THE USE OF MACHINE LEARNING ALGORITHMS TO ASSESS THE IMPACTS OF DROUGHTS ON COMMERCIAL FORESTS IN KWAZULU-NATAL, SOUTH AFRICA

By

Mthokozisi Ndumiso Mzuzuwentokozo Buthelezi

Supervisor: Dr RT Lottering

Submitted in partial fulfilment of the academic requirements for the degree of Master of Science in the Discipline of Applied Environment Science

School of Agricultural, Earth Environmental Sciences

University of KwaZulu-Natal

Pietermaritzburg

May 2020

UNIVERSITY OF KWAZULU-NATAL ™

INYUVESI YAKWAZULU-NATALI

Abstract

Droughts are a non-selective natural disaster in that their occurrence can be in both high and low precipitation areas. However, this study acknowledged that droughts are more recurrent and a regular feature in arid and semi-arid climates such as that of Southern Africa. Some of these countries rely strongly on commercial forests for their gross domestic product (GDP), especially South Africa and Mozambique which means droughts pose a significant threat to their economy and the society that depends on this economy. The risks associated with droughts have consequently created an increased demand for an efficient method of analysing and investigating droughts and the impacts they impose on forest vegetation. Therefore, this study aimed to examine the effects of droughts on all commercial forests within the province of KwaZulu-Natal (KZN) at a catchment and provincial scale by employing Kernel Support Vector Machine (Kernel –SVM), Rotation Forests (RTF) and Extreme Gradient Boosting (XGBoost) algorithms. These were based on Landsat and MODIS derived vegetation and conditional drought indices. The main aim of this study was achieved by the following objectives: (i) to improve methods for classifying droughts; (ii) to achieve medium spatial resolution drought analysis using Landsat sensors; (iii) to determine the accuracy of machine learning algorithms (MLAs) when employed on remote sensing data and (iv) to improve the usability of conditional drought indices and vegetation indices. The results obtained there-after demonstrated that the objectives of this study were met. With the MLAs performing better when using conditional drought indices compared to vegetation indices, therefore, highlighting drawbacks already associated with vegetation indices. Where at the catchment scale, Kernel – support vector machine (SVM) produced an overall accuracy (OA) of 94.44% when based on conditional drought indices compared to 81.48% when based on vegetation indices. On the same scale, Rotation forests (RTF) produced 96.30% and 81.84% when using conditional drought indices and vegetation indices, respectively. At a provincial scale, RTF produced an OA of 76.6% and 70.7% when using conditional drought indices and vegetation indices respectively. This was compared to extreme gradient boosting (XGBoost) which produced an OA of 81.9% and 69.3% when using conditional drought indices and vegetation indices respectively. These results also indicate that it is possible to analyse droughts at provincial and catchment scale. Although the results presented in this study were promising, more research is still required to improve the applicability of MLAs in drought analysis.

Keywords: Drought, Machine Learning, Rotation Forests, Kernel, Extreme Gradient Boosting

“

For I know the plans I have for you, declares the Lord, plans to prosper you, plans to give you hope and a future.

- Jeremiah 29:11

Dedication

To my late great grandmother, my family, my girlfriend, my friends, my team (Ababulali Football Club) and all the people impacted by the COVID-19 pandemic.

Acknowledgements

From start to finish, this project would not have been a success without the assistance of many people.

To Dr Romano Lottering and Mr Sizwe Hlatshwayo, my supervisors, without you this project would not have been a possibility. Thank you for your guidance, support, advice and for pushing me to reach this point of my thesis.

To my mother, words cannot express how grateful I am for everything you do for me. I am unable to say I completed my Masters because, instead I can only say we completed this Masters.

To Miss Swelihle Msomi and Miss Sphiwokuhle Buthelezi, thank you for caring and supporting me throughout the course of this project.

To Mrs Sindisiwe Lembethe and Miss Nontobeko Zondo, thank you for helping get to school and taking my stuff that I could not carry in a taxi.

To my friends, Mr Lonwabo Manqana, Miss Lindiwe Nkabane, Miss Samukelisiwe Vilakazi, Mr Nhlakanipho Mahlalela, Mr Ntando and Andile Ndlazi, Mr Phakamani Madlala, Miss Nosiphiwo Nodada, Miss Vuyokazi Zungu and Mr Nkosikhona Zondo, thank you for being good examples, for encouraging and talking to me during the lockdown and most importantly for being my role models.

Special thanks to Dr Kabir Peerbhay and the staff at Sappi for collecting and providing field data used for this project.

Also, I would like to acknowledge the Moses Kotane Institute for providing financial relief for the completion of this project.

Finally, I would like to thank my family for being there for me when we endured the hardest times when our grandmother and her daughter Nompumelelo Nkosi passed on. They meant the world to us.

Preface

The work described in this paper was completed at the University of KwaZulu-Natal, Pietermaritzburg, from March 2019 to May 2020, under the supervision of Dr Lottering.

The research represented in this document is original work by the author and has not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others it is duly acknowledged in the text.

.... 03 May 2020

Signature (Student) Date

Mthokozisi Ndumiso Mzuzuwentokozo Buthelezi

.... 04 May 2020

Signature (Supervisor) Date

Dr Lottering

Declaration

I, Mthokozisi Ndumiso Mzuzuwentokozo Buthelezi declare that

1. The research reported in this paper, except where otherwise indicated, is my original work.
2. This project has not been submitted for any degree or examination at any other university.
3. This paper does not contain other persons' data, pictures, graphs or additional information unless specifically acknowledged as being sourced from other persons.
4. This paper does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a) Their words have been re-written, but the general information attributed to them has been referenced.
    b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.
5. This paper does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source is detailed in the paper and References sections.

Signed: ....

Table of Contents

List of Figures

List of Tables

1. CHAPTER ONE: GENERAL INTRODUCTION

1.1 Introduction

Commercial forests commonly referred to as plantation forests are often a topic of controversy in environmental and hydrological sectors, especially in a country like South Africa which is not forest-rich (Tewari, 2001; Tuswa *et al.*, 2019). This is due to these forests being associated with negative implications such as environmental degradation, high water usage and streamflow reduction (Scott *et al.*, 1998; Wei *et al.*, 2018; Tuswa *et al.*, 2019). These implications are primarily due to commercial forests being composed of mainly alien invasive species (Shackleton *et al.*, 2019). However, in the South African economic sector, commercial forests are highly valuable (Tewari, 2001; Wei *et al.*, 2018). That is because of the increased demand for forest products and the ability of commercial forests to generate employment opportunities and enable foreign exchange (Tewari, 2001; Naidoo *et al.*, 2013; Davis and Vincent, 2017). South Africa has thus invested heavily in commercial forestry. Hence, commercial forests account for one percent (1.273 million ha) of the country's land area and approximately half of all commercial forests in the Southern African Development Community (SADC) (Naidoo *et al.*, 2013; Davis and Vincent, 2017). Therefore, it can be argued that these governments prioritised the economy over the environment.

However, South Africa is susceptible to recurring droughts and is relatively dry (Scott *et al.*, 1998; Rouault and Richard, 2003). Droughts generally have major impacts on socio-economic, ecological and agricultural systems which are interlinked with forest ecosystems and therefore, droughts have been determined to have negative impacts on forests (Dale *et al.*, 2001; Wu *et al.*, 2013). A continental summary by Allen *et al.* (2010) established that increased forest mortality in Africa is due to extensive droughts. This is further compounded by diseases and insect outbreaks induced by droughts which intensify tree mortality (Ganey and Vojta, 2011; Wu *et al.*, 2013). Therefore, there is a need to research, monitor and analyse the impacts of droughts on forests.

Monitoring and analysing droughts require some numerical standard; that will aid in understanding and quantifying the impacts of droughts (Heim Jr, 2002; Hayes *et al.*, 2011). Gaining this understanding and quantification would allow for timely responses to droughts (Peters *et al.*, 2002). Compared to raw indicator data, this standard is also more readily useable (Zargar *et al.*, 2011). Establishing this standard is, however, hindered by the complexity of droughts that arise from a lack of a definitive definition, the difficulty in determining its severity and the difficulty to accurately predict its onset and ending (Peters *et al.*, 2002; Hayes *et al.*, 2011; Baniya *et al.*, 2019).

The difficulty in determining the severity of droughts stems from the fact that it is not only dependent on the spatial extent, duration and intensity but also the demand on water resources by the type of vegetation, economic and human activities within a region experiencing the drought (Wilhite and Glantz, 1985). The complexity of droughts and the existence of multiple definitions has therefore resulted in the development of multiple drought numerical standards of which most are only applicable for a specific region and application (Heim Jr, 2002; Zhuo *et al.*, 2016). These include conventional indices that are based on meteorological indices (MI) such as Palmer drought severity index (PDSI), standardized precipitation evapotranspiration index (SPEI) and standardized precipitation index (SPI) (Peters *et al.*, 2002; Wu *et al.*, 2013; Zhang *et al.*, 2017b). These indices are limited by the sparse and limited availability of weather station networks, especially in forested areas (Caccamo *et al.*, 2011).

The limitations associated with MI can be overcome by remote sensing which can monitor meteorological, agricultural and hydrological droughts (Wu *et al.*, 2013). Remote sensing enables spatial and temporal analysis of drought impacts through repeated image acquisition over a given location (Dube *et al.*, 2016). Thus, there has been increased development of multiple drought indices that are based on remote sensing (Peters *et al.*, 2002; Zhuo *et al.*, 2016; Zhang *et al.*, 2017b). Remote sensing indices include evapotranspiration (ET), Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) (Zhuo *et al.*, 2016). From the indices mentioned above, conditional indices can be derived. Conditional indices include the LST based temperature condition index (TCI) and NDVI based vegetation condition index (VCI) which are prevalent conditional indices for monitoring the impacts, duration and severity of droughts (Du *et al.*, 2013; Zhuo *et al.*, 2016).

Multiple studies have analysed droughts at continental, national and regional scale; these include Ganey and Vojta (2011) who studied droughts by investigating their influence on tree mortality and determined that alongside climate change, they pose a significant threat for forests at a global scale. Zhang and Jia (2013) who monitored drought utilising multi-sensor microwave remote sensing data in semi-arid northern China. Baniya *et al.* (2019) studied temporal and spatial variation of drought by applying satellite derived VCI in Nepal from 1982–2015. This current study deviates from the aforementioned studies in that it investigates the impacts of droughts at a catchment scale, which will yield more detailed and location-specific results. The results obtained from the catchment analysis will then be compared with those obtained from the provincial scale analysis.

Remote sensing has improved considerably with temporal resolutions approaching real-time image acquisition and spatial resolutions now at sub-meter level (Nouri *et al.*, 2014). However, limited studies are using high temporal and spatial resolution sensors because of their limited availability and high costs (Dube *et al.*, 2016). Given the economic status of South Africa as a developing country, it is important to utilise cheap and easily accessible remote sensing imagery from sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat (Dube and Mutanga, 2015; Dube *et al.*, 2016). Therefore, Landsat sensors (Landsat 4 and 5, 7 and 8) and MODIS will be utilised for the purposes of this study. Information from these sensors will yield high-volume data sets that will require classification into sensible groups that can be interpreted as meaningful information which is required to make conclusive arguments (Jain, 2010; Mountrakis *et al.*, 2011).

There are multiple methods available for organising data; such include factor analysis, principal component analysis (PCA), linear regression and cluster analysis amongst others (Jain, 2010). These methods are generally classified as supervised when reference data is required or unsupervised when image statistics are utilised to obtain the required parameters (Belgiu and Drăguţ, 2014). Also, according to Belgiu and Drăguţ (2014) there had been no study comparing supervised and unsupervised data segmentation methods before their study. However, most studies have tended to utilise and recommend supervised object-based techniques if there is accurate and efficient reference data (Ma *et al.*, 2017).

