64.7	64.4	65.3	64.9	66	64.4
	LR	50	50.5	51.1	49.9	49.5	50.1	49	48.9	49.9
	NB	50.5	50.5	50.6	50.5	50.6	50.6	50.5	50.5	50.5
	NN	61.4	62.9	63.3	64.3	64	64.4	64.7	65	63.7
	RF	77.1	77.9	78.3	78.8	80.4	80.7	80.7	80.9	79.4
	SVM GS	63.9	65.2	64.9	67	67.2	68.6	67.6	70.1	66.8
	SVM L	61.7	61.9	63.9	63.6	63.9	64.2	66.5	66.1	64
	SVM RBF	52.2	52.4	52	52.6	51.1	51.5	51.3	52	51.9
	XGBoost	76	76.6	77	78	79.1	79.3	79.6	79.9	78.2
	Mean	62.3	63	63.4	63.9	64.1	64.6	64.6	65.1	

Note: Deep neural network failed on unstandardised features

Figure 3.2 is a visual representation of the RF classification output performed on the unstandardised 1.5 m resolution features, which were generalised using a 5x5 window size. The area covers 25 km^2 and was randomly selected for a qualitative assessment. The classification (Figure 3.2a) agrees very well with the aerial photograph of the same area (Figure 3.2b), although some false positives are noticeable in some areas and the map suffers from the typical salt-and-pepper (noise) effect of per-pixel classifications. Despite the false positives and noise in Figure 3.2a, vineyards are the only land cover represented by relatively uniform areas, which suggests the classification was a success.

Figure 3.2: Random forest classification result (a) performed on the LiDAR dataset resampled to 1.5 m and generalised using a 5x5-window size compared to (b) an aerial photograph of the same area

3.5 DISCUSSION

Our experiments showed that classifier performance varied substantially and that they can be ranked (based on OAs when input features were standardised) in descending order of accuracy as: RF > XGBoost > NN > d-NN > SVM RBF > k-NN > SVM GS > LR > SVM L > DT > NB. OAs of the three best performing classifiers in this study (RF, XGBoost and NN) were very similar, with the largest difference being 1.6% between RF and NN. The difference between RF and both XGBoost and NN was statistically significant, while the difference between XGBoost and NN was not. This confirms that the RF was the superior classifier in this study. These results agree with those of Ma et al. (2017) who reviewed several studies and found that RF generally outperforms other classifiers. Reviewing several studies, Khatami, Mountrakis & Stehman (2016) also noted RF's superiority over DT, and that k-NN performed slightly better than DT in most cases. However, the relatively poor performance of SVM in our study does not agree with their findings. Jahan & Awrangjeb (2017) applied SVM and DT on LiDAR features only – to differentiate several land covers – and found that DT outperformed SVM, which disagrees with our results.

Kappa, f-score, SD and AUC measures provided in Appendix A are generally in agreement with the OA results shown in Table 3.3. However, kappa values were generally lower, with the highest being 0.61 and lowest 0. This is likely attributed the dataset only contains two classes. Since kappa compares the classification to a classification by chance, there is a better chance that the

classification by chance would result in a correct classification when two classes are used and thus a less than favourable kappa value.

Very little is known about the benefits of standardising remote sensing data for input to machine learning classifiers. In contrast to Jonsson et al. (2002) who noted reduced accuracies when zero mean and unit variance standardisation was employed on imagery for face authentication (using SVM as classifier), our results show that most of the classifiers (NN, d-NN, k-NN, LR, NB, SVM L, SVM RBF and SVM GS) benefitted significantly from standardisation. In fact, standardisation had a more significant and pronounced positive effect on the SVM classifications than the grid search parameter optimisation (compare SVM RBF results in Table 3.3). Grid search involves an exhaustive search through a manually specified subset of parameters to find the parameter values that will result in the highest accuracies (Lin et al. 2008) and is as such computationally expensive. Standardisation is a much simpler procedure and consequently requires less processing time. According to Pedregosa et al. (2012), the objective function (such as the RBF kernel of SVM) of a learning algorithm expects all the features to be centred around zero and have a variance of the same order, thus an increase in accuracies is expected after standardisation. Furthermore, Shalabi, Shaaban & Kasabeh (Shalabi, Shaaban & Kasasbeh 2006) stated that standardisation improves the performance of distance-based classifiers (nearest neighbour or clustering classifiers) because it prevents features with large ranges from outweighing (have a greater influence than) features with smaller ranges. Standardisation may also improve the efficiency of classification algorithms such as NN, k-NN and clustering classifiers, which could explain the increase of stability for NN, LR, SVM L and SVM RBF (Shalabi, Shaaban & Kasasbeh 2006). Furthermore, the results of the standardised and unstandardised datasets were very similar to each other for the RF and XGBoost classifiers, which suggests that these classification algorithms are more robust when classifying vineyards using LiDAR data.

The experiments in which 5x5 window sizes were used for image texture calculations performed consistently better than when 3x3 window sizes were employed. This is attributed to the 5x5-window size being large enough to geographically cover multiple vine rows and thereby reducing the effect inter-row gaps. This is in agreement with Warner & Steinmaus (2005), who stated that the windows size only needs to be large enough to capture repeating features of interest.

The differences among the results of the experiments carried out at different resolutions were not statistical significant when a 5x5-window size was used, but differed significantly when a 3x3-window size was used. Generally, resampling the LiDAR data to lower resolutions produced higher accuracies, but at the expense of lower spatial precision. From these results, it would seem

that a resolution of 1.5 m with a 5x5-window size is most suitable as this combination produced high classification accuracies, while maintaining an effective spatial precision of 7.5 m.

The salt-and-pepper-effect visible in Figure 3.2a can be reduced by using a post-classification majority (mode) filter. This will likely increase the accuracies reported in Table 3.3. Alternatively, a geographic object-based image analysis (GEOBIA) approach can be used instead of using a per-pixel approach as it has been shown to reduce the salt-and-pepper effect and could thus improve the OAs obtained from the classifiers (Blaschke 2010; Whiteside, Boggs & Maier 2011).

The accuracies obtained in this study were generally lower than those obtained by Mathews & Jensen (2012). The complexity of our study area may have affected overall accuracies as it contains a large range of grape varieties, surrounded with many other land cover/uses. Our classifications were also applied at regional scale (595 km^2), whereas Mathews & Jenson (2012) considered a study area of approximately 7.8 km^2 with a relatively small number of vineyards. The way in which the accuracy assessments were carried out could also have contributed to the difference in accuracy. Mathews & Jensen (2012) used an area-based (vineyard level) accuracy assessment approach, whereas in this study a per-pixel approach was used, which makes direct comparison between the studies difficult. Despite these differences, our study agrees with their primary finding that LiDAR is an effective source of data for discriminating vineyards from other land covers. Mathews & Jenson (2012) found that a 0.6 m resolution nDSM generalised with a 12x12-window focal filter produced the best results, which provides an effective spatial resolution of 7.2 m. This compares well with the 7.5 m effective spatial resolution of our recommended configuration (resulting from 1.5 m resolution data and a 5x5-window size) and shows that an effective resolution around 7.5 m is low enough to capture the spatial pattern of the vineyards, but not so large that it would result in over generalisation. We were unable to replicate their recommended configuration as the density of our LiDAR data were insufficient to produce a 0.6 m nDSM, but based on our findings we believe that the use of a 1.5 m nDSM is sufficient. In addition, using a lower resolution nDSM has two main benefits: 1) it reduces the size of the dataset and thus decreases processing time; and 2) it reduces interpolation errors, especially when the nDSM cell size is larger than the mean LiDAR point spacing (Liu & Zhang 2007).

The availability of LiDAR data in agricultural areas will likely increase as its value becomes more apparent. Crop type differentiation is only one potential application of such data. Other uses include crop (mainly tree crops) health assessments, canopy volume assessments, erosion monitoring, drainage system design and field layout planning (Howard 2015; Jang et al. 2008). As an example, a LiDAR survey was recently completed for the Vaalharts irrigation scheme (South Africa) and several other areas are being targeted for future campaigns. The planned GEDI satellite

(which is successfully deployed in late 2018) may also assist in regional crop type mapping. This study provides a good foundation for additional research on the use of such data (along with other remotely sensed data) for this purpose.

3.6 CONCLUSION

This study investigated the performance of LiDAR data and various machine learning algorithms for classifying vineyards. The main purpose was to determine which LiDAR derivative resolutions and image texture window sizes produced the best accuracies when used as input to a range of machine learning algorithms. The results showed that the best classifications were obtained when a 5x5 pixel window was used to derive features from a 1.5 m resolution nDSM, as this configuration balances classification accuracy with spatial precision. Most of the machine learning algorithms performed better with standardised input data, while the DT, RF and XGBoost algorithms did not benefit from standardisation. RF is the recommended classifier for vineyard mapping using LiDAR data due to the high accuracies obtained, its robustness and ability to handle unstandardised data.

Having up-to-date regional maps of vineyards and other crops is essential for yield estimations, water management and economical modelling. The findings of this study provides a good motivation for incorporating LiDAR data into operational vineyard (and similar crops) mapping at regional scales. Although the focus of this study was on the use of LiDAR data on its own (and on finding the best data configurations) for vineyard mapping, combining such data with other remotely sensed data (e.g. satellite imagery) will likely improve classification accuracies. However, it is conceivable that LiDAR data will be of critical value in such implementations, especially if imagery with spatial resolutions lower than the row spacing of crops (e.g. cell sizes of more than 2 m) are used.

CHAPTER 4: CROP TYPE MAPPING USING LIDAR, SENTINEL-1 AND AERIAL IMAGERY WITH MACHINE LEARNING ALGORITHMS[1]

4.1 INTRODUCTION

Remotely sensed data acquired by satellites or aircraft (manned and unmanned) are frequently used for generating crop type maps (Pádua et al. 2017), with multispectral sensors mounted on satellites being the most popular. Examples of satellite imagery used for crop type mapping include those acquired by SPOT 4 and 5 (Turker & Kok 2013; Turker & Ozdarici 2011), Landsat 8 (Gilbertson, Kemp & Van Niekerk 2017; Liaqat et al. 2017; Siachalou, Mallinis & Tsakiri-Strati 2015), Quickbird (Senthilnath et al. 2016; Turker & Ozdarici 2011), RapidEye (Siachalou, Mallinis & Tsakiri-Strati 2015), and MODIS (Dell'Acqua et al. 2018; Liaqat et al. 2017). Aerial multispectral sensors are also frequently employed, especially if very high spatial resolution imagery is required (Rajan & Maas 2009; Mattupalli et al. 2018; Wu et al. 2017; Vega et al. 2015). Fewer examples of the use of hyperspectral imagery are available (Jahan & Awrangjeb 2017; Liu & Bo 2015; Yang et al. 2013), mainly due to the expense in obtaining such data. Similarly, the use of LiDAR data for crop type mapping is uncommon, with Mathews & Jensen (2012) and Estrada et al. (2017) being notable exceptions.

LiDAR data are becoming increasingly available as more aerial surveys are carried out and Earth observation satellites fitted with LiDAR sensors are launched. For instance, the recent launch of the ICESat-2 LiDAR satellite (September 2018) and the attachment of the GEDI LiDAR sensor to the international space station (December 2018) has opened up many new avenues for research and will provide the first opportunity to map vegetation structure at global scale and at high resolutions (Escobar & Brown 2014). Small factor LiDAR sensors mountable on unmanned aerial vehicles (UAV) will also contribute to increased data availability. These new sources of LiDAR data bode well for the agricultural sector (Sankey et al. 2017) as it will be invaluable for crop type classifications, especially when combined with high resolution, multispectral and multi-temporal optical images such as those provided by the Sentinel-2 constellation. The two Sentinel-2 satellites carry 13-band multispectral sensors with swathe widths of 290 km and resolutions of 10 m, 20 m and 60 m, depending on the wavelength, which is ideal for crop type mapping.

Machine learning has been widely used in remote sensing (Lary et al. 2016). Non-parametric machine learning algorithms are capable of dealing with high-dimensional datasets with non-normal distributed data (Al-doski et al. 2013; Gilbertson, Kemp & Van Niekerk 2017). Commonly

[1] This chapter was published as a journal article in Geo-spatial Information Science (https://doi.org/10.1080/10095020.2020.1782776).

used machine learning algorithms are decision trees (DTs), random forest (RF), neural network (NN), and support vector machines (SVM) (Al-doski et al. 2013; Lary et al. 2016). Given that the Sentinel-2 satellites have been in operation for a relatively short time (since 2015), the number of studies that have used the data for crop type classifications are limited. The combination of this data with machine learning algorithms is particularly scarce. However, three studies, namely Inglada et al. (2015), Matton et al. (2015), and Valero et al. (2016), used SPOT 4-Take 5 and Landsat 8 data to emulate Sentinel-2 data. The studies used machine learning classifiers to create crop type maps with Inglada et al. (2015) using RF and SVM, Matton et al. (2015) applying maximum likelihood (ML) and K-means, and Valero et al. (2016) applying the RF classifier. Inglada et al. (2015) and Valero et al. (2016) both created crop type maps for 12 sites, each in a different country. Inglada et al. (2015) obtained overall accuracies of above 80% for seven of the sites while three of the sites had accuracies of between 50% and 70%. Valero et al. (2016) achieved accuracies of around 80%. Matton et al. (2015) considered eight sites, each in a different country, and obtained accuracies of above 75% for all sites, except in one where an accuracy of 65% was achieved. Two studies, Immitzer, Vuolo & Atzberger (2016) and Estrada et al. (2017), classified crops using Sentinel-2 data with the former using RF to classify seven crop types and the latter using DTs to classify five crop types. Immitzer, Vuolo & Atzberger (2016) compared an object-based image analysis (OBIA) and a per-pixel approach, with GEOBIA obtaining an overall accuracy (OA) of 76.8% and the per-pixel classification obtaining an OA of 83.2%. Estrada et al. (2017) considered two study areas and obtained OAs of 85.6% and 95.6% respectively.

Aerial imagery is obtained from sensors mounted on either a piloted (manned) aircraft or a UAV (Matese et al. 2015). An UAV has the benefit of a low operational cost, but is limited to short flight times that limit the area that can be surveyed (Matese et al. 2015). Vega et al. (2015) used machine learning to classify two crop types (sunflowers and non-sunflowers) using multispectral UAV imagery as input. Three resolutions (0.01 m, 0.3 m and 1 m) were evaluated. OAs of above 85% were obtained at all three resolutions (Vega et al. 2015). Wu et al. (2017) performed a GEOBIA SVM and a per-pixel machine learning classification on 0.4 m multispectral UAV imagery. They achieved an OA of 95% for the object-based SVM classification and the per-pixel machine learning classification obtained an OA of 75% when the multispectral imagery was used (Wu et al. 2017). Mattupalli et al. (2018) tested imagery from two different sensors, with the first sensor mounted on a UAV and the second mounted on a manned aircraft. The imagery from the two sensors were resampled to 0.1 m and then used as input to machine learning to classify three crop types. The OAs achieved ranged from 89.6% to 98.1% with the imagery from the manned aircraft achieving outperforming the UAV imagery.

Several studies have combined aerial and satellite imagery with height data to improve crop type and other land cover classifications. Height data can be obtained using stereo photogrammetry techniques or by using data obtained from RADAR (SAR) or LiDAR. Stereo photogrammetry uses overlapping images to create a digital surface model (DSM) (Wu et al. 2017). SAR imagery can be used to create a digital elevation model (DEM); however, SAR does not provide discrete returns (like LiDAR) that can be used to create detailed digital surface models (DSM) and digital terrain models (DTM). LiDAR is also a form of active remote sensing that captures 3D point clouds of the earth's surface by transmitting and receiving energy pulses in a narrow range of frequencies (Campbell & Wynne 2011; Ismail et al. 2016). LiDAR is commonly used to derive surface height information by either using the 3D point cloud or by interpolating a DSM or digital terrain model (DTM) (Zhou 2013). A normalized DSM (nDSM), or canopy height model (CHM), can be created by subtracting the DSM from the DTM. LiDAR has an advantage over photogrammetric methods in that it generally provides more accurate height measurements, especially for areas with dense vegetation (Satale & Kulkarni 2003). In addition, LiDAR is less affected by weather conditions (Satale & Kulkarni 2003). LiDAR can also penetrate vegetation canopies and obtain height information of the terrain below. The terrain heights can then be used to create a DTM and nDSM with higher accuracy and less effort. Besides height information, LiDAR can also provide returned intensity information, which can be used to discriminate between different land covers. For instance, water results in low intensity returns, while the intensity of returns from vegetation is high (Antonarakis, Richards & Brasington 2008).

The majority of the studies that use LiDAR data for classification derived an nDSM or CHM as they represent only the aboveground features (Yan, Shaker & El-Ashmawy 2015). Chen et al. (2009) combined very high resolution (VHR) Quickbird imagery with LiDAR data to classify land cover in an urban setting. The OA increased from 69.1% to 89.4% when the LiDAR derived nDSM was used and added to the imagery as input to the classifier (Chen et al. 2009). They attributed the increase in accuracy to the height data, which made it easier to differentiate between land covers that have similar spectral signatures. Wu et al. (2017) derived an orthomosaic and a DSM from aerial imagery to classify crop types and obtained an overall increase in OA of 3% when the DSM was used along with the orthomosaics. Estrada et al. (2017) classified a LiDAR point cloud, which was used in an interpolation procedure to produce a rasterised elevation model. The latter was then combined with Sentinel-2 imagery to classify crops. However, unlike Chen et al. (2009) and Wu et al. (2017) who combined the LiDAR data with imagery as input features to the classifiers, Estrada et al. (2017) classified the LiDAR and Sentinel-2 data separately and then combined the results using a post-classification aggregation procedure.

Liu & Bo (2015) classified crops using airborne hyperspectral and a LiDAR derived CHM. They compared five different classification schemes, using SVM as classifier in a GEOBIA environment. The OA increased by 8.2% when VHR hyperspectral data were combined with the canopy height model (CHM) and 9.2% when the CHM was combined with minimum noise fraction transformed (MNF) hyperspectral data (Liu & Bo 2015). The highest OA (90.3%) was achieved when the geometric properties of objects and image textures (that were applied on the CHM) were combined with the untransformed CHM and MNF data. This increase was attributed to the ability of image textures to quantify the structural arrangements of objects and their relationship to the environment. They thus provide supplementary information related to the variability of land cover classes and can be used to discriminate between heterogeneous crop-fields (Chica-Olmo & Abarca-Hernández 2000; Peña-Barragán et al. 2011; Zhang & Zhu 2011). Jahan & Awrangjeb (2017) used hyperspectral imagery and a LiDAR-derived DSM, nDSM and intensity raster to classify five land covers. Their study considered two machine learning classifiers (SVM and DTs) and tested nine different combinations of the hyperspectral and LiDAR data. They found that, for the SVM experiments, OA increased by 7.6% when the DSM was added to the hyperspectral data and by 8.4% when image textures (performed on the DSM) were added. Similar but more modest increases (3% and 4.3% respectively) were observed for the DT experiments.

Although several studies have combined LiDAR data with imagery to classify land cover and crop types, very little work has been done on using LiDAR data on its own for this purpose. A notable exception is Brennan & Webster (2006) who used four LiDAR derivatives (intensity, multiple return, normalized DSM and a DSM) to classify land cover and obtained an OA of 94.3% when targeting ten land cover classes and 98.1% when targeting seven classes. Charaniya & Manduchi (2004) classified four land covers using four LiDAR derivatives (normalized DSM, height variation, multiple returns and intensity) and obtained an OA of 85%. Also using LiDAR derivatives, Mathews & Jensen (2012) obtained an OA of 98.2% when differentiating vineyards from other land covers.

From the literature, it seems that the use of LiDAR data as additional input variables (along with imagery) improves land cover classifications. It is even possible to extract individual crop types (e.g. wine grapes) using LiDAR derivatives only, i.e. without using optical imagery as additional input data to classification algorithms. However, it is not clear what value LiDAR data provide to differentiate different types of crops – when used on its own and when it is combined with satellite and aerial imagery. This study investigates the performance of various machine learning algorithms on different combinations of multispectral and LiDAR data for crop type mapping in Vaalharts, the largest irrigation scheme in South Africa. To our knowledge, no study has assessed

the use of LiDAR data on its own for differentiating multiple crop types. Given that LiDAR data are becoming increasingly available at regional scales – and the likelihood that data from space borne LiDAR (e.g. GEDI, IceSat-2) will soon become common – an improved understanding of the value of such data for crop type classification is needed. The crop type maps produced using LiDAR data are compared to crop type maps that were produced using 20cm aerial and 10m Sentinel-2 imagery. In addition, the LiDAR data are used in different combinations with the Sentinel-2 and aerial imagery as input to the machine learning classifiers. Ten machine learning classification algorithms, namely RF, DTs, extreme gradient boosting (XGBoost), k-nearest neighbour (k-NN), logistic regression (LR), naïve bayes (NB), NN, deep neural network (d-NN), SVM with a linear kernel (SVM-L), and SVM with a radial basis function kernel (SVM RBF), are used to determine which of these classifiers are most effective for crop type differentiation using the selected datasets.

4.2 MATERIALS AND METHODS

4.2.1 Study area

The study area (Figure 4.1) is located in the Vaalharts irrigation scheme in the Northern Cape Province of South Africa. The irrigation scheme is situated at the confluence of the Harts and Vaal Rivers and has a scheduled area of 291.81 km^2 (Van Vuuren 2010). The region has a steppe climate with an average annual temperature of 18.6°C and an average annual rainfall of 437 mm. The selected study site is 303.12 km^2 in size and contains a variety of land covers, including indigenous vegetation, built up, bare ground, water and crops. Cotton, maize, wheat, barley, lucerne, groundnuts, canola and pecan nuts are all grown in the area on a crop rotation basis (Muller & Van Niekerk 2016; Nel & Lamprecht 2011).

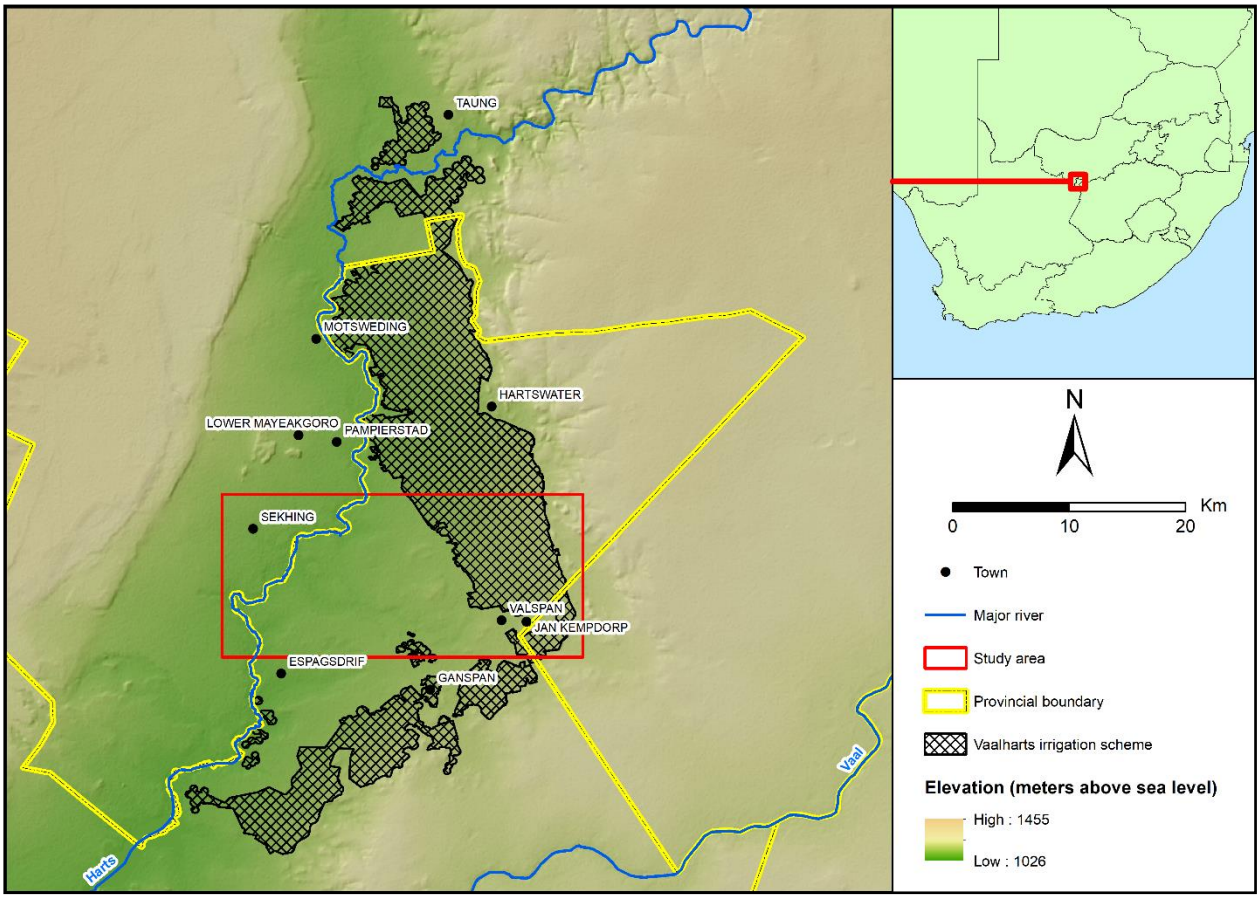


Figure 4.1: Study area Vaalharts irrigation scheme (380 km^2), Northern Cape, South Africa

4.2.2 Data acquisition and pre-processing

Three remote sensing datasets were used in this study, namely LiDAR data, aerial photographs and satellite imagery.

The LiDAR data and aerial imagery were collected by Land Resources International for the Northern Cape Department of Agriculture, Land Reform and Rural Development. The data were acquired between 19 and 29 February 2016 with a Leica ALS50-II LiDAR sensor at an altitude of 4500 ft. The LiDAR data have an average point spacing of 0.7 m and an average point density of 2.04 m^2. The aerial imagery was acquired between 22 February and 18 March 2016 with a PhaseOne iXA sensor at an altitude of 7500 ft. The imagery consisted of four bands, namely blue, green, red and near-infrared (NIR) with the RGB bands having a ground sampling distance (GSD) of 0.1 m and the NIR a GSD of 0.5 m. The RGB bands of the aerial images were resampled to 0.5 m to match the resolution of the NIR band. This dataset was labelled A2. The analysis was also performed on the aerial imagery (A1) at its original resolution (0.1 m for the red, green and blue bands and 0.5 m for the NIR band) in order to assess whether down sampling makes any statistically significant difference. The spectral information of the aerial imagery is shown in Table 4.1.

Table 4.1: Spectral information for the aerial imagery

Bands	Wavelength (nm)	Resolution (m)
Blue	450-480	0.1
Green	550-580	0.1
Red	650-680	0.1
NIR	720-2500	0.5

The Sentinel-2 image was acquired on 10 February 2016. This image was selected because it was the closest cloud-free temporal match to the LiDAR data and aerial photography. Ten bands were used for analysis as shown in Table 4.2. The Sentinel-2 image was atmospherically corrected using ATCOR in PCI Geomatica 2018. Since the Sentinel-2 image was acquired at level-1C orthorectification was already performed on the imagery and thus not required.

Table 4.2: Sentinel-2 bands used for analysis

Bands	Central wavelength (µm)	Resolution (m)	Bandwidth (nm)
Band 2 – Blue	0.490	10	65
Band 3 – Green	0.560	10	35
Band 4 – Red	0.665	10	30
Band 5 – Vegetation red edge	0.705	20	15
Band 6 – Vegetation red edge	0.740	20	15
Band 7 – Vegetation red edge	0.783	20	20
Band 8 – NIR	0.842	10	115
Band 8A – Narrow NIR	0.865	20	20
Band 11 – SWIR	1.610	20	90
Band 12 – SWIR	2.190	20	180

Four features were derived from the LiDAR data, namely an nDSM, generalised nDSM (gen-nDSM), an intensity image and a multi-return value raster. The nDSM was created by interpolating a 2 m resolution DSM and DTM and subtracting the DTM from the DSM. Inverse distance weighted (IDW) interpolation was used for the interpolations. A generalized DSM (gen-nDSM) was created by calculating the range of values within a 5x5 moving window. The range was selected as suggested by Mathews & Jensen (2012), who found that nDSM range within a small window is useful for differentiating between low and high vegetation. The intensity image was interpolated at a resolution of 2 m using IDW. The nDSM, gen-nDSM, and intensity image were created using ArcGIS 10.4. A 10 m resolution multi-return value raster was created by using LiDAR360 1.3. PCI Geomatica was used to apply histogram-based texture measures (HISTEX) and texture analysis (TEX) on the nDSM and intensity image, using a 5x5 window size. Highly

correlated texture features were excluded. The LiDAR derivatives were interpolated/created to match the spatial resolution of the Sentinel-2 bands (2, 3, 4 and 8), which has a resolution of 10 m. The image texture was incorporated in accordance with Liu & Bo (2015). The LiDAR-based features considered in this study are listed in Table 4.3.