Unsupervised methods are recently becoming widely accepted and utilised by the science community (Afanador *et al.*, 2016). However, unsupervised methods are difficult and challenging (Jain, 2010), and more studies should be conducted to determine their accuracy when classifying large volumes of data. Therefore, this study utilised a supervised approach through supervised MLAs, namely, Kernel - SVM, RTF and XGBoost. MLAs are gaining popularity within remote sensing because they can accurately classify non-linear and high dimensional data, they accept varying input predictor data and are non-parametric (Maxwell *et al.*, 2018; Shiferaw *et al.*, 2019). Multiple studies have also determined that MLAs can match if not exceed the accuracies obtained by conventional remote sensing classification methods (Mountrakis *et al.*, 2011; Ghimire *et al.*, 2012).

However, there is often limited reference data in remote sensing which makes SVMs more applicable given their performance gain and the fact that researchers have evolved them to be employable on low quality and quantity training data (Mountrakis *et al.*, 2011). Kernel – SVM is one of the better versions of SVMs, and its basis is that each classifier is trained on samples derived from a pair of labelled classes that the classifier must learn to distinguish (Hobbs, 2018). RTF is defined by Blaser and Fryzlewicz (2016) as "an ensemble method which combines the predictions of multiple base learners to create more accurate predictions." RTF also assumes that one classifier is independently constructed utilising the decision tree technique where each tree is trained in a rotated feature space on the training samples (Xia *et al.*, 2013). XGBoost is based on gradient boosting machines (GBMs), which have been applied widely on multiple remote sensing applications (Georganos *et al.*, 2018). It is adjusted for large tree structures and yields good classifications with high execution speeds (Sandino *et al.*, 2018). Utilising these three MLAs will give insight to future prospects of ML in remote sensing.

1.2 Aims and Objectives

This study aims to analyse the impacts of droughts on all commercial forests within the province of KwaZulu-Natal (KZN) at a catchment and provincial scale by employing MLAs on Landsat and MODIS derived from vegetation and conditional drought indices.

To meet the above aim, the following objectives were determined:

- To evaluate ML image classification methods for drought mapping.
- To achieve high spatial resolution drought analysis using Landsat sensors.

- To implement machine learning approaches to classify drought damage on commercial forests.
- To improve the usability of conditional drought indices and vegetation indices.

1.3 Research Questions

- Can ML image classification methods be used for drought mapping?
- Can Landsat be used to analyse the impacts of droughts on commercial forests at a catchment scale?
- How does the drought damage classification conducted at catchment scale compare with those conducted at a provincial scale?
- Can ML algorithms improve on results achieved using conventional classification methods?

1.4 Significance of the Study

The findings from this study are very beneficial to the commercial forestry sector and therefore the economy of South Africa because this sector contributes significantly to the GDP of the country. However, droughts have been found to be more recurrent in South Africa and therefore, assessing and understanding their impacts on forest vegetation is very important for the commercial forestry sector to adapt to this phenomenon. Thus, researchers that apply methods recommended in this study will be able to assess drought impacts more effectively and efficiently.

1.5 Limitations of the Study

Though significant, some limitations of this study should be noted. The study methodology did not account for different species within the Sappi forests. That was because the reference data was not mapped out to indicate the grouping of different species. Another limitation is that the study utilised only the reference data and no field visits were conducted due to the scope of the study.

1.6 Scope of the Study

The study assessed the impacts of droughts on commercial forests in KZN at catchment and provincial scales. The 2015 – 2016 drought was the main focus of the study; hence the duration of the study was the beginning of 2015 to the end of 2017.

1.7 Structure of the Thesis

Four chapters form this thesis where the first chapter presents the general introduction and the last chapter presents the synthesis and concluding remarks. Chapter Two and Three are made up of two research papers which seek to meet objectives mentioned in 1.2 and answer research questions mentioned in 1.3.

Chapter Two classifies the impacts of droughts on commercial forests at a catchment scale by employing Kernel – SVM and RTF based on Landsat 4 and 5, 7 and 8 data. The study also gives insight on the different accuracies obtained when using ML algorithms on vegetation and conditional indices. Chapter Three focuses on the classification of commercial forest trees impacted by droughts and those which were not; the analysis was conducted at a provincial scale. RTF and XGBoost were employed on vegetation and conditional drought indices derived from MODIS imagery to achieve the results.

2. CHARPTER TWO: IMPLEMENTING MACHINE LEARNING APPROACHES TO CLASSIFY DROUGHT DAMAGE ON COMMERCIAL FORESTS.

Abstract

South Africa and other countries within Southern Africa are susceptible to recurrent droughts which pose a major threat to commercial forests. Commercial forestry contributes significantly to the economy of these countries as well as the livelihood of people that live in them. Droughts enhance forest degradation and exacerbate wildfire outbreaks and destructive pest outbreak. It is these events or the possibility of their occurrence that can or could reduce the productivity and profitability of commercial forests, hence there is an increased demand for an efficient method of analysing and investigating droughts and the impacts they impose on forest vegetation. This study, therefore, utilised two MLAs, namely, RTF and Kernel – SVM for the classification of drought damaged trees and those that were not damaged within the small Sappi Shafton plantation located in the Natal Midlands during the intense and prolonged 2015 – 2016 drought. The algorithms were employed on conditional drought (VCI and VMCI) and vegetation (NDVI, EVI, SAVI, NDWI, SR and GCHL) indices derived from Landsat 4 and 5, 7 and 8 imageries from 2001 to 2017. Kernel - SVM produced an OA of 94.44% when based on conditional drought indices compared to 81.48% when based on vegetation indices. RTF produced 96.30% and 81.84% when using conditional drought indices and vegetation indices, respectively. These results showed that both algorithms are capable of accurately detecting trees that exhibit drought damage and those that do not, more so RTF and when the algorithms were employed on conditional drought indices data. When applied on vegetation indices data, the performance of the algorithms decreased considerably. Overall, the study demonstrated that MLAs can be employed when classifying drought damage on forest vegetation, Landsat data can be used for analysing drought damage at a catchment or regional scale, drought analysis can be undertaken at a catchment scale and that conditional drought indices are a better option for drought analysis than vegetation indices.

Keywords: drought, machine learning, rotation forest, Kernel support vector machine, catchment

2.1 Introduction

Commercial forests are crucial for South Africa's economy (Führer, 2000). In 2015, commercial forestry had a contribution approximated to R31.1 billion towards the South African GDP and brought employment to 158 400 people (Xulu *et al.*, 2019). However, with droughts occurring more frequently; these forests are faced with aggravated degradation which is further compounded by drought-induced wildfires and pests outbreaks (Kirilenko and Sedjo, 2007). It has been reported that the rate of occurrence of new destructive pest in South Africa is alarming (SA Forestry, 2016) and wildfires do not only exacerbate forest degradation, but they also result in deforestation which is a concern on its own (Poursanidis and Chrysoulakis, 2017). Forest degradation is a complex process, defined by a permanent or a temporary deterioration in the structure or density of the vegetation cover, where the deterioration is not necessarily due to land-use change (Lambin, 1999; Hosonuma *et al.*, 2012). The complexity of forest degradation arises as a result of its strong interaction with climatic fluctuations that generally occurred naturally but have become strongly influenced by human-induced climate change (Lambin, 1999).

Climate change has also, resulted in declining precipitation for semi-arid regions such as South Africa (Hans, 1948; Nicholson, 1986; Olsson, 1993). This declining precipitation has led to significant drought periods in the country, namely, 1950-1969, 1970-1988, 1992-1995 and 2015-2016 droughts (Nicholson, 1986; Richard *et al.*, 2001; Baudoin *et al.*, 2017). Thus, the characterisation of South Africa as a drought susceptible country (Vogel, 1994; Rouault and Richard, 2003). Studies by Breshears *et al.* (2005), Nepstad *et al.* (2007), Asner and Alencar (2010), da Costa *et al.* (2010), Saatchi *et al.* (2013) and Martin-Benito *et al.* (2017) found that the impacts of droughts on forests is reduced canopy cover, tree growth, and above-ground biomass which are components of forest degradation. From these findings, it can be concluded that forest degradation resulting from droughts will harm commercial forest productivity and profitability (Warburton and Schulze, 2008). This does not bode well for a country like South Africa, which is dependent on commercial forestry for its GDP.

Climate change further compounds the recurrent nature of droughts in South Africa by increasing their severity, extent and frequency (Asner and Alencar, 2010). It is these drought conditions that are necessary for the manifestation of drought impacts on forest trees (Warburton and Schulze, 2008; Gazol *et al.*, 2017; Xulu *et al.*, 2019). The 2015-2016 drought which resulted due to an extreme El Niño (Bahta *et al.*, 2016) embodied these conditions. Agri SA (2016) reported that this drought had such high magnitude, intensity and temperatures that planning was beyond any farmer or ministry with any resource base. The year 2015 also saw the lowest mean annual precipitation on record for South Africa since 1904 which was 403 mm (De Jager, 2016). The losses accumulated by farmers costed up to an estimated R10 million (Bahta *et al.*, 2016). Therefore monitoring droughts and the resulting forest degradation and other associated impacts of droughts has become inevitable and a necessity to limit the losses incurred by farmers and the government (Gazol *et al.*, 2017).

Monitoring will also aid the consideration of potential drought adaptation measures (Warburton and Schulze, 2008). However, like forest degradation, droughts are also complex. Their complexity arises from their complex impacts, slow onset and lack of a specific definition (Botai *et al.*, 2016). This is further compounded by different species or individual trees exhibiting different responses under similar drought stresses (Martin-Benito *et al.*, 2017). This complexity has made it difficult to detect droughts (Xulu *et al.*, 2019). As a result, multiple drought indices are developed and most are only applicable to a specific region and application (Heim Jr, 2002; Zhuo *et al.*, 2016).

Remote sensing, however, uses optical sensors to detect droughts directly and can, therefore, overcome the hassle of selecting a suitable drought index (De Sy *et al.*, 2012; Wu *et al.*, 2013; Mladenova *et al.*, 2014; Xulu *et al.*, 2019). Also, drought impacts occur slowly through land cover modifications which require integration of temporal, spatial and spectral information that is only possible with remote sensing (Lambin, 1999). This is due to remote making it possible to access systematic observation systems and historical data archives (Rosenqvist *et al.*, 2003).

However, the application of remote sensing is hindered by limited access to required data, especially in developing countries where funding for projects is generally lacking (Van Westen and Soeters, 2002; De Sy *et al.*, 2012; Bello and Aina, 2014). Lack of funds meant researchers could not access remote sensing data products which had been commercialised since 1948 and each scene was valued at approximately $4400; that was until the introduction of Landsat-7 which decreased the cost of data to approximately $660 per scene (Wulder *et al.*, 2012). In January of 2008 the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS) released the Landsat Data Distribution Policy which enabled Landsat products starting from 1972 to be freely distributed (Woodcock *et al.*, 2008). This policy has resulted in higher usage of Landsat data products and has enabled the undertaking of studies that seek to monitor the impacts of droughts.

Multiple drought and vegetation indices were derived for purposes of analysing the impact of droughts on forest vegetation using Landsat data. Several studies as reported by Du *et al.* (2013) have utilised conditional drought indices with great success, namely, NDVI based VCI (Kogan and Sullivan, 1993; Kogan, 1995b) and LST based TCI (Kogan, 1997). Though accurate separately, Kogan (1997) and Du *et al.* (2013) found that the combination of TCI and VCI yielded the vegetation health index (VHI) which is a more accurate index, therefore, increasing the accuracy of the results. Also, using a single drought index may not be as accurate as using multiple indices for the drought analysis (Hao and AghaKouchak, 2013). Therefore, this study will utilise multiple drought indices, namely, VCI and vegetation moisture condition index (VMCI) based on the normalised difference moisture index (NDMI). Results obtained from the drought conditional indices will be compared with raw vegetation indices to determine which can accurately detect the impacts of droughts on forest vegetation. However, obtaining these results will require the employment of two MLAs that are also capable of visualising these results.