Table 4.3: LiDAR features used as input to the classifiers

Type	Features	Number of features
LiDAR Derivatives	nDSM	4
	Intensity	
	Focal nDSM	
	Multi-returns	
Textural features	HISTEX: mean, median, mean deviation from mean, mean deviation from median, entropy, weighted-rank fill ratio	12
	TEX: homogeneity, contrast, dissimilarity, mean, variance, entropy, angular second moment, correlation, inverse difference	18
	Total number of features:	34

A principle component analysis (PCA) was performed on the aerial imagery bands. The same texture features that were used for the LiDAR data were applied to the first principal component (PC1). To accommodate the higher resolution of the aerial imagery, a 51x51 window size was used for generating the texture features from the A1 dataset. The features generated from the aerial imagery are listed in Table 4.4. Due to the lower resolution of the A2 dataset, texture feature were not generated for the A2 dataset.

Table 4.4: Aerial features used as input to the classifiers

Type	Features	Number of features
Spectral bands	Blue	4
	Green	
	Red	
	NIR	
Textural features	HISTEX: mean, median, mean deviation from mean, mean deviation from median, entropy, weighted-rank fill ratio	6
	TEX: homogeneity, contrast, dissimilarity, mean, variance, entropy, angular second moment, correlation, inverse difference	9
	Total number of features:	19

The features stemming from the Sentinel-2 imagery are the 10 bands that had resolutions equal or higher than 20 m (Table 4.2).

By using three data sources (LiDAR, aerial, and satellite data) individually and in combination, seven different experiments were configured, namely aerial (A2 and A1), LiDAR (L), Sentinel-2 (S), aerial and Sentinel-2 (A-S), aerial and LiDAR (A-L), LiDAR and Sentinel-2 (L-S), and lastly LiDAR, aerial and Sentinel-2 (A-S-L). Table 4.5 lists the eight input datasets considered.

Table 4.5: Summary of the different experiment of datasets.

ID	Dataset	Number of features
A1	Aerial	14
A2	Aerial	19
S	Sentinel-2	10
L	LiDAR	34
A-S	Aerial (A2) & Sentinel-2	23
A-L	Aerial (A2) & LiDAR	47
L-S	LIDAR & Sentinel-2	44
A-S-L	Aerial (A2), Sentinel-2 and LiDAR	57

Note: The first column shows the unique identifier, the characters represent the data contained in the dataset with A indicating aerial imagery, S indicating Sentinel-2 imagery, and L indicating LiDAR data.

The datasets were standardised using zero-mean and unit variance standardisation (Equation 4.1) (Jonsson et al. 2002):

$$x' = \frac{x - \bar{x}}{\sigma} \qquad \text{Equation 4.1}$$

where x is the original value;

$\bar{x}$ is the mean of the feature; and

σ is the standard deviation of the feature.

4.2.3 Reference data

A vector database containing crop type data were obtained from GWK (www.gwk.co.za), an agribusiness operating in the study area. The database contains polygons for three crop types, namely, maize, cotton and groundnuts. A fourth crop type, *orchard*, was added to the database by visual image interpretation and manual digitising from aerial imagery. Stratified random sampling was used to create 1000 data points from the crop type database. Each target class (*maize*, *cotton*, *groundnuts*, *orchard*, and *non-agriculture*) was allocated 200 random sample points.

4.2.4 Classification and accuracy assessment

The classifications were performed using the Scikit-learn 0.18.2 Python library. Scikit-learn is an open-source machine learning library developed by Pedregosa et al. (2012) and includes a wide range of classification algorithms (RF, DTs, k-NN, LR, NB, NN, SVM-L, SVM RBF) and metrics, including OA and kappa (K). The Tensorflow 1.2.1 library (Abadi et al. 2016) was used to perform

a deep neural network (d-NN) classification, while XGBoost 0.7 (Chen et al. 2017) was used to perform the XGBoost classification. The d-NN classifier was set to three hidden layers and the other classifiers were configured to use the default parameters.

The algorithms were iterated a hundred times in order to cross-validate and assess the stability of the models. For each iteration, the reference dataset was randomly split into a training (70% of samples) and accuracy assessment (30% of samples) subset. Classification algorithm performance was assessed using four metrics, namely, OA, K, f-score, and standard deviation (SD) of OA. The latter metric was used to assess model stability. The thematic crop type maps resulting from the classifications were also qualitatively assessed by means of visual comparisons.

The Friedman test (Zimmerman & Zumbo 1993) was used to compare the results from the different classifiers and experiments. The Friedman test is a non-parametric alternative to a repeated-measure ANOVA and can be used with ordinal, interval and ratio data (Sheldon, Fillyaw & Thompson 1996; Zimmerman & Zumbo 1993). P-values lower than 0.05 were considered significant.

4.3 RESULTS

The results are summarised in Table 4.6. For sake of readability, only the OAs are shown. Other metrics (Kappa, f-score and standard deviation of the OA) are provided in Appendix B.

4.3.1 Individual dataset–classifier combinations

The most accurate individual classification (OA of 94.6%) was achieved when the A-S-L (aerial, Sentinel-2 and LiDAR) dataset was used as input to the RF classifier (A-S-L>RF scenario). This was followed by the combination of XGBoost with A-S-L (OA of 94.1%); SVM-L with A-S-L (OA of 93.5%); SVM-L with L-S (OA of 93.5%); NN with A-S-L (OA of 93.4%); RF with L-S (OA of 93.2%); and RF with A-S (OA of 93.1%). Although there were no significant differences among the accuracies of the top three results ($P>0.05$), the difference between the A-S-L>RF and L-S>SVM-L scenarios was significant ($P = 0.022$).

Table 4.6: Overall accuracy results for the seven datasets and the ten different classifiers

Classifier	Dataset								Mean	Stdev
	A1	A2	S	L	A-S	A-L	L-S	A-S-L		
d-NN	81	55.2	90.8	83.2	92.3	88.2	91.5	92.2	84.3	11.7
DT	72.2	46.1	81	82.3	86.2	84.7	90.2	90	79.1	13.6
k-NN	77.1	54.5	85.8	83.9	88.9	87.7	91.2	92.1	82.7	11.5
LR	73.2	44.5	85.3	84.9	91.6	86.8	92.2	92.9	81.4	15.2
NB	62.5	46.7	67.9	77.7	74.8	81.2	84.7	86	72.7	12.4
NN	81.2	56.5	88.2	86.3	92.7	89.8	92.8	93.4	85.1	11.5
RF	81.9	54.4	86.5	87.3	93.1	90.7	93.2	94.6	85.2	12.3
SVM-L	73.4	44	88.3	86.2	92.6	88.2	93.5	93.5	82.5	15.8
SVM RBF	72	50.5	75.9	83.4	87	86.6	89.8	90.1	79.4	12.5
XGBoost	81.3	56.1	86.3	87.8	91.9	91.3	93	94.1	85.2	11.7
Mean	75.6	50.9	83.6	84.3	89.1	87.5	91.2	91.9		
Std dev	5.8	4.8	6.6	2.8	5.3	2.8	2.5	2.4		

4.3.2 Dataset performance

Overall, the A-S-L and L-S datasets produced the highest mean OAs, with 91.9% and 91.2% respectively (see last two rows in Table 4.6). The A-S was the third best performing dataset with a mean OA of 89.1%. The mean OA of the A-S dataset was significantly lower than those of the A-S-L (P = 0.011) and L-S (P = 0.011) datasets, while the difference between A-S-L and L-S were not significant (P = 0.002). On average, the A2 dataset performed the worst with a mean OA of 50.9%. The A-S-L dataset obtained the lowest OA standard deviation values (2.4%) and all the datasets containing LiDAR data (L, A-L, L-S, A-S-L) obtained OA standard deviation values of 2.8% or lower. The datasets containing aerial data obtained OA standard deviation values between 4.8 and 5.8 and the S dataset obtained the highest OA standard deviation value of 6.6.

4.3.3 Classifier performance

Overall, RF (85.2%), XGboost (85.2%), NN (85.1%), and d-NN (84.3%) were the best performing classification algorithms (see last two columns in Table 4.6). The differences in mean OA of these classifiers were statistically insignificant (P = 0.119), which suggests that they performed on par with one another. When all five best performing classification algorithms (RF, XGboost, NN, d-NN and LR) were compared, the differences in mean OA were statistically significant (P = 0.001). Similarly, when all the classification algorithms were compared the difference in mean OA were statistically significant (P = 0). The standard deviations of OAs for the classifiers is high (11-16%), mainly due to the A2 dataset's poor classification results. When the A2 dataset is omitted, the standard deviations drop sharply (4-9%), with RF, NN, XGboost, d-NN and k-NN having standard

deviation values of 4-5%. This suggests that RF, NN, XGboost, d-NN and k-NN performed more consistently among different datasets compared to the other classification algorithms.

The performance of the classifiers showed the same pattern as the results in Chapter 3, where RF and XGBoost were the best performing classifiers. However, in this study, d-NN and NN were also among the best performing classifiers. NN and d-NN obtained higher OAs when used to classify the Sentinel-2 data, while RF and XGBoost obtained higher OAs when used to classify the LiDAR data. When the LiDAR and Sentinel-2 data were combined, RF and XGBoost classifiers obtained similar OAs, with the former outperforming the latter by only 0.2%. Furthermore, nine classifiers obtained their best OAs when performed on the aerial imagery, LiDAR and Sentinel-2 experiment (A-S-L), with only DT obtaining its highest OA for L-S. Eight of the classifiers (RF, XGBoost, DTs, k-NN, LR, d-NN, SVM-L and NN) obtained OAs higher than 90% when performed on the L-S experiment, whereas SVM RBF and NB obtained OAs of 84.7% and 89.8% respectively.

The following sections focus on the classification accuracies per crop type. For the sake of brevity, only the results of the best-performing classifier, RF, are shown.

4.3.4 RF per-class performance (per dataset)

A confusion matrix for each of the RF experiments is provided in Appendix C. Table 4.6 summarizes the per-class performances of all experiments, while the errors of commission and omission are listed in Table 4.7.

Table 4.7: Error of commission and omission for all five class. Only the errors of commission and omission for the random forest classifier are shown

Class	Error (%)	A1	S	L	A-S	A-L	L-S	A-S-L
Non-Agri	Commission	21.8	8.3	23.2	5.8	17.2	8.8	5.8
	Omission	22.0	13.0	27.7	10.1	18.9	10.6	10.0
Maize	Commission	17.4	8.3	6.7	7.4	6.3	6.0	5.3
	Omission	19.4	2.4	3.1	1.0	2.2	1.5	1.5
Orchard	Commission	4.8	9.4	4.3	4.0	3.3	3.5	3.1
	Omission	12.3	21.2	6.2	6.4	6.5	3.9	3.7
Groundnut	Commission	23.7	19.8	16.1	7.8	10.7	6.8	5.1
	Omission	21.3	14.3	17.4	9.9	12.4	8.4	6.0
Cotton	Commission	23.1	21.6	13.6	9.8	8.8	8.9	7.6
	Omission	16.2	15.2	8.2	6.8	6.0	9.2	5.6

The A-S-L>RF experiment performed the best and was the only scenario in which the omission and commission errors were equal to or below 11% for all five classes. Generally, the *non-agri* class was the most confused with other classes. For instance, the *non-agri* class was most

frequently misclassified as orchards, with the highest number (160) of false positives (FP) followed by *groundnuts* (108). *Cotton* and *maize* obtained low FP values of 47 and 26 respectively.

Similar to the A-S-L>RF experiment, the *non-agri* class was the most confused with other classes when the L-S dataset was used as input to the RF classifier. The errors of commission and omission of 8.8% and 10.6% respectively are also comparable to those obtained with the A-S-L>RF experiment.

The A-L>RF experiment performed relatively poorly, with the highest OA being 91.3% (XGBoost). *Maize*, *orchard,* and *cotton* performed on par with one another with and all three obtaining errors of commission and omission below 9%. *Groundnuts* was the second-worst performing crop type, with error of commission and omission values of 10.7% and 12.4% respectively. *Non-agri* was the worst performing class for the A-L>RF experiment with error of commission and omission values of 17.2% and 18.9% respectively.

In the L>RF experiment, *maize*, *orchard* and *cotton* were the most accurately classified, with errors of commission and omission below 10%, except for *cotton* which obtained an error of commission of 13.6%. *Groundnuts* obtained the highest error of commission (16.1%) and omission (17.4%) out of the crop classes. As with previous experiments, the *non-agri* was the most difficult to classify in this experiment, with error of commission and omission values of 23.2% and 27.7% respectively.

Out of all the single-sensor experiments, the S>RF experiment returned the lowest error of commission and omission values for the *non-agri* class. The *non-agri* class was the second-best performing class, with *maize* being the most successfully classified. *Groundnuts* and *cotton* performed on par with each other, but where the most difficult crop types to differentiate in the S>RF experiment.

Overall, A2>RF was the worst performing experiment, with *orchard* being the class that was most accurately classified. *Maize* was the second-best performing class for the A2>RF experiment, followed by *non-agri* and *cotton* (both obtained similar results). *Groundnuts* was the worst performing class. When the A2 dataset was used as input to RF, m*aize* was most frequently confused with cotton, while *non-agri* and *groundnuts* were also often confused.