Machine learning is focused on computer-based systems that improve and adapt through experience and it has progressed considerably over the past two decades (Brunton *et al.*, 2020). Thus, it has become a preferred method for multiple applications including remote sensing (Jordan and Mitchell, 2015). Multiple algorithms based on machine learning have been developed and evolved and are said to yield higher accuracies than conventional classification methods (Arganda-Carreras *et al.*, 2017). MLAs include RTF developed by Rodriguez *et al.* (2006) based on random forests (RF) and Kernel – SVM which is a supervised learning method (Hobbs, 2018) were utilised to classify trees affected by droughts within the Sappi Shafton plantation between 2013 and 2018.

RTF is classified as an ensemble method that combines the predictions made by multiple base learners to create predictions that are more accurate (Blaser and Fryzlewicz, 2016). RTF also assumes that one classifier is independently created utilising the decision tree method where each tree is trained in a rotated feature space on the training samples (Xia *et al.*, 2013). Multiple studies which include Xia *et al.* (2013), Du *et al.* (2015) and Khamar and Eftekhari (2018) have utilised RTF for the classification of remote sensing imagery and determined that it generally performed better than other ensemble methods such as RF and SVM. To verify the accuracy and performance of RTF, it was compared with Kernel – SVM because; SVMs generally produce high accuracies (Roli and Fumera, 2001; Melgani and Bruzzone, 2004). Comparing these algorithms will aid in determining the shortcomings of each algorithm over the other and therefore assist researchers when selecting an algorithm to use for future studies. The basis for Kernel-SVM is that each classifier is trained on samples derived from a pair of labelled classes which the classifier must learn to distinguish (Hobbs, 2018). Kernel – SVM is employable on non-linearly separable data (Melgani and Bruzzone, 2004); which is important because data is generally non-linearly separable.

Against this background and the fact that machine learning's potential to analyse droughts and their impacts has not been investigated thoroughly, this study aimed to analyse the impacts of droughts on commercial forests over time at a catchment scale. To achieve this, conditional drought and vegetation indices were derived from scenes produced by Landsat 4 and 5, 7 and 8. Scenes from Landsat are obtained at a 30-meter resolution and 16-day revisit cycles, thus making it possible to analyse and visualise vegetation and land cover changes at a catchment scale using RTF and Kernel – SVM algorithms (Zhu et al., 2016). Therefore, the objectives of this study were adapting machine learning through RTF and Kernel – SVM for drought analysis, localising drought analysis to a catchment scale and determining how different indices affect the accuracy of RTF and Kernel – SVM.

2.2 Methodology

2.2.1 Study area description

The Sappi Forests Natal Midlands District or commonly the Sappi Shafton plantation (29°25'17.77"S and 30°11'14.43"E) is located within the Natal Midlands, which is approximately 70 km from Pietermaritzburg (Clulow *et al.*, 2011) (Figure 2.1). This region is defined by its subtropical oceanic climate that yield mild summers and cool dry winters (Hitayezu *et al.*, 2014). The Natal Midland's mean annual temperature is 17°C and this region receives approximately 659 mm to 1280 mm of rainfall per year, which is mostly during the wet summer months (Bulcock and Jewitt, 2012; Duncan, 2019). Summer rainfall in the Natal Midlands is characterised by many low-intensity events and some high-intensity storms (Bulcock and Jewitt, 2012). This region also gets frequent and heavy mist which has a considerable contribution to the area's precipitation (Clulow *et al.*, 2011).

The conditions mentioned above make the Natal Midlands an optimal region for perennial crops and commercial forestry with *Pinus patula (*pine), *Euculyptus grandis* (gum tree) and *Acacia mearnsii* (wattle) being the dominant plantation species (Bulcock and Jewitt, 2012; Albaugh *et al.*, 2013; Duncan, 2019). However, this region has a history of droughts and with persisting climate change, the conditions that make the Natal Midlands a mainstay for commercial forestry are slowly deteriorating (Warburton and Schulze, 2008; Duncan, 2019). Droughts decrease precipitation, increase soil moisture deficit and create water stress on the forest vegetation (Schönau and Schulze, 1984).

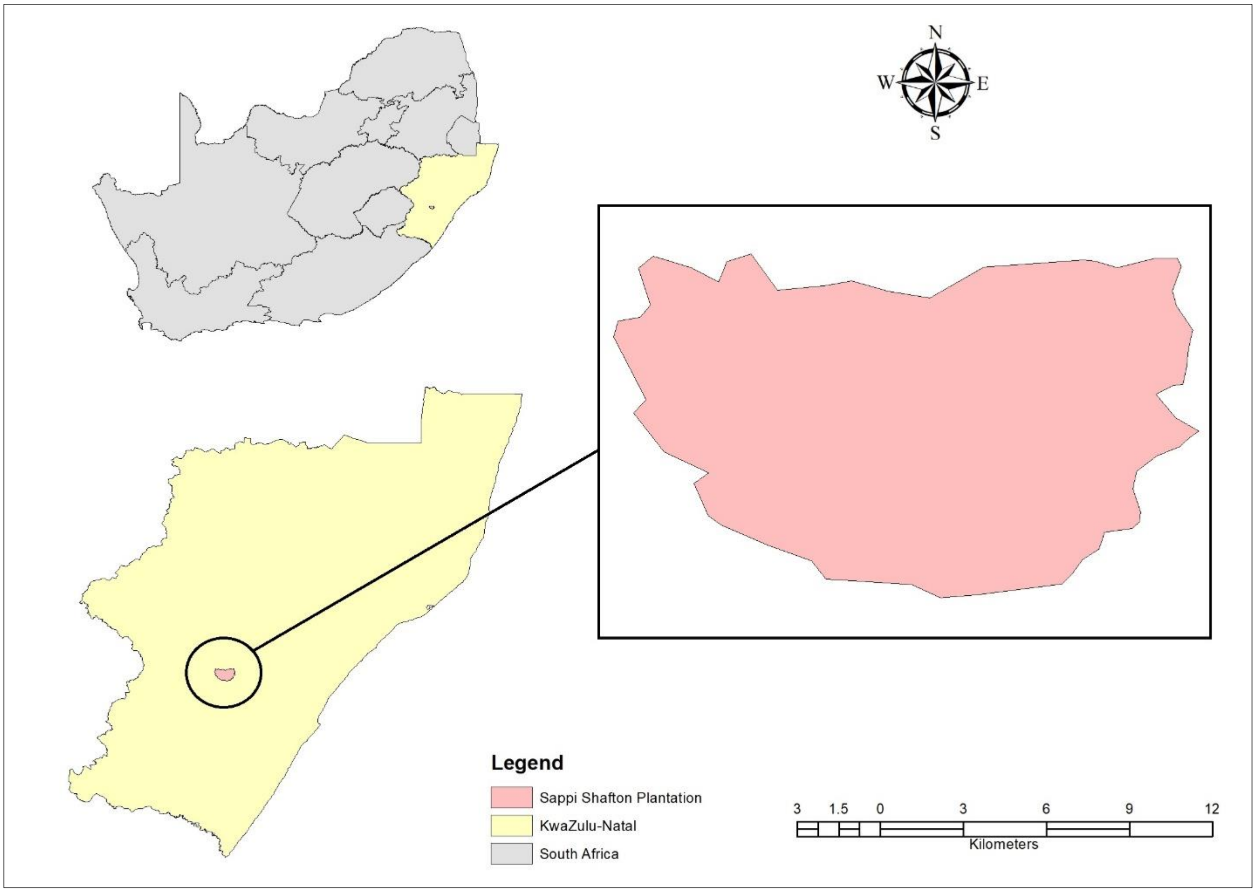


Figure 2.1 The location of the Sappi Shafton plantation within KZN, South Africa.

2.2.2 Data acquisition

Three Landsat sensors were used for this study, namely, Landsat 4 and 5, 7 and 8. Figure 2.2 shows the path to acquire scenes from these sensors freely from https://earthexplorer.usgs.gov/ which is an archive created and maintained by the USGS Earth Resources Observation and Science (EROS) Centre. The Landsat images were downloaded for January, February and December beginning from 2001 to 2018. Landsat 8 was launched in 2013. Hence, its scenes were downloaded from 2014 to 2018. The three selected months were used because it is during these months that the

Natal Midlands gets most of its annual rainfall. Comparing these three months, February had to some extent more usable scenes.

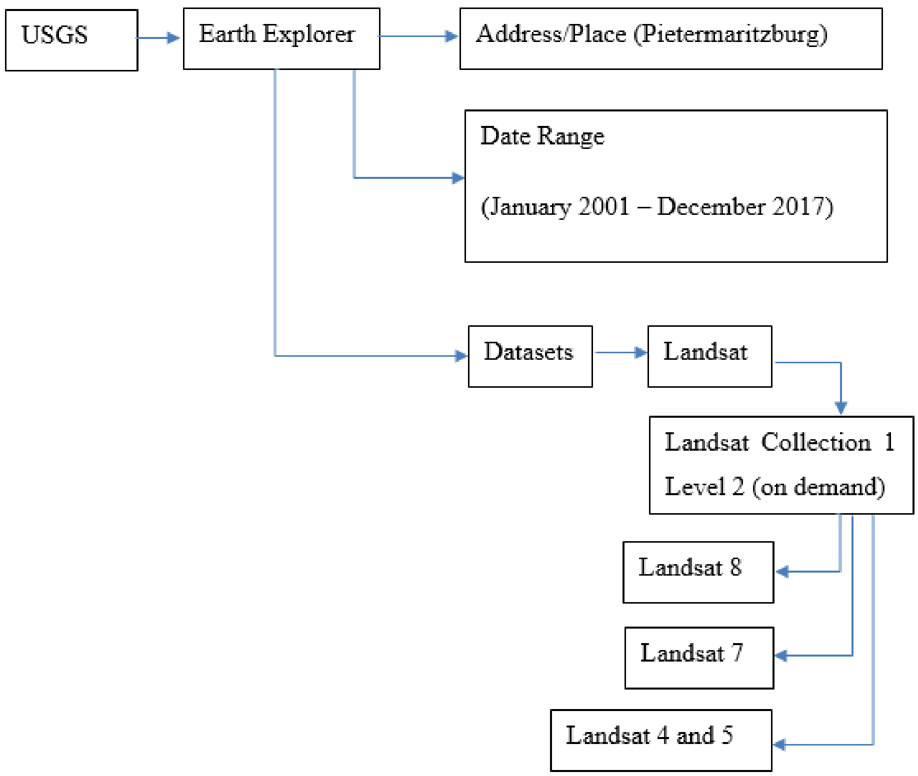


Figure 2.2 Path used to acquire data from USGS.

Table 2.1 shows different electromagnetic spectrum frequencies measured by the Landsat sensors utilised for this study. Frequencies are separated into ranges that are then designated into bands. These bands include red, blue, green, near-infrared (NIR), short-wavelength infrared (SWIR) and thermal infrared (TIR). The bands that are bold in Table 2.1 were used to calculate NDMI, NDVI, normalised difference water index (NDWI), enhanced vegetation index (EVI), soil adjusted vegetation index (SAVI), simple ratio (SR) and green channel index (GCHL).

Table 2.1 Landsat 4 and 5 Thematic Mapper (TM), 7 Enhanced Thematic Mapper Plus (ETM+) and 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) band designations (Zhu *et al.*, 2016).