4.3.5 Qualitative evaluation

Figure 4.2 shows a visual comparison of seven experiments (A2>RF, S>RF, L>RF, A-S>RF, A-L>RF, L-S>RF and A-S-L>RF) and an RGB image for orientation. The main purpose of this qualitative evaluation is to compare the quantitative results to the spatially represented classified data. From visual inspection, the RF maps compare well with local knowledge. The only exception

is in the A>RF experiment, in which *non-agri* was relatively well differentiated, while the majority of the fields were classified as either *maize* or *cotton*, which is not realistic. The L>RF experiment often misclassified *non-agri* as *orchard* and *groundnuts*. Natural vegetation, power lines and urban areas were most often confused with these classes. As shown in the confusion matrices, the S>RF experiment had the lowest number of misclassifications for *non-agri*, which seems to be in good agreement with the thematic map produced from the experiment. Based on a visual inspection, the L>RF experiment seems to have generated the most misclassifications. A comparison of the thematic maps produced by the S>RF and the L>RF experiments reveals that the former classified *non-agri* best. Conversely, the L>RF experiment classified crops better (more evenly). Combining the L and S datasets improved the crop type classifications, which is in agreement with the quantitative assessments (error of commission and omission values of below 11%). The A-S>RF and S>RF maps seem very similar, although the *orchard* class seems to be better depicted in the latter experiment. There is little difference between the A-L>RF, L-S>RF and A-S-L>RF maps, apart from in the *non-agri* class which seems to be better classified in the latter experiment, which corresponds with the quantitative assessments

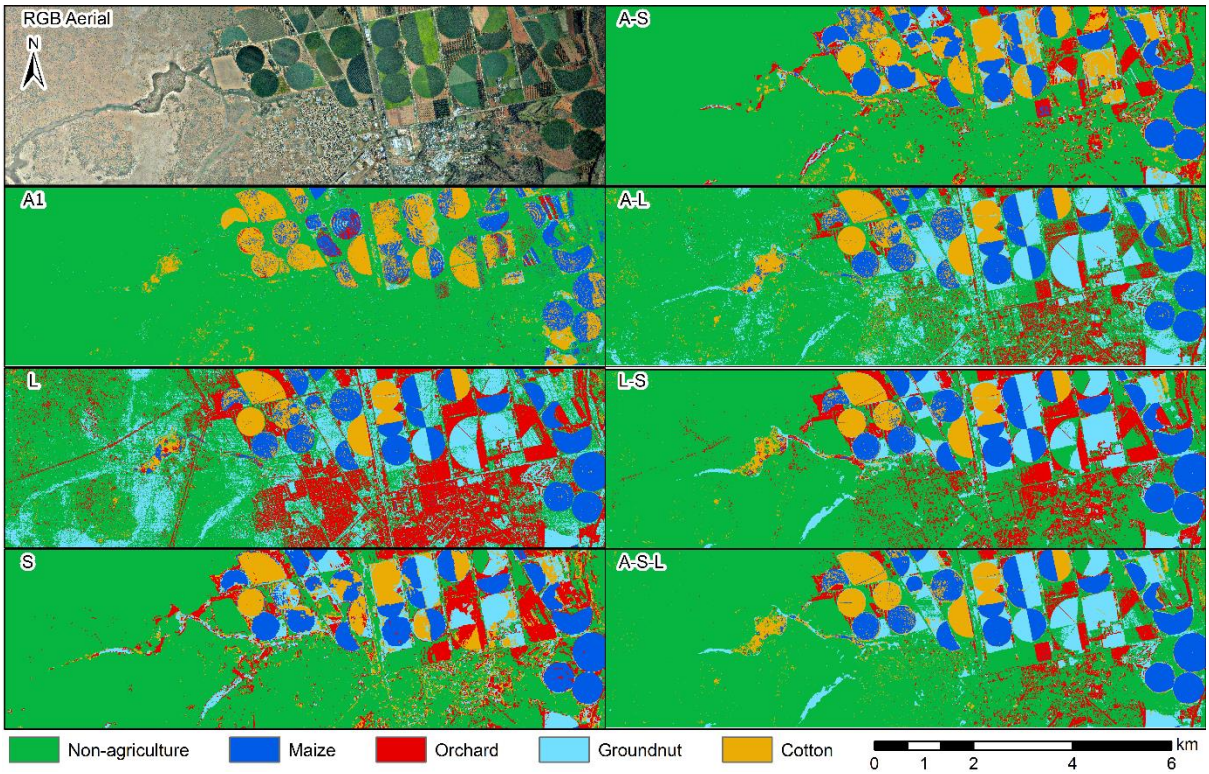


Figure 4.2: Visual comparison of the random forest classification algorithm for the seven experiments, with the RGB aerial photograph in the top left corner for orientation.

4.4 DISCUSSION

The experiments showed that five classes (*non-agri*, *groundnuts*, *cotton*, *maize* and *orchards*) can be accurately classified using machine learning and different combinations of data (aerial, LiDAR and Sentinel-2 data). Nine of the ten machine learning classification algorithms were able to obtain OAs of above 90%, with RF obtaining the highest OA (94.6%). The datasets used in this study were able to obtain acceptable OAs when used on their own as input for the machine learning algorithms, with LiDAR and Sentinel-2 obtaining similar OAs. However, when the datasets were combined, specifically the LiDAR and Sentinel-2 data, higher OAs were obtained.

Using only Sentinel-2 data (S>RF experiment) or aerial imagery (A1>RF and A2>RF experiments) as input to the RF classifier resulted in relatively high misclassifications among the *orchard*, *groundnuts*, *maize* and *cotton* classes. *Groundnuts* were mostly misclassified as *cotton* or *orchard* in the S>RF experiment, which was likely due to the similar spectral signatures of these classes (Figure 4.3). However, *maize* was the least misclassified in this experiment, obtaining low errors of commission (0.08%) and omission (0.02%) compared the other classes, despite having a spectral signature similar to *cotton* and *groundnuts*. These low errors were mainly due to *maize* being misclassified as *non agri* only 84 times, while other classes were misclassified as *non agri* at least 4 times more frequently. For the A1>RF experiment, *groundnuts*, *non-agri*, *maize* and *cotton* were frequently confused, while the *orchard* class was the most accurately classified (error of commission of 4.8% and error of omission of 12.3%). The relatively good performance of the VHR aerial imagery for mapping orchards was attributed to the ability of the texture features to represent the structural (spatial) characteristics of this class. Tree crops (mostly pecan nuts and fruit trees in the study area) are usually planted in rows and about 5-10 m apart. These rows are clearly visible in the VHR aerial imagery. In contrast, the resolution of the Sentinel-2 imagery is too low (10 m) to adequately represent the row structure of the orchards. This finding is in agreement with Warner & Steinmaus (2005) who achieved a UA of 97.3% and a PA of 88.7% when classifying *orchard* using VHR imagery.

The non-agricultural (*non-agri*) class is much more heterogeneous than the other classes and as such is expected to have some spectral and structural overlap with the crop type classes. This explains why the *non-agri* class was often misclassified. Nevertheless, the Sentinel-2 data performed relatively well and even outperformed several of the other datasets that included additional features (e.g. texture measures), indicating the potential of Sentinel-2 data for crop type classification. This agrees with Vuolo et al. (2018), who obtained OAs of above 90% when using Sentinel-2 imagery for crop type classification. Similarly, Belgiu & Csillik (2018) created crop type maps using Sentinel-2 data in three test areas and obtained OAs ranging from 75% to 98%.

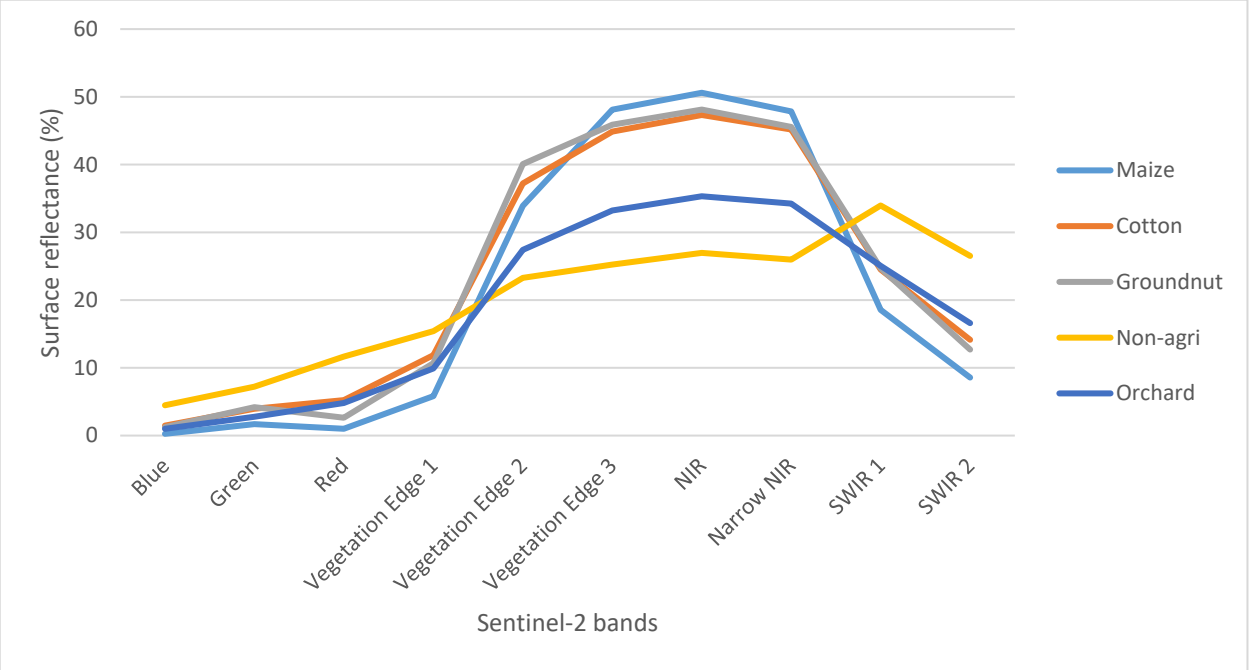


Figure 4.3: Spectral responses of the five crop type classes based on the Sentinel-2 bands considered

Groundnuts obtained the highest error of commission and omission in the A2>RF experiment, which was unexpected due to *groundnuts* being planted in rows, which would have been best represented by the texture features. However, *groundnuts* are planted in the study area with row spacings of 45cm up to 76cm, which is likely too narrow to be adequately depicted by the resolution of the aerial imagery. Similarly, the *maize* and *cotton* crop classes did not benefit from the texture features as they are planted in too narrow rows.

The LiDAR data (L dataset) performed well on its own, despite not having the benefit of spectral information. However, it only performed well when there were substantial height differences between the target crops (e.g. between *orchards* and *groundnuts*). Based on the visual and quantitative analyses it is clear that most misclassifications involving the LiDAR data could be attributed to within-field height variations (canopy gaps or areas with poor crop growth), which causes tall crops (*orchards*) to have similar height values to short crops such as *groundnuts* (around zero height) and intermediately tall crops such as *maize* and *cotton* (0-2 m). Heights of *cotton* and *maize* varied substantially within and among fields across the study area, which resulted in many areas within *maize* fields having the same height as *cotton*. These variations in heights could explain why there were misclassifications between *cotton* and *maize* for the L experiment. For the same reason, the *non-agri* class was the most confused when the LiDAR data were used on its own as it contains land covers (natural vegetation, trees, man-made structures) that have similar heights to many of the crop type classes considered. For instance, because it has a height of close to zero in the LiDAR data, *groundnuts* were most frequently misclassified as short vegetation and bare areas within the *non-agri* class. This observation is in agreement with Mathews & Jensen (2012),

who found that *non-agri* and other crops are often confused for vineyards when only LiDAR data are used as classifier input.

The L-S dataset, which is the combination of the L and S datasets, seemed have to retained the benefits of the two data sources (high OA of 93.5% using the SVM-L classifier). The addition of the L dataset to the S dataset helped to minimise misclassifications among the four crop types (resulting in errors of commission and omission of below 10.6%). A mean OA increase of 7.1% when the LiDAR data were added to the Sentinel-2 image corresponds with other studies (Antonarakis, Richards & Brasington 2008; Chen et al. 2009; Jahan & Awrangjeb 2017; Liu & Bo 2015; Matikainen et al. 2017; Wu et al. 2017) where substantially higher OAs were obtained when LiDAR data were added to spectral data. Similar improvements in accuracy were observed in the A-L experiment, which corresponds well with Chen et al. (2009).

The Sentinel-2 imagery performed the best of all the single-source datasets considered. Although the LiDAR data performed on par with the Sentinel-2 imagery for crop type differentiation, the latter data have the advantage of being regularly updated (once every five days, depending on cloud cover), while LiDAR data are typically updated less frequently (once every few years when obtained using aircraft). Vuolo et al. (2018) showed that crop type classification accuracies increased when multiple Sentinel-2 images, collected over a growing season, were used as input to machine learning classifiers. Consequently, the value of using LiDAR data in combination with multi-temporal Sentinel-2 data may be worth investigating in future work.

The availability of LiDAR data are likely to exponentially increase as satellite-based systems become operational and as LiDAR sensors mounted on UAVs become more common. However, LiDAR data are only recommended for crop type classification if the crop types being classified have sufficient height differences. It is clear from our findings that, if it is available, LiDAR data should be combined with spectral data to improve classifications, especially for differentiating crop types that have similar spectral properties, but are also structurally different (i.e. have different heights).

The aerial imagery (A1 and A2) did not perform as well as the Sentinel and LiDAR datasets. Based on our results, the use of such data for crop type classification is not recommended, especially given that Sentinel-2 data generally performed better and are readily (and more frequently) available.

Although the main focus of the study was not to compare the accuracies of different classification algorithms, we can recommend RF or XGBoost when only LiDAR data are available, while the d-NN or SVM-L algorithms are most suitable for when Sentinel-2 imagery is the only available data

source. When using a combination of LiDAR and Sentinel-2 data, either XGBoost, RF, d-NN or SVM-L performed well with our data.

4.5 CONCLUSION

To our knowledge, this is the first study in which LiDAR data were used on its own as input to machine learning algorithms for differentiating multiple crop types. Experiments involving combinations of aerial, Sentinel-2 and LiDAR data were carried out to assess the impact of combining the different data sources on classification accuracies. It was shown that most crops could be differentiated with LiDAR data on its own, with XGBoost providing the best accuracies (87.8%). The LiDAR data proved particularly useful for differentiating crops with substantial height differences (e.g. *orchards* from *groundnuts*). In general, the classifications in which LiDAR derivatives were used as the only predictor variables were comparable with those in which Sentinel-2 data were used on its own. However, it is clear from the results that the machine learning classifiers were most effective when the different data sources were combined, with the combination of all three sources (i.e. aerial imagery, LiDAR and Sentinel-2 data) providing the highest accuracies (94.6% when RF was used as classifier). Using the aerial imagery on its own produced the lowest accuracies (mean OA of 75.6%).

The findings of this research can aid in solving the real world problem of monitoring crop production, since this research has provided valuable information on crop type classification. The research provided information on which data to use, either on its own or in combination with other data. The study also provides insight into which machine learning algorithms are more effective and robust for crop type classifications, which will likely inform future operational crop type mapping implementations.

CHAPTER 5: DISCUSSION AND CONCLUSION

This chapter summarises and critically evaluates the findings of the research. The aims and objectives are revisited in the first section, while the section that follows outlines the findings of the two main experiments (Chapters 3 and 4). In Section 5.3 the limitations of the research are discussed and recommendation for future work made, followed by conclusions in the final section.

5.1 REVISITING THE AIM AND OBJECTIVES

This study aimed to evaluate a method of classifying crops with LiDAR data, multispectral imagery and machine learning. The study was motivated by the importance of using accurate crop type maps for regional crop analysis such as crop yield and water use estimations. LiDAR is an active sensor that provides detailed elevation point clouds, which can be used to derive DSMs, DTMs and intensity rasters. LiDAR data are commonly used as an additional feature for land cover classification, but have also been used as an additional feature for crop type classification (Liu & Bo 2015). LiDAR data add an additional dimension to imagery by providing height information (vegetation height). Furthermore, by analysing the brightness, angular position, change in frequency and the timing of the LiDAR pulses, the structure of the terrain and vegetation can be obtained, something that is not possible with conventional optical sensors. The increasing availability of LiDAR data, as well as the existence of the GEDI satellite (successfully launched in late 2018) and ICESat-2 (mostly used for monitoring sea ice but can provide detailed elevation measurements of land surfaces) will likely lead to more frequent use of LiDAR data for crop type classification.

The literature review (Objective 1) was presented in Chapter 2 and focussed on the remote sensing data and methods used for crop type mapping. The research showed that crop type classification has been performed by various studies and that satisfactory accuracies have been obtained by using multispectral, hyperspectral and SAR data. However, the literature review revealed a knowledge gap when it came to the value of LiDAR data for crop type classification. LiDAR data were used as an additional feature in studies that employed it. An nDSM is the most common LiDAR derivative used, while DSMs and DTMs have also been employed. The literature review showed that texture measures performed on LiDAR derivatives, specifically the nDSM, hold much potential for crop type differentiation (Jahan & Awrangjeb 2017; Liu & Bo 2015).

Two main experiments were carried out in this study. The first experiment focussed on the use of LiDAR data for mapping vineyards, while the second experiment applied the methods of the first experiment for differentiating several crops and testing different combinations of Earth observation data (satellite imagery, aerial photographs and LiDAR data).

The data collection (Objective 2) was accomplished by obtaining LiDAR data for the first experiment from the City of Cape Town. The LiDAR data and aerial imagery used in the second experiment were obtained from the Northern Cape Department of Agriculture, Land Reform and Rural Development, while the Sentinel-2 data (second experiment) was obtained from ESA.

For Objective 3, an nDSM and intensity raster were derived from the LiDAR data. Texture measures were applied on both of these derivatives. Different resolutions and window sizes were applied to assess their value for classifying crop types. Four different resolutions (1.5 m, 2 m, 2.5 m, 3 m) were employed for the LiDAR derivatives and two window sizes (3x3 and 5x5) were used for the texture measures. The derivatives were then used as input to 11 machine learning classification algorithms. The results (Chapter 3) showed that the larger (5x5) window size obtained the highest accuracy (80.9%), regardless of the resolution of the LiDAR derivatives.

In Chapter 4, Sentinel-2 and aerial imagery were compared and combined with LiDAR data in order to address Objective 5. The datasets were used as input for the classification algorithms on their own and in different combinations. The two datasets that obtained the highest mean OAs were a combination of LiDAR data and Sentinel-2 (91.2%), and a combination of LiDAR data, Sentinel-2 and aerial imagery (91.9%). These two combinations of datasets obtained similar OAs and were not significantly different.

Objective 4 was formulated to evaluate a range of machine learning classification algorithms (RF, DT, k-NN, NN, d-NN, LR, NB, SVM L, SVM RBF and XGBoost) for the use of multiple crop type classifications. Overall, the RF and XGBoost algorithms performed the best in both experiments. RF obtained the highest OAs, namely 80.9% (Experiment 1) and 94.6% (Experiment 2). Chapters 3 and 4 both contributed to Objective 4.

The results from Chapters 3 and 4 were used to address Objective 6 (see Section 5.3 for details).

5.2 FINDINGS OF THE RESEARCH

Chapter 3 demonstrated that vineyards can be successfully (80.4%) classified at a regional scale using LiDAR data and machine learning algorithms. Furthermore, it was shown that a window size of 5x5 pixels for texture measures resulted in the highest accuracy regardless of the resolution used for the LiDAR derivatives (nDSM and intensity raster). The four resolutions (1.5 m, 2 m, 2.5 m and 3 m) with a window size of 5x5 applied for the texture measures showed no significant statistical difference in OA. In contrast, the four resolutions with a window size of 3x3 applied for the texture measures showed a significant statistical difference in OA. The results indicated that, when using LiDAR data, the window size is more important than resolution when classifying vineyards.

The results obtained in Chapter 3 showed agreement with Mathews & Jenson (2012) who also used LiDAR data for classifying vineyards. However, they performed an unsupervised classification on a smaller study area (7.8 km^2) with a substantially lower number of vineyards, did not perform texture measures on the LiDAR data, and only used an area-based accuracy assessment. The unsupervised classification consisted of six classes, which had to be manually labelled as either vineyards or non-vineyards, whereas the supervised classifications implemented in this study performed this function inherently. Furthermore, based a visual (qualitative) inspection, the results in Chapter 3 appear to have classified non-vineyards more accurately than in Mathews & Jenson (2012). Despite the differences between the two studies, both studies agree that LiDAR is an effective data source for discriminating vineyards from other land cover/use classes as both studies were able to obtain accuracies of above 80%.

Chapter 4 used the methods identified in Chapter 3 to evaluate how well different datasets and dataset combinations can differentiate multiple crop types. Five classes were considered, four of which were crop types (maize, cotton, groundnuts and orchards), while the fifth class represented non-agricultural land cover/use. The datasets consisted of LiDAR data, a Sentinel-2 image and aerial photographs. Every possible combination (seven in total) of the datasets was assessed. The combination of all three datasets obtained the highest mean OA of 91.9%, but similar OAs were obtained by combining LiDAR data and Sentinel-2 imagery (mean OA of 91.2%). OAs of 84.3% and 83.6% were respectively obtained when the LiDAR and Sentinel-2 datasets were used on their own. The aerial imagery (on its own) obtained the lowest OA out of all the datasets, with a mean OA of 75.6%. These results showed that LiDAR data can match or outperform spectral data for differentiating crop types. This finding corresponds with those of other studies (Jahan & Awrangjeb 2017; Liu & Bo 2015; Matikainen et al. 2017), which showed that the combination of height data (LiDAR) and spectral data (Sentinel-2 or aerial imagery) leads to a better outcome compared to when using spectral data alone. However, the fact that LiDAR data on its own can perform as well or better than spectral data sources is a novel finding.

The machine learning classification algorithms performed similarly in both experiments, with RF and XGBoost generally producing the best results. However, unlike in Experiment 1 (Chapter 3), the variation between the OAs of the three best performing classifiers (RF, XGBoost and NN) in Experiment 2 (Chapter 4) was statistically significant. The results from Chapters 3 and 4 suggest that RF is robust and can effectively classify LiDAR data, either on its own or in combination with spectral data. These results are in agreement with Ma et al. (2017) and Khatami, Mountrakis & Stehman (2016) who performed a review of several studies and concluded that RF generally outperforms other machine learning classifiers. A major contribution of the current study is that

performing the zero mean and unit variance standardisation on the LiDAR data improved the performance of all the classifiers, with the exception of RF, DT and XGBoost.

In summary, the results from this research provided new insights into how well LiDAR data can classify crops by using machine learning classification algorithms. Furthermore, the study showed how LiDAR data perform compared to spectral data (Sentinel-2 imagery and aerial photographs) and how LiDAR data perform in combination with these data sources.

5.3 LIMITATIONS AND RECOMMENDATIONS

This study provides a good foundation for LiDAR-based crop type classification. However, the study was carried out in two areas where the crops had either distinct planting patterns or different heights. Future research should investigate the performance of LiDAR data in classifying a wider variety of crops with similar heights and planting patterns.

The image analysis in this study used a per-pixel approach, which resulted in a noticeable salt-and-pepper effect in classified maps. Although this effect can be reduced with post-processing (e.g. majority filter), an alternative is to use a GEOBIA approach. Many studies have shown that accuracies can be improved when a GEOBIA approach (instead of a per-pixel approach) is used. However, GEOBIA does add another level of complexity, in particular regarding the selection of segmentation parameters (Gilbertson, Kemp & Van Niekerk 2017). Research is needed to investigate whether a GEOBIA approach will outperform the per-pixel approach applied in this study, and if it does, whether the improvements in accuracies are worth the added complexity.

Another limitation of using LiDAR data is the cost of acquisition (aerial surveys), which makes frequent surveys to increase the data's temporal resolution unfeasible for many applications. In contrast, multispectral imagery captured by the Sentinel-2 constellation provides high spatial resolution imagery at high temporal frequencies (five-day interval). Other crop type mapping studies have shown increases in OAs when a multi-temporal approach is used (Gilbertson, Kemp & Van Niekerk 2017; Vuolo et al. 2018). Future research should analyse how LiDAR data compare to a multi-temporal image analyses using Sentinel-2 data and whether LiDAR data can improve the OAs of multi-temporal approaches.

The machine learning classification algorithms used in this study were configured to use the default parameters for all the classifications. Parameter optimisation for the best performing classifiers (RF, XGBoost, NN and SVM) should be investigated in future research in order to assess if the OAs will increase and, if they do, whether the increases are worth the added effort and computational costs.

5.4 CONCLUSIONS

The research presented in this thesis aimed to evaluate the use of LiDAR data and machine learning for crop type classification. In addition, it evaluated the effects of combining LiDAR data with spectral imagery (Sentinel-2 and aerial imagery). The thesis classified crops in two areas, namely the City of Cape Town (Chapter 3) and the Vaalharts irrigation scheme (Chapter 4). The aim of the thesis was achieved by six objectives listed in Section 1.7. All of the objectives were met.

Four main conclusions can be drawn from the results of this research. First, the window size used for the texture measures had a bigger impact on the OAs than the resolution of the LiDAR derivatives, with larger window sizes resulting in higher OAs. Second, the RF and XGBoost classifiers were the best performing classifiers. In addition, RF and XGBoost also proved to be the most robust classifiers as, unlike the rest of the classifiers (excluding DT), they did not require the data to be standardised. Third, LiDAR data performed well for classifying crop types, and in most cases its performance equalled that of the Sentinel-2 imagery when the crops had either a distinct planting pattern (i.e. rows) or where the height values of the different crops varied. Fourth, the combination of LiDAR data, Sentinel-2 and aerial imagery produced the highest OA (94.6%) in Experiment 2. However, the combination of LiDAR and Sentinel-2 imagery produced similar results and had the added benefit of requiring less processing time and storage space. Although computation efficiencies were not a focus of this research, they should be considered for operational implementations involving large datasets.

The findings of this study provide new insight into the value of LiDAR data (on its own and in combination with spectral data) for crop type classifications. Accurate crop type maps are invaluable for regional crop analyses. With LiDAR data becoming increasingly available, it is likely that such data will become a primary data source in future crop type mapping workflows. Hopefully the findings of this research will contribute to more efficient management of South Africa’s limited agricultural resources and improve food security in the future.

REFERENCES

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y & Zheng X 2016. TensorFlow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*: 265–284.

Adejuwon A & Mosavi A 2010. Domain Driven Data Mining – Application to Business. *Journal of Computer Science* 7, 4: 2–5.

Al-doski J, Mansor SB, Zulhaidi H & Shafri M 2013. Image Classification in Remote Sensing. *Journal of Environment and Earth Science* 3, 10: 141–147.

Antonarakis AS, Richards KS & Brasington J 2008. Object-based land cover classification using airborne LiDAR. *Remote Sensing of Environment* 112, 6: 2988–2998.

Arvor D, Durieux L, Andrês S & Laporte MA 2013. Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing* 82: 125–137.

Bannari A, Pacheco A, Staenz K, McNairn H & Omari K 2006. Estimating and mapping crop residues cover on agricultural lands using hyperspectral and IKONOS data. *Remote Sensing of Environment* 104, 4: 447–459.

Bargiel D 2017. Remote Sensing of Environment A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment* 198: 369–383.

Bauer ME, Cipra JE, Anuta PE & Etheridge JB 1979. Identification and area estimation of agricultural crops by computer classification of LANDSAT MSS data. *Remote Sensing of Environment* 8, 1: 77–92.

Belgiu M & Csillik O 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment* 204, January 2017: 509–523.

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

Bendig J, Bolten A & Bareth G 2013. UAV-based Imaging for Multi-Temporal, very high

Resolution Crop Surface Models to monitor Crop Growth Variability. *Photogrammetrie - Fernerkundung - Geoinformation* 2013, 6: 551–562.

Bietresato M, Carabin G, Vidoni R, Gasparetto A & Mazzetto F 2016. Evaluation of a LiDAR-based 3D-stereoscopic vision system for crop-monitoring applications. *Computers and Electronics in Agriculture* 124: 1–13.

Blaes X, Vanhalle L & Defourny P 2005. Efficiency of crop identification based on optical and SAR image time series. *Remote Sensing of Environment* 96, 3–4: 352–365.

Blaschke T 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 1: 2–16.

Blaschke T, Hay GJ, Kelly M, Lang S, Hofmann P, Addink E, Queiroz Feitosa R, van der Meer F, van der Werff H, van Coillie F & Tiede D 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 180–191.

Bolton DK & Friedl MA 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology* 173: 74–84.

Bork EW & Su JG 2007. Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. *Remote Sensing of Environment* 111, 1: 11–24.

Bradley AP & Bradley AP 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 7: 1145–1159.

Brennan R & Webster TL 2006. Object-oriented land cover classification of lidar-derived surfaces. *Canadian Journal of Remote Sensing* 32, 2: 162–172.