Landsat 4 and 5 (TM)			**Landsat 7 (ETM+)**			**Landsat 8 (OLI & TIRS)**		
Band Description	Wavelength (µm)	Resolution (meters)	Band Description	Wavelength (µm)	Resolution	Band Description	Wavelength	Resolution
Band 1 - Blue	0.45 – 0.52	30	Band 1 - Blue	0.45 – 0.52	30	Band 1 – Ultra Blue	0.43 – 0.45	30
Band 2 - Green	**0.52 – 0.60**	**30**	**Band 2 - Green**	**0.52 – 0.60**	**30**	Band 2 - Blue	0.45 – 0.51	30
Band 3 - Red	**0.63 – 0.69**	**30**	**Band 3 - Red**	**0.63 – 0.69**	**30**	**Band 3 - Green**	**0.53 – 0.59**	**30**
Band 4 - NIR	**0.76 - 0.90**	**30**	**Band 4 - NIR**	**0.77 – 0.90**	**30**	**Band 4 - Red**	**0.64 – 0.67**	**30**
Band 5 - SWIR 1	1.55 – 1.75	30	Band 5 - SWIR 1	1.55 – 1.75	30	**Band 5 - NIR**	**0.85 – 0.88**	**30**
Band 7 – SWIR 2	2.08 – 12.50	30	Band 7 – SWIR 2	2.09 – 2.35	30	Band 6 – SWIR 1	1.57 – 1.65	30
			Band 8 - Pan	0.52 – 0.90	15	Band 7 – SWIR 2	2.11 – 2.29	30
						Band 8 - Pan	0.50 – 0.68	15
						Band 9 - Cirrus	1.36 – 1.38	30
			Band 61 - TIR	10.40 – 12.50 (low gain)	60	Band 10 - TIRS	10.60 – 11.19	100
Band 6 - TIR	10.40 – 12.50	120* (30)	Band 62 - TIR	10.40 – 12.50 (high gain)	30	Band 11 - TIRS	11.50 – 12.51	100

One issue encountered during data acquisition was that most scenes were unusable due to cloud cover. This is, however, a known problem as it was previously alluded by Asner (2001) and Schleeweis *et al.* (2016) who stated that optical imagery is disadvantaged by cloud cover which persists even during dry conditions. It is for that reason; the three Landsat sensors were utilised with a probability that there will be fewer gaps in the data. EROS automatic cloud cover assessment (ACCA) algorithms were developed and evolved over the years to be compatible with Landsat-7 ETM+ and Landsat-4 and -5 TM images (Schleeweis *et al.*, 2016). However, it should be noted that images that were processed before January of 2009 possess inaccurate ACCA scores. Therefore, Landsat images that had no cloud cover nor cloud shadows were utilised to obtain the results.

2.2.3 Image pre-processing

Images obtained from Landsat are susceptible to distortions caused by atmospheric, solar, sensor and topographic effects (Young *et al.*, 2017). To make sure the images utilised in this study did not have these distortions, the images were pre-processed through geometric, solar and absolute radiometric correction using ArcGIS version 10.4. Schroeder *et al.* (2006) argued that atmospheric correction could introduce new errors. Therefore to avoid incurring these errors, Young *et al.* (2017) recommended that high – level products be used instead of low-level products. Therefore, to avoid introducing new errors to the data extracted from the Landsat imagery through atmospheric correction, high-level products were downloaded from the USGS database for this study.

2.2.4 Data processing

The indices used in this study were selected based on their ability to monitor drought impacts on vegetation with information derived from {Svoboda, 2016 #61}. Also, all the indices were applied on a threshold of one to zero with zero signifying drought impact and one indicating no drought.

a) Conditional Drought Indices

Landsat scenes were first clipped using the boundary of the Sappi Shafton plantation. NDVI and NDMI were then calculated with the aim of obtaining the required drought conditional indices namely, VCI and VMCI.

i. Normalised Difference Vegetation Index

The fraction of radiation that is photosynthetically active and absorbed by vegetation yields NDVI; this makes it usable for monitoring long-term vegetation change such as its density and health (Jiao *et al.*, 2016; Novillo *et al.*, 2019). NDVI can also aid the calculation of the proportion of vegetation (*Pv*) (Fils *et al.*, 2018).

$$NDVI = \frac{NIR - RED}{NIR + RED} \ldots\ldots (1)$$

Where NIR and RED represent the near-infrared band and the red band respectively. For Landsat 4 and 5 and 7, NIR and RED are band 4 and 3, respectively. For Landsat 8, NIR and RED are bands 5 and 4 respectively.

ii. Normalised Difference Moisture Index

NDMI is the indicator of soil moisture content (Sahu, 2014). It is calculated using the following equation:

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \ldots\ldots (2)$$

Where NIR and SWIR are the bands designated for near-infrared and short-wave infrared respectively.

iii. Vegetation Condition Index

Drought impacts are difficult to detect directly from NDVI data, thus VCI was developed by Kogan (1995b) who scaled NDVI values from zero to one utilising the minimum and maximum NDVI for each location (Zhuo *et al.*, 2016).

$$VCI = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \ldots\ldots (3)$$

Where $NDVI_{max}$ is the maximum NDVI and $NDVI_{min}$ is the minimum NDVI, calculated for the corresponding pixels in the same month from the entire NDVI records (2001–2018) and *NDVI* is the actual NDVI for a pixel in a particular month. When VCI is closer to zero, the vegetation is in poor condition, however, when closer to one the vegetation is in a good condition.

iv. Vegetation Moisture Condition Index

The VMCI algorithm is similar to the algorithm used to obtain VCI, and the difference is that instead of using NDVI when calculating VCI, NDMI is used to calculate VMCI.

$$VMCI = \frac{NDMI - NDMI_{min}}{NDMI_{max} - NDMI_{min}} \ldots\ldots (4)$$

Where $NDMI_{max}$ is the maximum NDMI and $NDMI_{min}$ is the minimum NDMI respectively, calculated for the corresponding pixels in the same month from the entire Landsat NDMI records (2001–2018) and *NDMI* is the NDMI value for a pixel in a particular month.

b) Vegetation Indices

Vegetation indices have been mostly utilised for the derivation of vegetation structural parameters, phenological monitoring and vegetation classification (Huete *et al.*, 1999). With NDVI already calculated, EVI, SAVI, NDWI, SR and GCHL were calculated to bring a total of six vegetation indices to be used to analyse droughts.

i. Enhanced Vegetation Index

EVI is regarded as the improved NDVI because of its ability to reduce soil and atmospheric noise simultaneously; hence EVI possesses improved vegetation monitoring capabilities (Liu and Huete, 1995; Huete *et al.*, 1999; Matsushita *et al.*, 2007; Tenny and Hoffman, 2019).

$$EVI = 2.5 \times \left(\frac{NIR-R}{L+NIR+6R-7.5B}\right) \ldots\ldots (5)$$

Where *NIR* is the near-infrared, *R* represents the band designated for red, *L* is the soil adjustment parameter which equals to 1 and 6 and 7.5 are constant values.

ii. Soil Adjusted Vegetation Index

SAVI is an index developed to reduce canopy background noise (Huete, 1988).

$$SAVI = \frac{(NIR-R)\times(1+L)}{(NIR+R+L)} \ldots\ldots (6)$$

Where *NIR* is near-infrared, *R* is red and *L* is determined according to the specific environmental conditions. When the vegetation is dense, *L* approaches 1, meaning there is less canopy background noise (Tenny and Hoffman, 2019).

iii. Simple Ratio

Han *et al.* (2019) theorised that leaves absorb more red light compared to infrared light which meant with more leaves in the canopy, the higher the ratio.

$$SR = \frac{NIR}{R} \ldots\ldots (7)$$

Where *NIR* is near-infrared and *R* is the band designated for red.

iv. Normalised Difference Water Index

NDWI is defined by Ling *et al.* (2003) as a dimensionless index which determines the availability of water in a vegetation surface. The equation to calculate NDWI was proposed by Nanda *et al.* (2018):

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)} \ldots\ldots (8)$$

Where *NIR* is near-infrared and *SWIR* is the band designated for short wave infrared.

v. Green Channel Index

GCHL Index estimates the changes in the state of vegetation using the variety of the green channel instead of using the normalised difference between NIR and R (Pedregosa *et al.*, 2011).

$$GCHL = \frac{NIR}{G} - 1 \ldots\ldots (9)$$

Where *NIR* is the band designated for near-infrared and *G* is the band designated for the green channel.

2.2.5 Reference data

Sappi conducted field visits to monitor the condition of forest vegetation in terms of drought damage within the 517 compartments in KZN. However, only 30 of the 517 compartments were within the Sappi Shafton plantation. Therefore, data from those 30 compartments were utilised as reference data for extracting information from the Landsat imagery. R-Studio version 1.2.1335 was used to perform this extraction. The extracted information was then utilised by RTF and Kernel – SVM to perform drought damage classification within the plantation.

2.2.6 Statistical analysis

2.2.6.1 Kernel – SVM

SVM's are a powerful tool for diverse machine learning problems and outperform most conventional learning algorithms. However, they are not as compact as neural networks (Dutta *et al.*, 2015). SVM's also differ from neural networks in that they do not follow the empirical risk minimisation (ERM) rule which is focused on training error minimisation and often leads to overfitting problems (Yin and Yin, 2016). To overcome overfitting incurred during training error minimisation, SVM's follow the structural risk minimisation (SRM) rule which also improves generalisation abilities by restricting the complexity of the algorithm (Vapnik, 2013).

However, SVM's were initially proposed for binary classifications and assumed linear separation which rarely occurs in real-world tasks (Yin and Yin, 2016). Therefore, to broaden the applicability of SVM's to real world-tasks, advanced methods such as kernel clustering and kernel principal component analysis (KCPA) were introduced. These kernel (K) functions can separate non-linearly separable data by mapping it at a higher dimension. Kernel SVM's obtain more classes by utilising the "one versus one" reduction method where binary classifiers (K choose 2) are trained by training each classifier on the samples from a pair of classes and learn to distinguish between them (Hobbs, 2018).

2.2.6.2 Rotation forests

RTF is viewed as a typical supervised clustering method (Gu *et al.*, 2008). It is based on the decision tree method which is sensitive to the rotation of the feature axes which makes them more accurate compared to other decision tree based learning methods (Rodriguez *et al.*, 2006; Du *et al.*, 2015). That is because other decision tree based methods only split the class space vertically and horizontally which is less likely to produce better classification than performing rotation of the feature class (Kazllarof *et al.*, 2019). RTF generates the data set in an Independent Component Analysis (ICA) and PCA feature spaces and multiple classification trees are produced by utilising the data set produced in these feature spaces (Xia *et al.*, 2013; Gulácsi and Kovács, 2018). Rodriguez *et al.* (2006) selected PCA as a base for feature extraction because ensemble based on it performed better than ensemble based on random feature selection. However, Rodriguez *et al.* (2006) did acknowledge PCA limitations due to dimensionality reduction highlighted in the literature. To overcome reduced dimensionality, RTF randomly divides the feature into K subsets, and in each subset the linear transformation method is applied (Wang *et al.*, 2018). RTF also keeps all the principal components so that all the discriminatory information is preserved.

2.2.7 Algorithm parameter optimisation

The parameters evaluated for Kernel – SVM were cost (C) and gamma (γ) which are independent of each other and the algorithm returns the pair (C and γ) with the least error rate or best accuracy as the optimal parameter values (Sarkar *et al.*, 2016). According to Pedregosa *et al.* (2011), "the cost parameter defines the trading off of correct classification of training examples against maximisation of the decision function's margin and the gamma parameter defines the reach of a single training example". When C is high it means the algorithm prioritised error minimisation and when it is low, margin maximisation is prioritised (Nanda *et al.*, 2018). Low gamma means the single training sample is far-reaching and when high it does not reach far.

For RTF, the parameters that affect its performance are *K* and *L* which are the number of variable subsets and the number of base classifiers respectively (Wang *et al.*, 2018). *L* is a hyper parameter of RTF which can be optimised using cross-validation or a separate validation set and it is L that indicates how complex the ensemble is (Rodriguez *et al.*, 2006). Additionally, Wang *et al.* (2018) found that increased *L* and *K* lead to increased computational cost and compromise the classification accuracy while also noting that the accuracy is not affected by increasing *K* alone.