Bujan S, Gonzalez-Ferreiro E, Reyes-Bueno F, Barreiro-Fernandez L, Crecente R & Miranda D 2012. Land use classification from lidar data and ortho-images in a rural area. *The Photogrammetric Record* 27, 140: 401–422.

Campbell JB & Wynne RH 2011. *Introduction to Remote Sensing*. Fifth Edit. New York: The Guilford Press.

Castelluccio M, Poggi G, Sansone C, Verdoliva L & Aug C V 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *arXiv preprint*: 1–11.

Charaniya AP, Manduchi R & Lodha SK 2004. Supervised parametric classification of aerial LiDAR Data. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*: 30–30.

Chen T, He T, Benesty M, Khotilovich V & Tang Y 2017. Package xgboost - R implementation.

Chen X, Wang D, Chen J, Wang C & Shen M 2018. The mixed pixel effect in land surface phenology: A simulation study. *Remote Sensing of Environment* 211, June 2017: 338–344.

Chen Y, Su W, Li J & Sun Z 2009. Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research* 43, 7: 1101–1110.

Cheney M 2001. A mathematical tutorial on synthetic aperture radar. SIAM Rev. 43(2), 301-312. *SIAM Review* 43, 2: 301–312.

Chica-Olmo M & Abarca-Hernández F 2000. Computing geostatistical image texture for remotely sensed data classification. *Computers & Geosciences* 26, 4: 373–383.

Chirici G, Mura M, McInerney D, Py N, Tomppo EO, Waser LT, Travaglini D & McRoberts RE 2016. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. *Remote Sensing of Environment* 176: 282–294.

Clausi DA 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing* 28, 1: 45–62.

Conrad C, Fritsch S, Zeidler J, Rücker G & Dech S 2010. Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data. *Remote Sensing* 2, 4: 1035–1056.

Cousins SR, Witkowski ETF, Pfab MF, Riddles RE & Mycock DJ 2013. Reproductive ecology of Aloe plicatilis, a fynbos tree aloe endemic to the Cape Winelands, South Africa. *South African Journal of Botany* 87: 52–65.

Dadhwal VK, Singh RP, Dutta S & Parihar JS 2002. Remote sensing based crop inventory : A review of Indian experience. 43, 1: 107–122.

Dekker RJ 2003. Texture analysis and classification of ERS SAR images for map updating of urban areas in the Netherlands. *IEEE Transactions on Geoscience and Remote Sensing* 41, 9 PART I: 1950–1958.

Delenne C, Durrieu S, Rabatel G & Deshayes M 2010. From pixel to vine parcel: A complete methodology for vineyard delineation and characterization using remote-sensing data. *Computers and Electronics in Agriculture* 70, 1: 78–83.

Dell'Acqua F, Iannelli GC, Torres MA & Martina MLV 2018. A novel strategy for very-large-scale cash-crop mapping in the context of weather-related risk assessment, combining global satellite multispectral datasets, environmental constraints, and in situ acquisition of

geospatial data. *Sensors (Switzerland)* 18, 2.

Doluschitz R & Schmisseur WE 1988. Expert Systems: Applications to Agriculture and Farm Management. *Computers and Electronics in Agriculture Elsevier Science Publishers B.V* 2: 173–182.

Donoghue DNM, Watt PJ, Cox NJ & Wilson J 2007. Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment* 110, 4: 509–522.

Duan Z, Zhao D, Zeng Y, Zhao Y, Wu B & Zhu J 2015. Assessing and Correcting Topographic Effects on Forest Canopy Height Retrieval Using Airborne LiDAR Data. *Sensors* 15, 6: 12133–12155.

Duro DC, Franklin SE & Dub?? MG 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* 118: 259–272.

Eastman JR 2006. Guide to GIS and Image Processing. *Clark University: Worcester, MA, USA* 1, May: 87–131.

Escobar V & Brown M 2014. ICESat--2 Applications Vegetation Tutorial with Landsat 8 Report [online]. Available from: http://icesat.gsfc.nasa.gov/icesat2/applications/ICESat2_Vegetation_Report_FINAL.pdf

Estrada J, Sánchez H, Hernanz L, Checa M & Roman D 2017. Enabling the Use of Sentinel-2 and LiDAR Data for Common Agriculture Policy Funds Assignment. *ISPRS International Journal of Geo-Information* 6, 8: 255.

Etoughe Kongo UP 2015. Urban land cover classification from high resolution geoeye-1 imagery using a lidar based digital surface model.

Evans C, Jones R, Svalbe I & Berman M 2002. Segmenting multispectral landsat TM images into field units. *IEEE Transactions on Geoscience and Remote Sensing* 40, 5: 1054–1064.

Fiorillo E, Crisci A, De Filippis T, Di Gennaro SF, Di Blasi S, Matese A, Primicerio J, Vaccari FP & Genesio L 2012. Airborne high-resolution images for grape classification: Changes in correlation between technological and late maturity in a Sangiovese vineyard in Central Italy. *Australian Journal of Grape and Wine Research* 18, 1: 80–90.

Fountas S, Carli G, Sørensen CG, Tsiropoulos Z, Cavalaris C, Vatsanidou A, Liakos B, Canavari M, Wiebensohn J & Tisserye B 2015. Farm management information systems: Current situation and future perspectives. *Computers and Electronics in Agriculture* 115: 40–50.

Friedl M a. MA & Brodley CECE 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* 61, 3: 399–409.

Gaigher R & Samways MJ 2010. Surface-active arthropods in organic vineyards, integrated vineyards and natural habitat in the Cape Floristic Region. *Journal of Insect Conservation* 14, 6: 595–605.

Geerling GW, Labrador-Garcia M, Clevers JGPW, Ragas a. MJ & Smits a. JM 2007. Classification of floodplain vegetation by data fusion of spectral (CASI) and LiDAR data. *International Journal of Remote Sensing* 28, 19: 4263–4284.

Georganos S, Grippa T, Vanhuysse S, Lennert M, Shimoni M & Wolff E 2018. Very High Resolution Object-Based Land Use – Land Extreme Gradient Boosting. *IEEE Geoscience and Remote Sensing Letters* 15, 4: 607–611.

Ghariani K, Chehata N, Le Bris A & Lagacherie P 2014. Agricultural field delimitation using active learning and random forests margin. *IEEE Geoscience and Remote Sensing Symposium* , d: 1717–1720.

Gilbertson JK, Kemp J & Van Niekerk A 2017. Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. *Computers and Electronics in Agriculture* 134: 151–159.

Gilbertson JK & Van Niekerk A 2017. Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronics in Agriculture* 142: 50–58.

Gong P, Mahler SA, Biging GS & Newburn DA 2003. Vineyard identification in an oak woodland landscape with airborne digital camera imagery. *International Journal of Remote Sensing* 24, 6: 1303–1315.

Goodfellow I, Bengio Y & Courville A 2016. *Deep learning*. MIT Press.

Greyling J 2015. A look at the contribution of the agricultural sector to the South African economy [online]. Available from: http://www.grainsa.co.za/a-look-at-the-contribution-of-the-agricultural-sector-to-the-south-african-economy [Accessed 16 April 2018].

Hall A, Louis J & Lamb D 2003. Characterising and mapping vineyard canopy using high-spatial- resolution aerial multispectral images. *Computers and Geosciences* 29, 7: 813–822.

Hämmerle M & Höfle B 2014. Effects of reduced terrestrial LiDAR point density on high-resolution grain crop surface models in precision agriculture. *Sensors (Switzerland)* 14, 12: 24212–24230.

Han M, Zhu X & Yao W 2012. Neurocomputing Remote sensing image classification based on neural network ensemble algorithm. *Neurocomputing* 78, 1: 133–138.

Hartfield KA, Landau KI & van Leeuwen WJD 2011. Fusion of high resolution aerial multispectral and lidar data: Land cover in the context of urban mosquito habitat. *Remote Sensing* 3, 11: 2364–2383.

Heydari SS & Mountrakis G 2018. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. *Remote Sensing of Environment* 204, October 2017: 648–658.

Hoekman DH, Vissers MAM & Tran TN 2011. Unsupervised Full-Polarimetric SAR Data Segmentation as a Tool for Classification of Agricultural Areas. 4, 2: 402–411.

Howard B 2015. LIDAR and its use in agriculture. *Agricultural Innovation Submisstion 27* , 9.

Hsu C-W, Chang C-C & Lin C-J 2008. A Practical Guide to Support Vector Classification. *BJU international* 101, 1: 1396–400.

Hubert-Moy L, Cotonnec A, Du L Le, Chardin A & Perez P 2001. A Comparison of Parametric Classification Procedures of Remotely Sensed Data Applied on Different Landscape Units. *Remote Sensing of Environment* 4257, 75: 174–187.

Immitzer M, Vuolo F & Atzberger C 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing* 8, 3.

Inglada J, Arias M, Tardy B, Hagolle O, Valero S, Morin D, Dedieu G, Sepulcre G, Bontemps S, Defourny P & Koetz B 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing* 7, 9: 12356–12379.

Ishida T, Itagaki S, Sasaki Y & Ando H 2004. Application of wavelet transform for extracting edges of paddy fields from remotely sensed images. *International Journal of Remote Sensing* 25, 2: 347–357.

Ismail Z, Khanan MFA, Omar FZ, Rahman MZA & Salleh MRM 2016. EVALUATING ERROR OF LIDAR DERIVED DEM INTERPOLATION FOR VEGETATION AREA. XLII, October: 3–5.

Jahan F & Awrangjeb M 2017. Pixel-based land cover classification by fusing hyperspectral and LiDAR data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42, 2W7: 711–718.

Jang JD, Payan V, Viau AA & Devost A 2008. The use of airborne lidar for orchard tree

inventory. *International Journal of Remote Sensing* 29, 6: 1767–1780.

Jonsson K, Kittler J, Li Y. & Matas J 2002. Support vector machines for face authentication. *Image and Vision Computing* 20, 4: 369–375.

Khanal S, Fulton J & Shearer S 2017. An overview of current and potential applications of thermal remote sensing in precision agriculture. *Computers and Electronics in Agriculture* 139: 22–32.

Khatami R, Mountrakis G & Stehman S V. 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment* 177: 89–100.

Kotsiantis S. 2007. Supervised Machine Learning: A Review of Classification Techniques. *Emerging Artificial Intelligence Applications in Computer Engineering* 160.

Kussul N, Lavreniuk M, Skakun S & Shelestov A 2017. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters* 14, 5: 778–782.

Lary DJ, Alavi AH, Gandomi AH & Walker AL 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* 7, 1: 3–10.

Lee H, Slatton KC, Roth BE & Cropper WP 2010. Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. *International Journal of Remote Sensing* 31, 1: 117–139.

Lee S 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing* 26, 7: 1477–1491.

Lee WS, Alchanatis V, Yang C, Hirafuji M, Moshou D & Li C 2010. Sensing technologies for precision specialty crop production. *Computers and Electronics in Agriculture* 74, 1: 2–33.

Li Q, Wang C, Zhang B & Lu L 2015. Object-Based Crop Classification with Landsat-MODIS Enhanced Time-Series Data. , November: 16091–16107.

Li W, Guo Q, Jakubowski MK & Kelly M 2012. A New Method for Segmenting Individual Trees from the Lidar Point Cloud. *Photogrammetric Engineering & Remote Sensing* 78, 1: 75–84.

Li Y & Cheng B 2009. An Improved k-Nearest Neighbor Algorithm and Its Application to High Resolution Remote Sensing Image Classification. *Geoinformatics*: 1–4.

Liaqat MU, Cheema MJM, Huang W, Mahmood T, Zaman M & Khan MM 2017. Evaluation of MODIS and Landsat multiband vegetation indices used for wheat yield estimation in irrigated Indus Basin. *Computers and Electronics in Agriculture* 138: 39–47.

Lim K, Treitz P, Wulder M, St-Onge B & Flood M 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography* 27, 1: 88–106.

Lin SW, Ying KC, Chen SC & Lee ZJ 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications* 35, 4: 1817–1824.

Liu X, Zhang Z, Peterson J & Chandra S 2007. The effect of LiDAR data density on DEM accuracy. *MODSIM07 - Land, Water and Environmental Management: Integrated Systems for Sustainability, Proceedings*: 1702–1708.

Liu X & Bo Y 2015. Object-based crop species classification based on the combination of airborne hyperspectral images and LiDAR data. *Remote Sensing* 7, 1: 922–950.

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.

Longley PA, Goodchild MF, Maguire DJ & Rhind DW 2005. *Geographic Information Systems & Science*. Thrid. Jefferson City: John Wiley & Sons.

Ma L, Li M, Ma X, Cheng L, Du P & Liu Y 2017. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 277–293.

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 & Jensen JLR 2012. An airborne LiDAR-based methodology for vineyard parcel detection and delineation. *International Journal of Remote Sensing* 33, 16: 5251–5267.

Matikainen L, Karila K, Hyyppä J, Litkey P, Puttonen E & Ahokas E 2017. Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating. *ISPRS Journal of Photogrammetry and Remote Sensing* 128: 298–313.

Matton N, Canto GS, Waldner F, Valero S, Morin D, Inglada J, Arias M, Bontemps S, Koetz B & Defourny P 2015. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. *Remote Sensing* 7, 10: 13208–13232.

Mattupalli C, Moffet C, Shah K & Young C 2018. Supervised Classification of RGB Aerial Imagery to Evaluate the Impact of a Root Rot Disease. *Remote Sensing* 10, 6: 917.

McCarley TR, Kolden CA, Vaillant NM, Hudak AT, Smith AMS, Wing BM, Kellogg BS & Kreitler J 2017. Multi-temporal LiDAR and Landsat quantification of fire-induced changes to forest structure. *Remote Sensing of Environment* , 191: 419–432.

Mcnairn H & Brisco B 2004. The application of C-band polarimetric SAR for agriculture : a review. 30, 3: 525–542.

Mcnairn H, Shang J, Jiao X & Champagne C 2009. The Contribution of ALOS PALSAR Multipolarization and Polarimetric Data to Crop Classification. 47, 12: 3981–3992.

Melgani F & Blanzieri E 2008. Nearest neighbor classification of remote sensing images with the maximal margin principle. *IEEE Transactions on Geoscience and Remote Sensing* 46, 6: 1804–1811.