Therefore, a tenfold cross-validation method was utilised to determine optimal parameters for both Kernel – SVM and RTF. Once the parameters were optimised, the classification analysis was conducted based on 30 forest compartments within the study area to create clusters of forest trees similarly affected by droughts and those that were not. Forest trees affected by droughts were determined by employing Kernel-SVM and RTF with optimal parameters based on conditional drought indices and vegetation indices. The study period stretched between 2013 and 2018; that incorporated the onset in 2013, the intensification in 2015 and the offset in 2017 of the drought (Xulu *et al.*, 2019). To compute RTF and Kernel-SVM classifications, *svm* and *rotationForest* functions within the latest version of R-Studio application version 1.2.1335 were used and the overall methodology is presented in Figure 2.3.

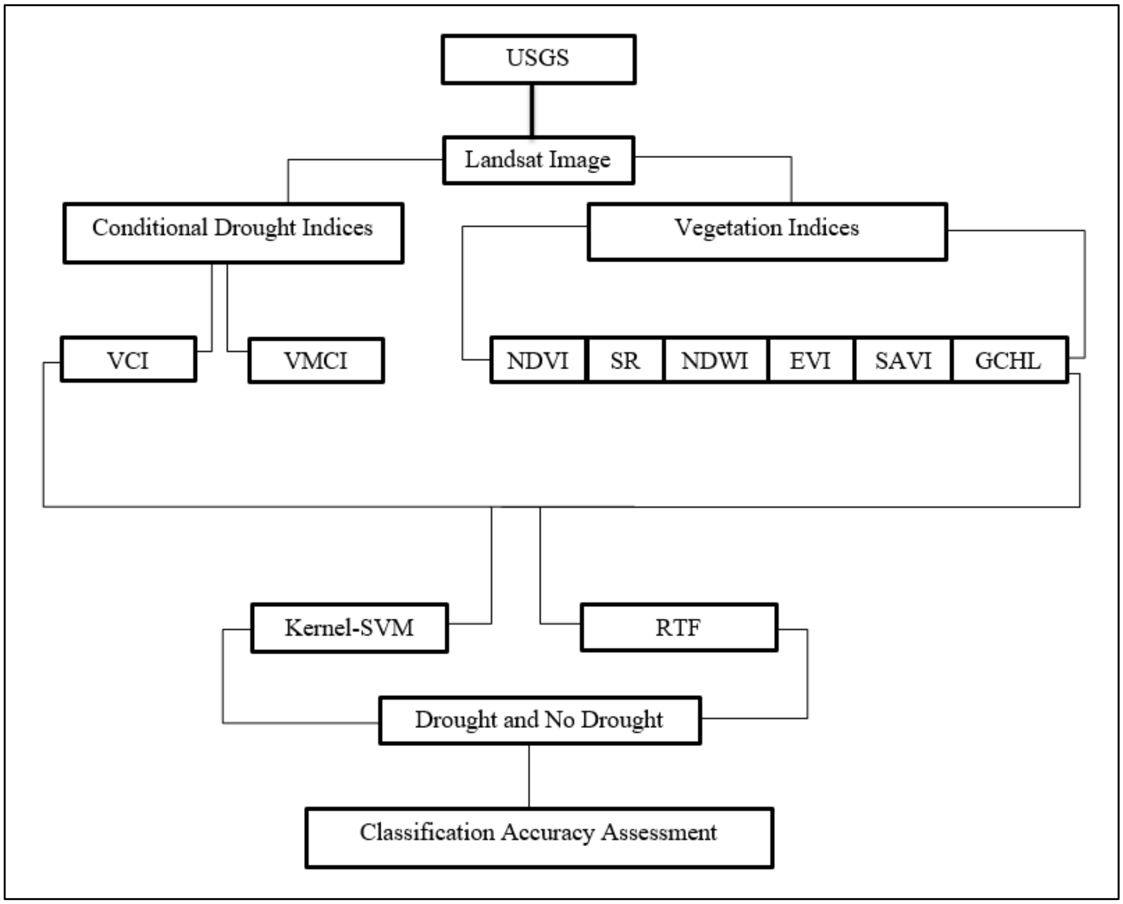


Figure 2.3 The procedure followed to obtain the result of the study.

2.2.1 Classification accuracy assessment

The classification accuracy assessments were conducted for the comparison of the performance of conditional drought indices and vegetation indices in detecting and mapping the extent of drought impacts over time using Kernel-SVM and RTF. Prior to training the model, the dataset (n = 240) extracted from the compartments for all tree species within the Sappi Shafton plantation was divided into 70% train and 30% test data. The data, both training and test data for drought damage (where level 1 represents the occurrence of drought and level 0 represents no drought) were resampled (10 fold and repeated 5 times) to ensure that the algorithm does not over fit the sample data (see Figure 2.4 and Figure 2.5). The Kernel-SVM and RTF were then calculated based on the classification results of the test dataset.

Figure 2.4 Drought damage data before resampling.

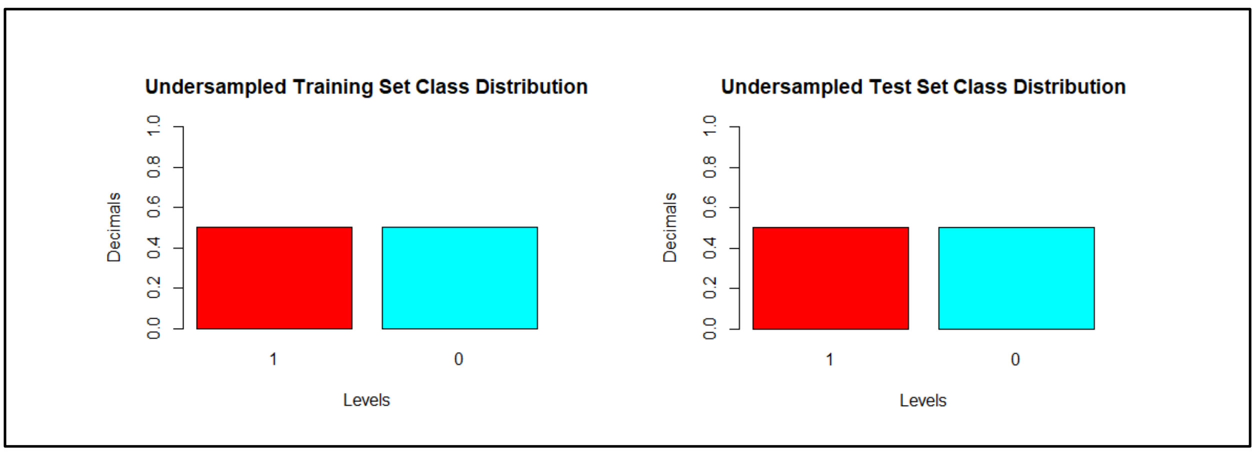


Figure 2.5 Drought damage data after resampling.

The OA from the confusion matrix, AUC and Cohen's Kappa coefficient were used as indicators of a good classification. The OA determines how often the model is accurate and will be utilised as the primary measure of accuracy for this study. The area under the receiver operating characteristic (ROC) curve is AUC which was computed using R Studio version 1.2.1335 and is considered the primary method of assessing the performance of predictive algorithms or models (Han *et al.*, 2019). Cohen's Kappa coefficient measures the performance of the classifier and compares it to how it would have performed merely by chance. Higher values of the Kappa coefficient ranging between 100% and 81 % indicate strong correlation and lower values that are less than 0% indicate no agreement at all.

2.3 Results

This section presents the results obtained from the classification accuracy assessment where the performance of Kernel - SVM was compared with RTF to classify droughts damage based on both conditional drought and vegetation indices.

2.3.1 Classification analysis

Kernel - SVM and RTF are both powerful algorithms capable of classifying and clustering non-linearly distributed data. Both Kernel - SVM and RTF were employed on identical datasets to determine which algorithm can accurately classify drought damaged trees from those that were not damaged based on conditional drought and vegetation indices. Figure 2.6 presents the decision boundaries for binary classification results obtained for both algorithms using conditional drought indices, where the points represents the actual observations and the area within the plot represents the predicted regions/boundaries where the algorithm predicts the trees will be impacted by droughts or not. The red points are trees observed to be impacted by droughts and the green points represent the trees that were not; the red area is where the algorithm predicts the trees will be impacted by droughts and the green area is where the algorithm predicts the trees will not be impacted by droughts.

Both algorithms were successful at classifying drought impacted trees from those that were not based on drought conditional indices; however, with a few points being misclassified. There was a considerable difference between the algorithms when predicting the area where trees not impacted by drought were found. Kernel - SVM predicted a relatively smaller area compared to RTF which predicted almost half the plot area.

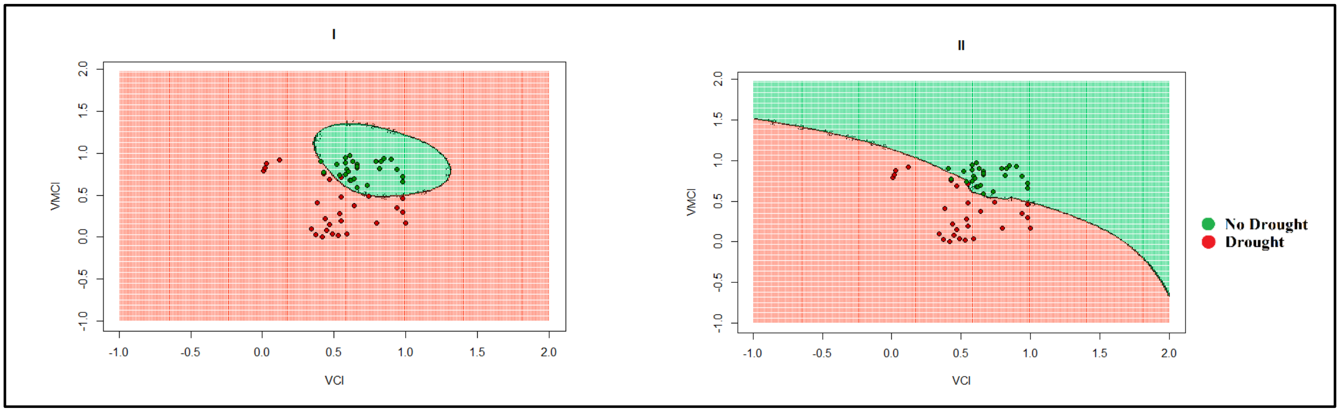


Figure 2.6 Kernel – SVM (I) and rotation forests (II) decision boundaries for conditional drought indices classification.

The algorithms were then used to classify drought damage data based on vegetation indices (Figure 2.7). The algorithms classification accuracy decreased compared to their performance when classifying drought damage using conditional drought indices. This was due to more points being outside the predicted decision boundaries where they belonged. The predicted areas were also different for both the algorithms with Kernel - SVM predicting a much smaller area for no drought occurrence compared to a larger area predicted by RTF.

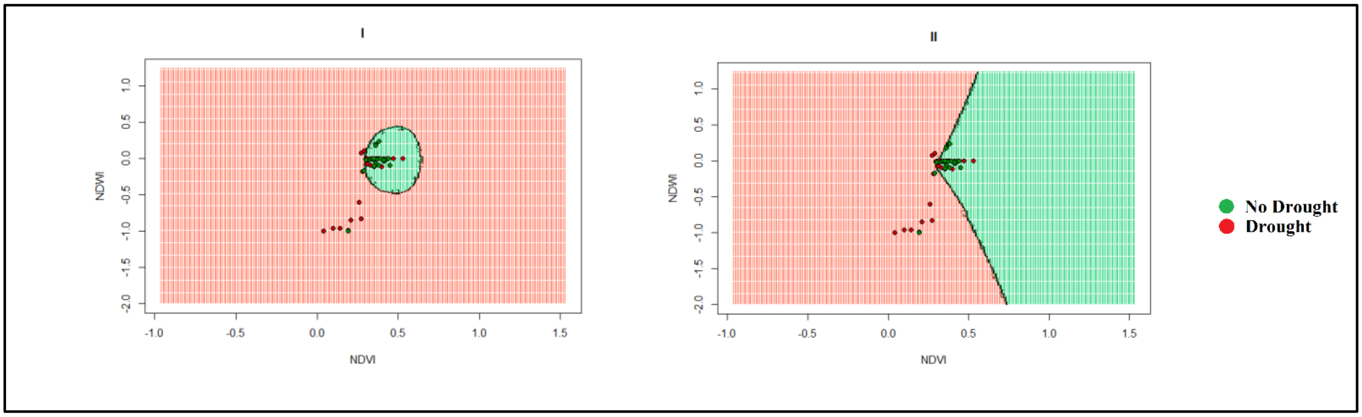


Figure 2.7 Kernel - SVM (I) and rotation forests (II) decision boundaries for the classifications based on vegetation indices.

Kernel - SVM and RTF were both successful at classifying the drought damage data; however, it is challenging to determine how accurate the classification was by looking at the decision boundaries alone. Therefore, the classification accuracy assessment was conducted.

2.3.2 Kernel – SVM and RTF optimal parameters

Table 2.2 presents the optimal parameter pairs for both algorithms and indices where the highest accuracy was obtained. The C for both applications of Kernel – SVM was 100% which meant error minimisation was emphasised. However, the gamma was different for conditional drought indices and vegetation indices. Kernel – SVM based on vegetation indices had a gamma of 16.7%, which means the single training example was far-reaching compared to 50% gamma for Kernel – SVM based on condition drought indices that implied a lower reach. RTF's K and L were identical for both indices at 2 and 3, respectively. These optimal parameters were then utilised to build RTF and Kernel – SVM classifiers based on conditional drought and vegetation indices.

Table 2.2 Algorithm parameters automatically optimised by both Kernel – SVM and RTF to generate predictions.

	Kernel SVM		Rotation Forests	
	Conditional Drought Indices	Vegetation Indices	Conditional Drought Indices	Vegetation Indices
Cost (C)	1	1	-	-
Gamma (γ)	0.5	0.167	-	-
K	-	-	2	2
L	-	-	3	3

2.3.3 Classification accuracy assessment

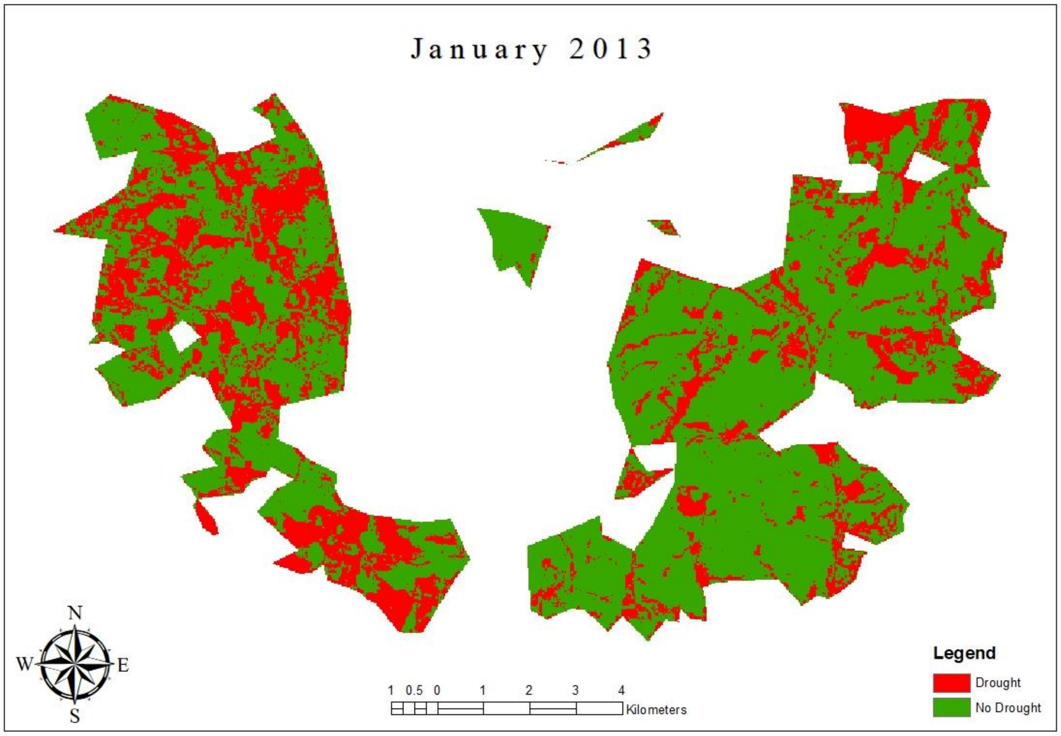
Table 2.3 presents the accuracy results obtained from Kernel - SVM and RTF when employed on both conditional drought indices and vegetation indices. RTF performed slightly better than Kernel – SVM when classifying based on conditional drought indices and arguably when classifying based vegetation indices. This conclusion was drawn from using OA as an indicator where Kernel - SVM had an OA of 94.44 % whereas RTF had 96.30% for conditional drought indices. However, for vegetation indices, they were almost equal where Kernel - SVM was 81.48 % and RTF was 81.84%. The AUC results for RTF classification were 96.3% compared to 94.4% obtained for Kernel – SVM based on conditional drought indices. This performance decreased when the classification was based on vegetation indices as both Kernel - SVM's and RTF produced an AUC of 81.5%.

Table 2.3 Results for Kernel -SVM and RTF accuracy assessment.

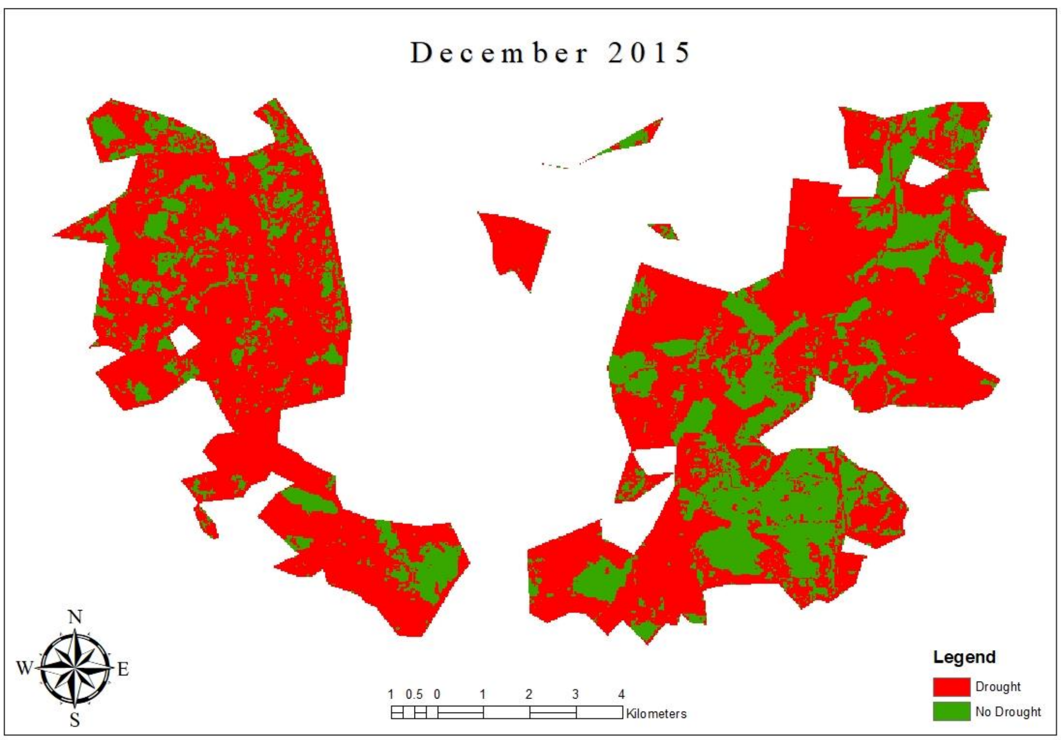
	Kernel - SVM		RTF	
	Conditional Drought Indices (%)	Vegetation Indices (%)	Conditional Drought Indices (%)	Vegetation Indices (%)
Overall Accuracy	94.44	81.48	96.3	81.84
AUC	94.4	81.5	96.3	81.5
Cohen's Kappa Coefficient	88.89	62.96	92.55	62.96

2.3.4 Drought damage maps

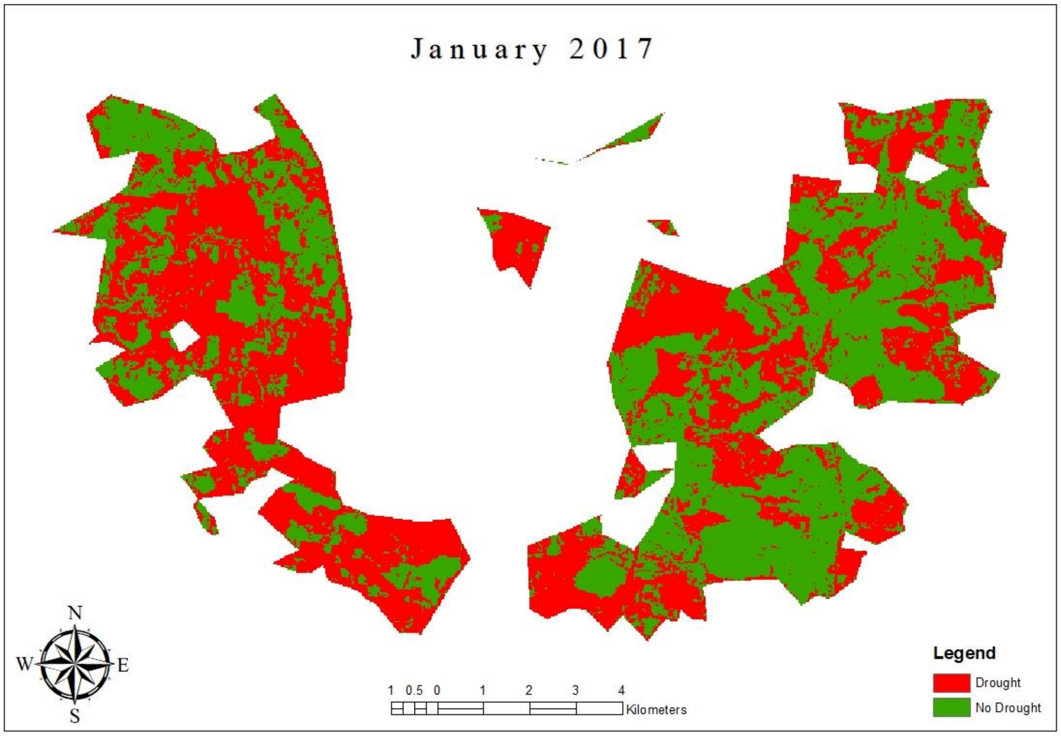
All drought damage maps were generated using ArcGIS version 10.4. Figure 2.8 and Figure 2.9 presents the resulting maps from the Kernel - SVM classification for both conditional drought and vegetation indices respectively. The maps include January 2013 (Figure 2.8a), which is probably the onset of the drought, December 2015 (Figure 2.8b) when the drought was very intense and January 2017 (Figure 2.8c) which was the probable offset period of the drought. From the predictive maps it can be observed that in 2013, there was less drought damage compared to 2015 when the drought was at its peak. January 2017 shows a considerable recovery by the forest vegetation given that the drought was ending.