Miller DM, Kaminsky EJ & Rana S 1995. Neural network classification of remote-sensing data. *Computers and Geosciences* 21, 3: 377–386.

Mo X, Liu S, Lin Z & Guo R 2009. Regional crop yield, water consumption and water use efficiency and their responses to climate change in the North China Plain. *Agriculture, Ecosystems and Environment* 134, 1–2: 67–78.

Mogala M 2012. A profile of the South African table grape market value chain. *Department of Agriculture Forestry and Fisheries Republic of South Africa*: 1–52.

Mohabeer H, Soyjaudah KMS & Pavaday N 2011. *Enhancing The Performance Of Neural Network Classifiers Using Selected Biometric Features*. SENSORCOMM 2011, The fifth International Conference on Sensors Technologies and Applications: 140–144.

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.

Möller M, Alchanatis V, Cohen Y, Meron M, Tsipris J, Naor A, Ostrovsky V, Sprintsin M & Cohen S 2007. Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. *Journal of Experimental Botany* 58, 4: 827–838.

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

Muller SJ & Van Niekerk A 2016. International Journal of Applied Earth Observation and

Geoinformation An evaluation of supervised classifiers for indirectly detecting salt-affected areas at irrigation scheme level. *International Journal of Applied Earth Observations and Geoinformation* 49: 138–150.

Myburgh PA 2015. *Irrigation management in table grape vineyards-South African experience*.

Myint SW, Gober P, Brazel A, Grossman-Clarke S & Weng Q 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment* 115, 5: 1145–1161.

Nel AA & Lamprecht SC 2011. Crop rotational effects on irrigated winter and summer grain crops at vaalharts. *South African Journal of Plant and Soil* 28, 2: 127–133.

Nogueira K, Penatti OAB & dos Santos JA 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition* 61: 539–556.

O'Neil-Dunne J, MacFaden SW, Royar AR & Pelletier KC 2012. An object-based system for LiDAR data fusion and feature extraction. *Geocarto International* 6049, April 2013: 1–16.

Ortiz MJ, Formaggio AR & Epiphanio JCN 1997. Classification of croplands through integration of remote sensing, GIS, and historical database. *International Journal of Remote Sensing* 18, 1: 95–105.

Pacifici F, Chini M & Emery WJ 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment* 113, 6: 1276–1292.

Pádua L, Adão T, Hruška J, Sousa JJ, Peres E, Morais R & Sousa A 2017. Very high resolution aerial data to support multi-temporal precision agriculture information management. *Procedia Computer Science* 121: 407–414.

Pandey A & Mishra A 2017. Application of artificial neural networks in yield prediction of potato crop. *Russian Agricultural Sciences* 43, 3: 266–272.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Louppe G, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M & Duchesnay É 2012. Scikit-learn: Machine Learning in Python. *Journal of machine learning research* 12: 2825–2830.

Pedroso M, Taylor J, Tisseyre B, Charnomordic B & Guillaume S 2010. A segmentation algorithm for the delineation of agricultural management zones. *Computers and Electronics in Agriculture* 70, 1: 199–208.

Pelletier C, Valero S, Inglada J, Champion N & Dedieu G 2016. Remote Sensing of

Environment Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment* 187: 156–168.

Peña-Barragán JM, Ngugi MK, Plant RE & Six J 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment* 115, 6: 1301–1316.

Pohl C & Van Genderen JL 1998. Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing* 19, 5: 823–854.

Pontius RG, Millones M, Pontius, Robert, Gilmore J, Millones M, Pontius RG & Millones M 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, 15: 4407–4429.

Popescu SC 2011. Lidar Remote Sensing. In *Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications*, 76–80.

Pradhan B 2010. Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. *Advances in Space Research* 45, 10: 1244–1256.

Prins AJ & Niekerk A Van 2020. Geo-spatial Information Science Crop type mapping using LiDAR , Sentinel-2 and aerial imagery with machine learning algorithms. *Geo-spatial Information Science* 00, 00: 1–13.

Qian Y, Zhou W, Yan J, Li W & Han L 2015. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing* 7, 1: 153–168.

De Rainville FM, Durand A, Fortin FA, Tanguy K, Maldague X, Panneton B & Simard MJ 2014. Bayesian classification and unsupervised learning for isolating weeds in row crops. *Pattern Analysis and Applications* 17, 2: 401–414.

Rajan N & Maas SJ 2009. Mapping crop ground cover using airborne multispectral digital imagery. *Precision Agriculture* 10, 4: 304–318.

Rutkowski L, Jaworski M, Pietruczuk L & Duda P 2014. The CART decision tree for mining data streams. *Information Sciences* 266: 1–15.

Rydberg A & Borgefors G 2001. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote*

Sensing 39, 11: 2514–2520.

Sankaran S, Khot LR, Espinoza CZ, Jarolmasjed S, Sathuvalli VR, Vandemark GJ, Miklas PN, Carter AH, Pumphrey MO, Knowles RRN & Pavek MJ 2015. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy* 70: 112–123.

Sankey T, Donager J, McVay J & Sankey JB 2017. UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. *Remote Sensing of Environment* 195: 30–43.

Sasaki T, Imanishi J, Ioki K, Morimoto Y & Kitada K 2012. Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data. *Landscape and Ecological Engineering* 8, 2: 157–171.

Satale DM & Kulkarni MN 2003. LiDAR in Mappingo [online]. Available from: https://www.geospatialworld.net/article/lidar-in-mapping/

Senthilnath J, Kulkarni S, Benediktsson JA & X-S. Y 2016. A Novel Approach for Multi-Spectral Satellite Image Classification Based on the Bat Algorithm. *IEEE Geoscience and Remote Sensing Letters* 13, 4: 599–603.

Senturk S, Bagis S & Berk Ustundag B 2014. Application of remote sensing techniques in locating dry and irrigated farmland parcels. *2014 The 3rd International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2014*.

Shalabi L Al & Shaaban Z 2006. *Normalization as a preprocessing engine for data mining and the approach of preference matrix*. International Conference on Dependability of Computer Systems: 207–214.

Shalabi L Al, Shaaban Z & Kasasbeh B 2006. Data Mining: A Preprocessing Engine. *Journal of Computer Science* 2, 9: 735–739.

Sheldon MR, Fillyaw MJ & Thompson WD 1996. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy research international : the journal for researchers and clinicians in physical therapy* 1, 4: 221–228.

Siachalou S, Mallinis G & Tsakiri-Strati M 2015. A hidden markov models approach for crop classification: Linking crop phenology to time series of multi-sensor remote sensing data. *Remote Sensing* 7, 4: 3633–3650.

Simonneaux V, Duchemin B, Helson D, Er-Raki S, Olioso A & Chehbouni AG 2008. The use of high-resolution image time series for crop classification and evapotranspiration estimate

over an irrigated area in central Morocco. *International Journal of Remote Sensing* 29, 1: 95–116.

Singh KK, Vogler JB, Shoemaker DA & Meentemeyer RK 2012. LiDAR-Landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing* 74: 110–121.

Solares C, Mar ANA & Sanz IA 2005. Bayesian Network Classifiers . An Application to Remote Sensing Image Classification. 2005: 62–67.

Solberg AHS, Jain AK & Taxt T 1994. Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 32, 4: 768–778.

Sonobe R, Tani H & Wang X 2017. An experimental comparison between KELM and CART for crop classification using Landsat-8 OLI data. *Geocarto International* 6049: 1–11.

Tatsumi K, Yamashiki Y, Canales Torres MA & Taipe CLR 2015. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture* 115: 171–179.

Tehrany MS, Pradhan B & Jebuv MN 2014. A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using SPOT 5 imagery. *Geocarto International* 29, 4: 351–369.

Tibane E 2016. Agriculture, Forestry and Fisheries. In *Pocket guide to South Africa 2015/16*, 13–20.

Turker M & Kok EH 2013. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS Journal of Photogrammetry and Remote Sensing* 79: 106–121.

Turker M & Ozdarici A 2011. Field-based crop classification using SPOT4, SPOT5, IKONOS and QuickBird imagery for agricultural areas: a comparison study. *International Journal of Remote Sensing* 32, 24: 9735–9768.

Ulaby FT, Li RY & Shanmugan KS 1982. Crop Classification Using Airborne Radat and Landsat Data. *IEEE Transactions on Geoscience and Remote Sensing* GE-20, 1.

Valero S, Morin D, Inglada J, Sepulcre G, Arias M, Hagolle O, Dedieu G, Bontemps S, Defourny P & Koetz B 2016. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. *Remote Sensing* 8, 1: 1–21.

Van Niekerk A, Jarmain C, Goudriaan R, Muller SJ, Ferreira F, Munch Z, Pauw T, Stephenson

G & Gibson L 2018. *An earth observation apporach towards mapping irrigated area and quantifying water use by irrigated crops in South Africa (TT 745/17)*.

Van Niel TG & McVicar TR 2004. Determining temporal windows for crop discrimination with remote sensing: A case study in south-eastern Australia. *Computers and Electronics in Agriculture* 45, 1–3: 91–108.

Vega FA, Ramírez FC, Saiz MP & Rosúa FO 2015. Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. *Biosystems Engineering* 132: 19–27.

Vuolo F, Neuwirth M, Immitzer M, Atzberger C & Ng WT 2018. How much does multi-temporal Sentinel-2 data improve crop type classification? *International Journal of Applied Earth Observation and Geoinformation* 72, July: 122–130.

Van Vuuren L 2010. Vaalharts - A garden in the desert. *Water Wheel* 9, 1: 20–24.

Waldhoff G, Lussem U & Bareth G 2017. Multi-Data Approach for remote sensing-based regional crop rotation mapping: A case study for the Rur catchment, Germany. *International Journal of Applied Earth Observation and Geoinformation* 61: 55–69.

Warner TA & Steinmaus K 2005. Spatial Classification of Orchards and Vineyards with High Spatial Resolution Panchromatic Imagery. *Photogrammetric Engineering & Remote Sensing* 71, 2: 179–187.

Weinberger K, Blitzer J & Saul L 2006. Distance metric learning for large margin nearest neighbor classification. *Advances in neural information processing systems* 18: 1473.

Whetton R, Zhao Y, Shaddad S & Mouazen AM 2017. Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI. *Computers and Electronics in Agriculture* 138: 127–136.

Whiteside TG, Boggs GS & Maier SW 2011. Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation* 13, 6: 884–893.

Wines of South Africa 2018. Statistics [online]. Available from: https://www.wosa.co.za/The-Industry/Statistics/World-Statistics/ [Accessed 12 March 2018].

World Wide Fund for Nature 2018. Agriculture: Facts and Trends, South Africa [online]. Available from: http://awsassets.wwf.org.za/downloads/facts_brochure_mockup_04_b.pdf [Accessed 16 April 2018].

Wu B, Meng J, Li Q, Yan N, Du X & Zhang M 2014. Remote sensing-based global crop monitoring: experiences with China's CropWatch system. *International Journal of Digital*

Earth 7, 2: 113–137.

Wu M, Yang C, Song X, Hoffmann WC, Huang W, Niu Z, Wang C & Li W 2017. Evaluation of orthomosics and digital surface models derived from aerial imagery for crop type mapping. *Remote Sensing* 9, 3: 1–14.

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

Yalcin H & Günay B 2016. Delineation of Parcels in Agricultural Lands in Remote Sensing. *Fifth International Conference on Agro-Geoinformatics*.

Yan WY, Shaker A & El-Ashmawy N 2015. Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment* 158: 295–310.

Yang C, Everitt JH, Du Q, Luo B & Chanussot J 2013. Using high-resolution airborne and satellite imagery to assess crop growth and yield variability for precision agriculture. *Proceedings of the IEEE* 101, 3: 582–592.

You J, Li X, Low M, Lobell D & Ermon S 2017. *Deep Gaussian process for crop yield prediction based on remote sensing data*. 31st AAAI Conference on Artificial Intelligence, AAAI 2017: 4559–4565.

Yue A, Zhang C, Yang J, Su W, Yun W & Zhu D 2013. Texture extraction for object-oriented classification of high spatial resolution remotely sensed images using a semivariogram. *International Journal of Remote Sensing* 34, 11: 3736–3759.

Zelinsky A 2009. Learning OpenCV: Computer Vision with the OpenCV Library. *IEEE Robotics and Automation Magazine* : 100

Zhang J 2010. Multi-source remote sensing data fusion: Status and trends. *International Journal of Image and Data Fusion* 1, 1: 5–24.

Zhang R & Zhu D 2011. Study of land cover classification based on knowledge rules using high-resolution remote sensing images. *Expert Systems with Applications* 38, 4: 3647–3652.

Zhang X, Wang J, Gao F, Liu Y, Schaaf C, Friedl M, Yu Y, Jayavelu S, Gray J, Liu L, Yan D & Henebry GM 2017. Exploration of scaling effects on coarse resolution land surface phenology. *Remote Sensing of Environment* 190: 318–330.

Zheng B, Myint SW, Thenkabail PS & Aggarwal RM 2015. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation* 34, 1: 103–112.

Zhou W 2013. An object-based approach for urban land cover classification: Integrating LiDAR height and intensity data. *IEEE Geoscience and Remote Sensing Letters* 10, 4: 928–931.

Zimmerman DW & Zumbo BD 1993. Relative Power of the Wilcoxon the Friedman Test , and Repeated Test , Measures ANOVA on Ranks. *Journal of Experimental Education* 62, 1: 75–86.

APPENDICES

APPENDIX A

Results for the different classification after standardisation. All the resolutions and window size classification results are shown.