January 2013
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

a)


December 2015
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

b)



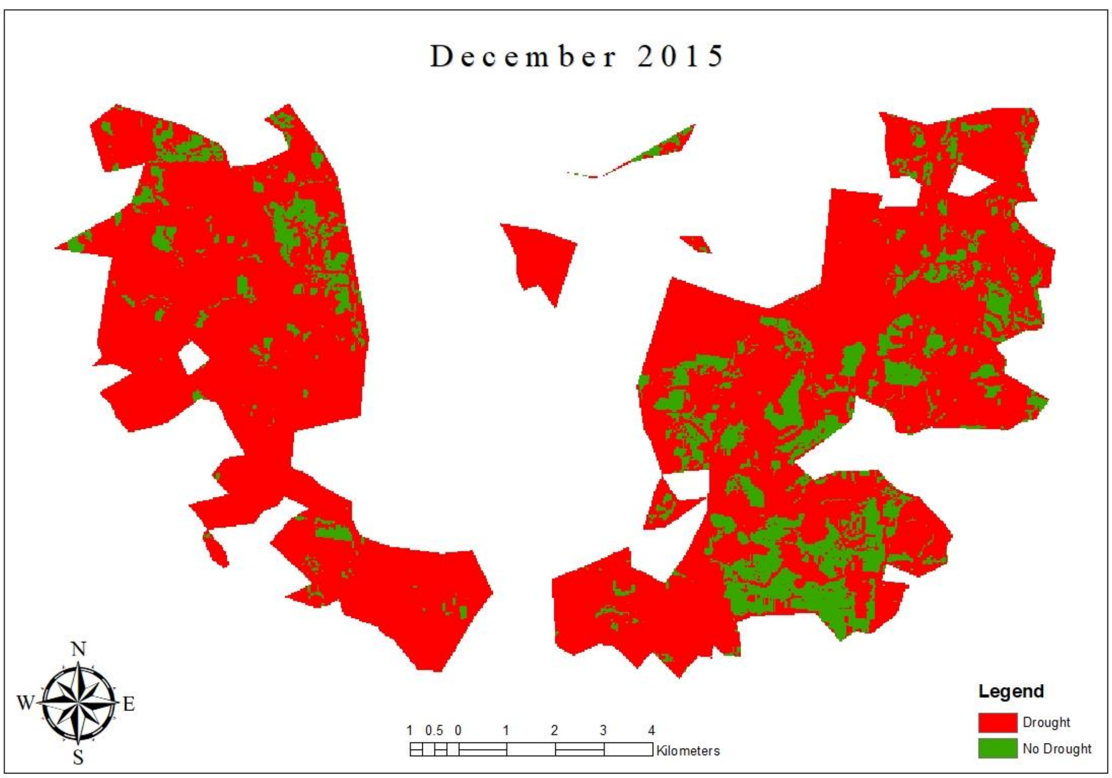
c)

Figure 2.8 Kernel - SVM classification maps based on conditional drought indices.

There are detectable differences between maps classified based on conditional drought indices (Figure 2.8) and those classified based on vegetation indices (Figure 2.8). The December 2015 map (Figure 2.8b) classified based on conditional drought indices contained more trees that were not impacted by droughts compared to the December 2015 map (Figure 2.9b) classified based on vegetation indices. There is also a noticeable difference between January 2017 maps (Figure 2.9c), where the January 2017 map classified based on conditional drought indices showed fewer trees that were not impacted by droughts compared to the January 2017 map (Figure 2.9c) classified based on vegetation indices which showed a huge improvement from what was observed by the end of 2015.


January 2013
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

a)


December 2015
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

b)

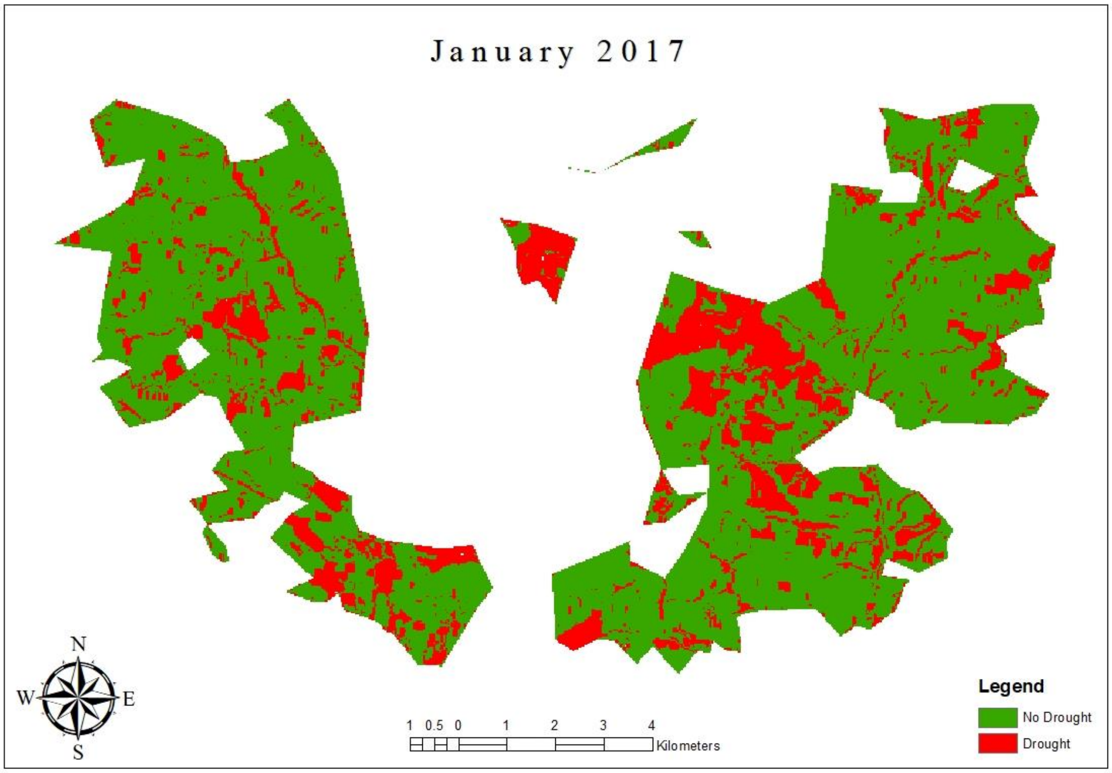
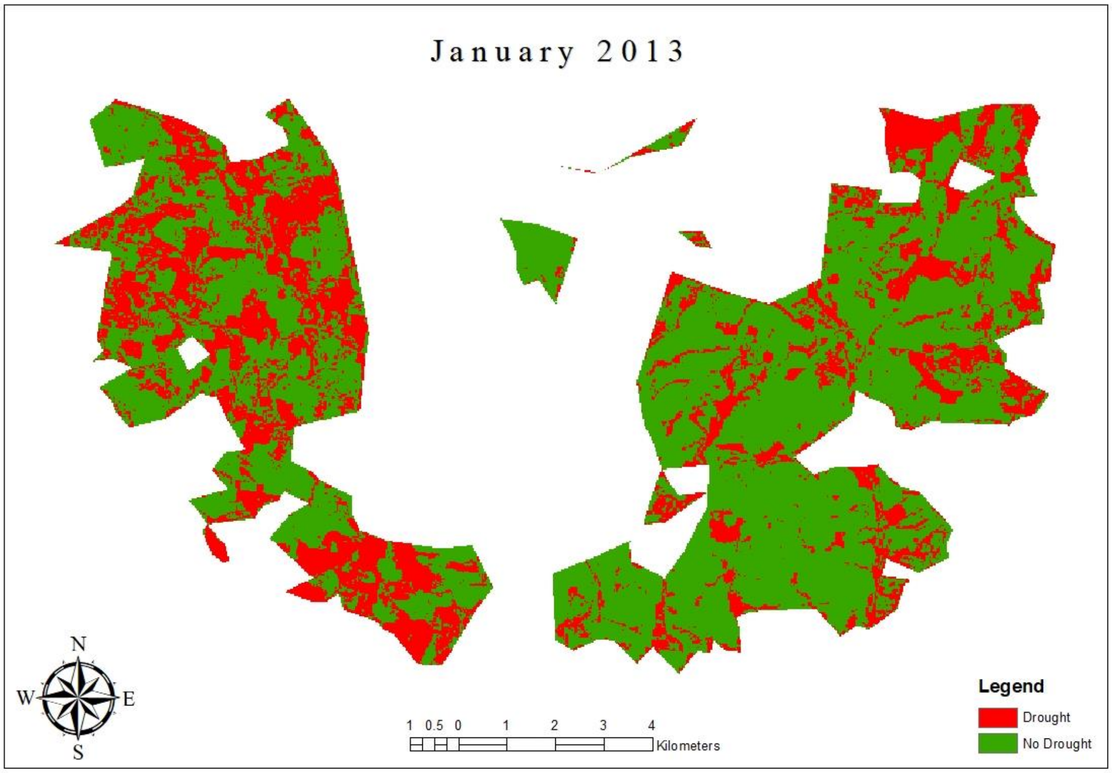
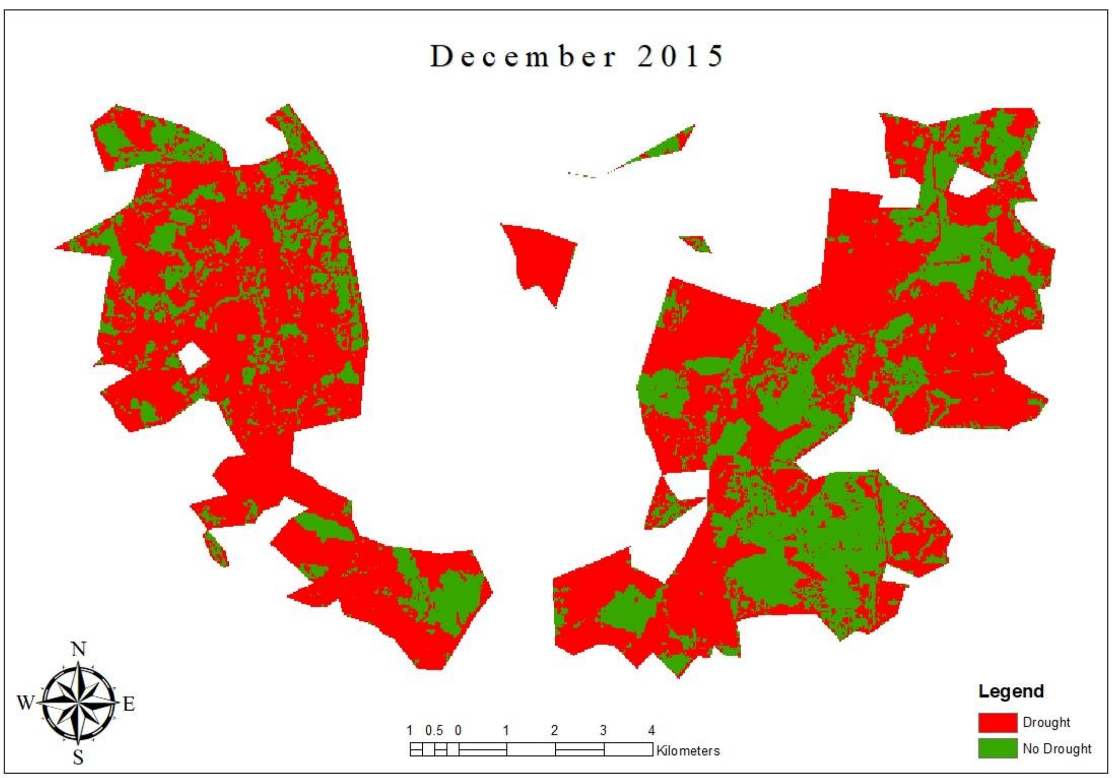
c) 

Figure 2.9 Kernel - SVM classification maps based on vegetation drought indices.

Figure 2.10 and Figure 2.11 present the resulting maps from the RTF classification for both conditional drought and vegetation indices, respectively. There are detectable differences in the drought damage maps classified using RTF. The December 2015 map (Figure 2.10b) classified based on conditional drought indices showed more non - drought impacted trees compared to the December 2015 map (Figure 2.11b) classified based on vegetation indices.