Random Forest		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	76.43	79.92
2.0m	77.47	80.19
2.5m	77.80	80.17
3.0m	78.06	80.36
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.53	0.60
2.0m	0.55	0.60
2.5m	0.56	0.60
3.0m	0.56	0.61
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.45	0.49
2.0m	0.51	0.56
2.5m	0.51	0.56
3.0m	0.47	0.53
AUC		
Resolution (m)	3x3	5x5
1.5m	0.76	0.80
2.0m	0.77	0.80
2.5m	0.78	0.80
3.0m	0.78	0.80
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.78	0.81
2.0m	0.79	0.81
2.5m	0.80	0.82
3.0m	0.80	0.82

Decision Trees		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	67.31	70.27
2.0m	68.36	70.76
2.5m	68.72	71.05
3.0m	69.20	71.34
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.35	0.41
2.0m	0.37	0.42
2.5m	0.37	0.42
3.0m	0.38	0.43
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.68	0.63
2.0m	0.67	0.59
2.5m	0.70	0.65
3.0m	0.61	0.67
AUC		
Resolution (m)	3x3	5x5
1.5m	0.67	0.70
2.0m	0.68	0.71
2.5m	0.69	0.71
3.0m	0.69	0.71
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.68	0.70
2.0m	0.68	0.71
2.5m	0.69	0.71
3.0m	0.69	0.71

XGBoost		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	75.37	78.74
2.0m	76.36	78.97
2.5m	76.77	79.20
3.0m	77.58	79.42
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.51	0.57
2.0m	0.53	0.58
2.5m	0.53	0.58
3.0m	0.55	0.59
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.53	0.49
2.0m	0.48	0.54
2.5m	0.49	0.49
3.0m	0.54	0.49
AUC		
Resolution (m)	3x3	5x5
1.5m	0.75	0.79
2.0m	0.76	0.79
2.5m	0.77	0.79
3.0m	0.77	0.79
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.78	0.80
2.0m	0.79	0.81
2.5m	0.79	0.81
3.0m	0.80	0.81

Neural Network		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	74.82	78.79
2.0m	76.17	79.20
2.5m	76.50	78.87
3.0m	76.81	79.18
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.50	0.58
2.0m	0.52	0.58
2.5m	0.53	0.58
3.0m	0.54	0.58
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.59	0.52
2.0m	0.56	0.49
2.5m	0.54	0.51
3.0m	0.48	0.58
AUC		
Resolution (m)	3x3	5x5
1.5m	0.75	0.79
2.0m	0.76	0.79
2.5m	0.76	0.79
3.0m	0.77	0.79
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.77	0.80
2.0m	0.78	0.80
2.5m	0.78	0.80
3.0m	0.79	0.81

Deep Neural Network		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	73.83	77.66
2.0m	75.15	78.06
2.5m	75.60	77.84
3.0m	75.72	77.86
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.48	0.55
2.0m	0.50	0.56
2.5m	0.51	0.56
3.0m	0.51	0.56
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.69	0.67
2.0m	0.65	0.62
2.5m	0.66	0.57
3.0m	0.65	0.61
AUC		
Resolution (m)	3x3	5x5
1.5m	0.74	0.78
2.0m	0.75	0.78
2.5m	0.76	0.78
3.0m	0.76	0.78
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.72	0.76
2.0m	0.73	0.77
2.5m	0.74	0.77
3.0m	0.74	0.77

K-Nearest Neighbor		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	69.29	75.01
2.0m	70.28	75.28
2.5m	70.38	75.11
3.0m	70.55	74.61
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.39	0.50
2.0m	0.40	0.51
2.5m	0.41	0.50
3.0m	0.41	0.49
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.55	0.50
2.0m	0.49	0.52
2.5m	0.56	0.51
3.0m	0.46	0.57
AUC		
Resolution (m)	3x3	5x5
1.5m	0.69	0.75
2.0m	0.70	0.75
2.5m	0.70	0.75
3.0m	0.70	0.75
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.72	0.77
2.0m	0.72	0.77
2.5m	0.73	0.77
3.0m	0.73	0.76

Logistic Regression		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	69.69	73.36
2.0m	70.38	73.06
2.5m	71.54	73.72
3.0m	71.73	73.58
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.39	0.47
2.0m	0.41	0.46
2.5m	0.43	0.47
3.0m	0.43	0.47
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.52	0.49
2.0m	0.58	0.59
2.5m	0.51	0.56
3.0m	0.49	0.48
AUC		
Resolution (m)	3x3	5x5
1.5m	0.70	0.73
2.0m	0.70	0.73
2.5m	0.71	0.74
3.0m	0.72	0.73
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.72	0.75
2.0m	0.73	0.75
2.5m	0.75	0.76
3.0m	0.75	0.76

Naïve Bayes		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	55.98	58.32
2.0m	56.28	56.97
2.5m	56.25	56.00
3.0m	60.43	58.80
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.11	0.16
2.0m	0.12	0.13
2.5m	0.12	0.11
3.0m	0.20	0.17
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	2.07	1.39
2.0m	2.03	1.23
2.5m	0.85	0.96
3.0m	2.50	1.30
AUC		
Resolution (m)	3x3	5x5
1.5m	0.56	0.58
2.0m	0.56	0.57
2.5m	0.56	0.56
3.0m	0.60	0.59
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.69	0.70
2.0m	0.69	0.70
2.5m	0.69	0.69
3.0m	0.71	0.70

SVM Linear		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	68.62	72.89
2.0m	69.42	72.51
2.5m	70.82	73.27
3.0m	70.98	73.14
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.37	0.46
2.0m	0.39	0.45
2.5m	0.41	0.46
3.0m	0.42	0.46
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.56	0.54
2.0m	0.58	0.49
2.5m	0.58	0.50
3.0m	0.56	0.53
AUC		
Resolution (m)	3x3	5x5
1.5m	0.69	0.73
2.0m	0.69	0.72
2.5m	0.71	0.73
3.0m	0.71	0.73
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.71	0.75
2.0m	0.73	0.75
2.5m	0.74	0.76
3.0m	0.75	0.76

SVM RBF		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	71.73	75.46
2.0m	72.43	75.88
2.5m	72.92	75.86
3.0m	73.04	75.76
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.43	0.51
2.0m	0.45	0.52
2.5m	0.46	0.52
3.0m	0.46	0.51
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.61	0.55
2.0m	0.54	0.49
2.5m	0.61	0.56
3.0m	0.58	0.59
AUC		
Resolution (m)	3x3	5x5
1.5m	0.72	0.75
2.0m	0.72	0.76
2.5m	0.73	0.76
3.0m	0.73	0.76
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.76	0.78
2.0m	0.76	0.78
2.5m	0.76	0.78
3.0m	0.77	0.78

SVM GRID SEARCH		
Overall Accuracy		
Resolution (m)	3x3	5x5
1.5m	69.91	74.06
2.0m	70.59	73.51
2.5m	71.56	73.56
3.0m	71.57	74.10
Kappa		
Resolution (m)	3x3	5x5
1.5m	0.40	0.48
2.0m	0.41	0.47
2.5m	0.43	0.47
3.0m	0.43	0.48
Standard Deviation		
Resolution (m)	3x3	5x5
1.5m	0.70	0.65
2.0m	0.54	0.47
2.5m	0.48	0.52
3.0m	0.66	0.69
AUC		
Resolution (m)	3x3	5x5
1.5m	0.70	0.74
2.0m	0.70	0.73
2.5m	0.71	0.74
3.0m	0.71	0.74
F-Score		
Resolution (m)	3x3	5x5
1.5m	0.63	0.70
2.0m	0.64	0.69
2.5m	73.56	0.68
3.0m	0.65	0.69

APPENDIX B

Confusion matrices for the RF classification algorithm and the seven experiments

A1								
	Non-Agri	Corn	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	3872	109	328	1012	634	5955	65,0	
Corn	271	3533	897	937	446	6084	58,1	
Orchard	773	1254	2858	174	963	6022	47,5	
Groundnut	1336	1135	79	2932	499	5981	49,0	
Cotton	889	761	670	519	3119	5958	52,3	
Total	7141	6792	4832	5574	5661	16314	OA	54,4
PA	54,2	52,0	59,1	52,6	55,1		Kappa	0,28

A2								
	Non-Agri	Maize	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	4667	56	514	596	134	5967	78.2	
Maize	165	5005	114	232	544	6060	82.6	
Orchard	223	0	5736	52	14	6025	95.2	
Groundnut	724	452	3	4482	210	5871	76.3	
Cotton	204	693	174	330	4676	6077	76.9	
Total	5983	6206	6541	5692	5578	24566	OA	81.9
PA	78.0	80.6	87.7	78.7	83.8		Kappa	0.76

S								
	Non-Agri	Maize	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	5538	0	273	65	164	6040	91.69	
Maize	84	5598	216	132	77	6107	91.67	
Orchard	198	29	5347	108	217	5899	90.64	
Groundnut	270	53	467	4753	387	5930	80.15	
Cotton	279	54	486	485	4720	6024	78.35	
Total	6369	5734	6789	5543	5565	25956	OA	86.5
PA	87.0	97.6	78.8	85.7	84.8		Kappa	0.82

L								
	Non-Agri	Maize	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	4560	25	309	850	190	5934	76.8	
Maize	143	5619	7	8	248	6025	93.3	
Orchard	195	0	5774	66	0	6035	95.7	
Groundnut	912	1	31	5049	26	6019	83.9	
Cotton	493	152	32	136	5174	5987	86.4	
Total	6303	5797	6153	6109	5638	26176	OA	87.3
PA	72.3	96.9	93.8	82.6	91.8		Kappa	0.83

A-S								
	Non-Agri	Maize	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	5595	7	183	105	49	5939	94.2	
Maize	81	5616	81	211	74	6063	92.6	
Orchard	176	4	5730	31	25	5966	96.0	
Groundnut	187	5	34	5580	244	6050	92.2	
Cotton	185	40	95	267	5395	5982	90.2	
Total	6224	5672	6123	6194	5787	27916	OA	93.1
PA	89.9	99.0	93.6	90.1	93.2		Kappa	0.91

A-L								
	Non-Agri	Maize	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	4941	33	321	570	104	5969	82.8	
Maize	161	5647	27	0	192	6027	93.7	
Orchard	160	0	5696	34	1	5891	96.7	
Groundnut	572	0	21	5399	54	6046	89.3	
Cotton	257	93	26	157	5534	6067	91.2	
Total	6091	5773	6091	6160	5885	27217	OA	90.7
PA	81.1	97.8	93.5	87.6	94.0		Kappa	0.88

L-S								
	Non-Agri	Maize	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	5455	24	172	177	151	5979	91.2	
Maize	15	5627	11	63	272	5988	94.0	
Orchard	162	0	5852	44	5	6063	96.5	
Groundnut	252	2	29	5555	124	5962	93.2	
Cotton	220	62	27	224	5475	6008	91.1	
Total	6104	5715	6091	6063	6027	27964	OA	93.2
PA	89.4	98.5	96.1	91.6	90.8		Kappa	0.91

A-S-L								
	Non-Agri	Maize	Orchard	Groundnut	Cotton	Total	UA	
Non-Agri	5555	26	160	108	47	5896	94.22	
Maize	39	5708	6	58	219	6030	94.66	
Orchard	156	0	5759	27	0	5942	96.92	
Groundnut	210	0	25	5671	69	5975	94.91	
Cotton	212	60	28	167	5690	6157	92.42	
Total	6172	5794	5978	6031	6025	28383	OA	94.6
PA	90.0	98.5	96.3	94.0	94.4		Kappa	0.93

APPENDIX C

The results for the different classifiers and the seven different experiment experiments. OA, Kappa, f-score and SD are shown.

DT								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	46,1	72.2	84.7	86.2	90.0	82.3	90.2	81.0
K	0,3	0.65	0.81	0.83	0.87	0.78	0.88	0.76
SD	2,6	2.75	2.31	2.04	1.72	2.14	1.73	2.12
f-score	0,5	0.72	0.85	0.86	0.90	0.82	0.90	0.81

RF								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	54,1	81.9	90.7	93.1	94.6	87.3	93.2	86.5
K	0,4	0.77	0.88	0.91	0.93	0.84	0.91	0.83
SD	2,2	2.15	1.53	1.30	1.25	2.01	1.15	1.80
f-score	0,5	0.82	0.91	0.93	0.95	0.87	0.93	0.87

XGBoost								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	56,1	81.3	91.3	91.9	94.1	87.8	93.0	86.3
K	0,5	0.77	0.89	0.90	0.93	0.85	0.91	0.83
SD	2,7	1.99	1.69	1.68	1.32	1.84	1.27	2.02
f-score	0,6	0.81	0.91	0.92	0.94	0.88	0.93	0.86

k-NN								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	54,5	77.1	87.7	88.9	92.1	83.9	91.2	85.8
K	0,4	0.71	0.85	0.86	0.90	0.80	0.89	0.82
SD	2,3	2.07	1.51	1.49	1.28	1.99	1.37	1.70
f-score	0,5	0.77	0.88	0.89	0.92	0.84	0.91	0.86

LR								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	44,5	73.2	86.8	91.6	92.9	84.9	92.2	85.3
K	0,3	0.67	0.83	0.89	0.91	0.81	0.90	0.82
SD	2,6	2.30	1.65	1.46	1.52	1.66	1.44	1.97
f-score	0,4	0.73	0.87	0.92	0.93	0.85	0.92	0.85

NB								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	46,7	62.5	81.2	74.8	86.0	77.7	84.7	67.9
K	0,3	0.53	0.76	0.69	0.82	0.72	0.81	0.60
SD	2,4	2.67	2.05	2.34	2.11	2.31	1.89	2.62
f-score	0,5	0.62	0.81	0.75	0.86	0.78	0.85	0.68

SVM-L								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	44	73.4	88.2	92.6	93.5	86.2	93.5	88.3
K	0,3	0.67	0.85	0.91	0.92	0.83	0.92	0.85
SD	2,5	2.20	1.70	1.30	1.34	1.77	1.34	1.68
f-score	0,4	0.73	0.88	0.93	0.94	0.86	0.94	0.88

SVM RBF								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	50,5	72.0	86.6	87.0	90.1	83.4	89.8	75.9
K	0,4	0.65	0.83	0.84	0.88	0.79	0.87	0.70
SD	2,6	2.69	1.78	1.67	1.63	1.96	1.70	2.34
f-score	0,5	0.72	0.87	0.87	0.90	0.83	0.90	0.76

NN								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	56,5	81.2	89.8	92.7	93.4	86.3	92.8	88.2
K	0,5	0.76	0.87	0.91	0.92	0.83	0.91	0.85
SD	2,6	2.10	1.43	1.41	1.38	1.85	1.32	1.52
f-score	0,6	0.81	0.90	0.93	0.93	0.86	0.93	0.88

d-NN								
Experiment	A1	A2	A-L	A-S	A-S-L	L	L-S	S
OA	56,3	81.0	88.2	92.3	92.2	83.2	91.5	90.8
K	0,5	0.76	0.85	0.90	0.90	0.79	0.89	0.88
SD	2,6	1.83	1.73	1.38	1.49	2.23	1.42	1.68
f-score	0,6	0.81	0.88	0.92	0.92	0.83	0.92	0.91