January 2013
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

a)


December 2015
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

b)

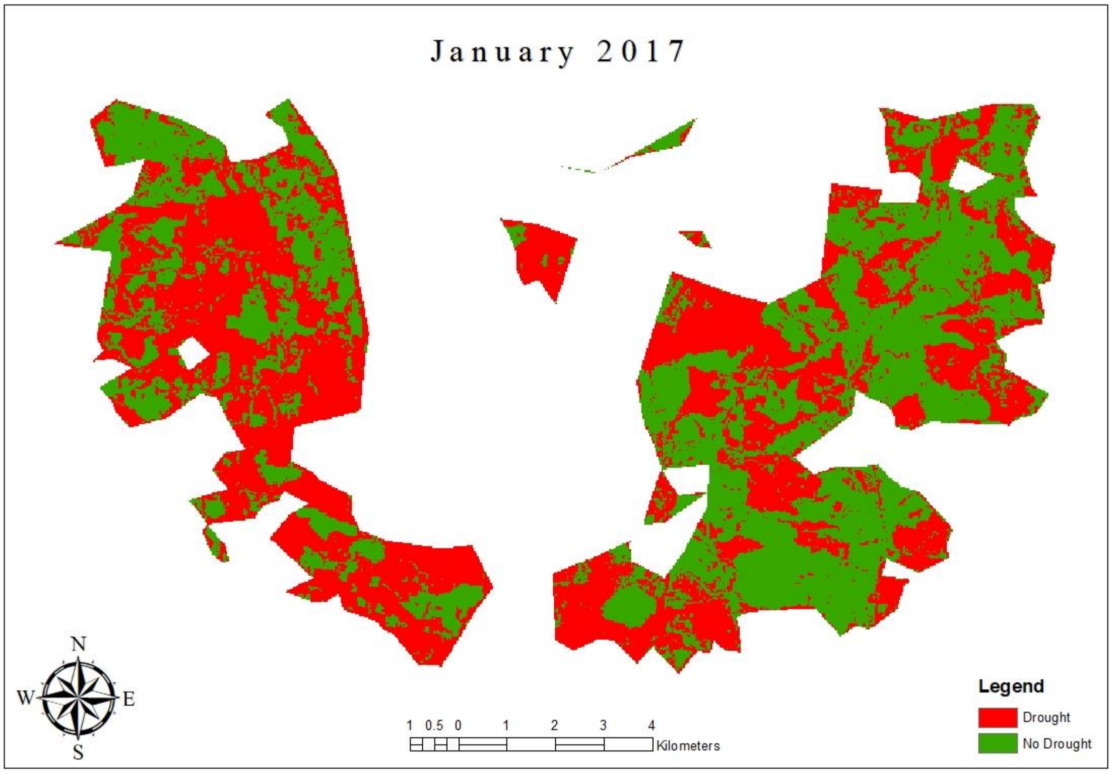
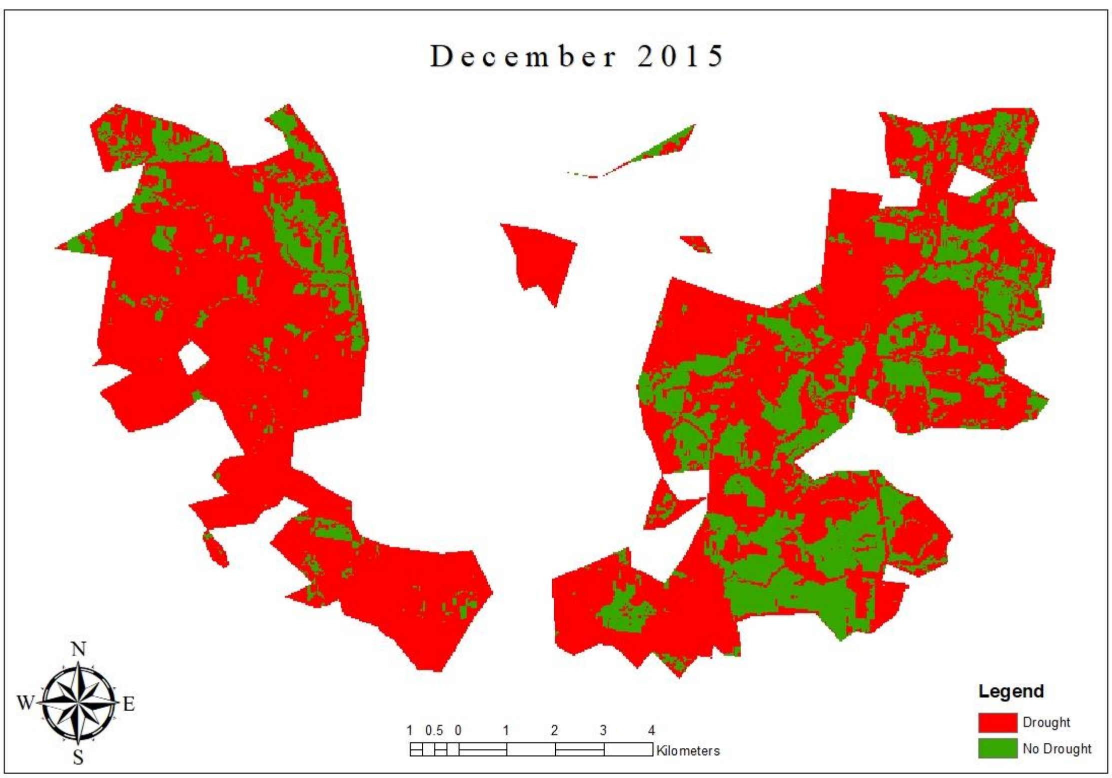
c) 

Figure 2.10 RTF classification maps based on conditional drought indices.

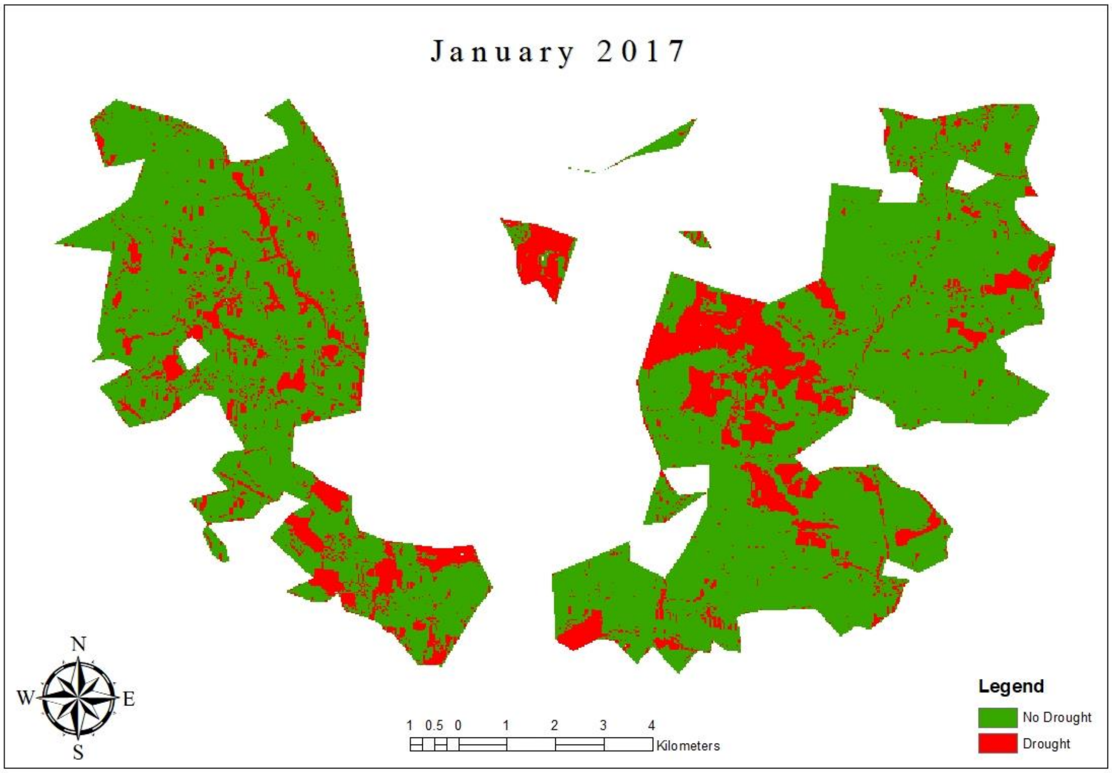
The differences are also detected for January 2017, where the January 2017 map (Figure 2.10b) classified based on conditional drought indices contained fewer trees that were not impacted by droughts compared to the January 2017 map (Figure 2.11b) classified based on vegetation indices.


January 2013
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

a)


December 2015
N
W
E
S
1 0.5 0 1 2 3 4
Kilometers
Legend
Drought
No Drought

b)



c)

Figure 2.11 RTF classification maps based on vegetation drought indices.

The overall results indicate that both Kernel – SVM and RTF classifications based on information derived from both conditional drought indices and vegetation indices can be utilised for analysing droughts and their impacts. It was also observed that when the algorithms were classifying based on conditional drought indices, they predicted more trees to be impacted by drought compared to when they were classifying based on vegetation indices.

2.4 Discussion

This study focused on machine learning as a tool for analysing droughts at a catchment scale. The results have shown that both MLAs, namely, Kernel – SVM and RTF are useful tools for analysing drought and their impacts on commercial forest vegetation. This is especially for RTF which showed a better accuracy compared to Kernel – SVM. However, RTF was relatively recently developed as an ensemble method by Rodriguez *et al.* (2006); hence it has been underutilised in studies focusing on drought assessment. The results also showed the potential for the application of conditional drought indices over vegetation indices for the analysis of droughts in forest ecosystems.

This study sought to explore the adaptation of Kernel – SVM and RTF for drought analysis, however, this required a thorough investigation of the accuracy of both these algorithms. Therefore, the ability of Kernel – SVM and RTF to accurately classify drought-impacted trees and trees that were not impacted by drought was assessed using the OA, AUC and the Cohen's Kappa index. The accuracy of these algorithms was investigated based on both conditional drought and vegetation indices. Using multiple indices ensured the reliability of the results as opposed to using a single index which Hao and AghaKouchak (2013) deemed insufficient for the classification and assessment of droughts. Similarly to our findings, Du *et al.* (2015) and Khamar and Eftekhari (2018) found RTF to perform better than the SVM algorithm.

Multiple studies including Zhu and Woodcock (2012), Zerrouki and Bouchaffra (2014), Han *et al.* (2015) and Chandrasekar *et al.* (2017) have utilised the confusion matrix and measures derived from it as the main accuracy assessment method. Ye *et al.* (2018) reviewed 209 studies for their study and determined that 146 studies used OA, which ranged between 42% and 96% as a measure of accuracy. The OA for this study was 94.44 % for Kernel – SVM and for RTF it was 96.30% when classifying based on conditional drought indices. For vegetation indices there was an observable decrease to the performance where Kernel - SVM was 81.48 % and RTF was 81.84%. These OA results demonstrate an excellent performance from both algorithms when classifying based on conditional drought indices data in comparison with the 85% average and the range determined by Ye *et al.* (2018) in their study.

Similar to OA, in many studies, AUC is recognised as the preferred accuracy measure of performance of classification algorithms (Cortes and Mohri, 2005). The AUC results for RTF classification were 96.3% compared to 94.4% obtained for Kernel – SVM based on conditional drought indices. This performance decreased when the classification was based on vegetation indices as both Kernel - SVM's and RTF produced an AUC of 81.5%.

The results demonstrated by the OA and AUC show that conditional drought indices were more accurate when classifying drought trees compared to vegetation indices. It should be taken into consideration that though accurate, conditional drought indices can misrepresent the actual drought extent and intensity (Jiao *et al.*, 2016). However, that does not mean that conditional drought indices results are not reliable as Dutta *et al.* (2015) and Zambrano *et al.* (2016) found that these indices strongly correlate with the prevalent and widely used in situ SPI. Multiple studies have also demonstrated the significance of vegetation indices for analysing droughts. These studies include Orhan *et al.* (2014), Jordan and Mitchell (2015), Arganda-Carreras *et al.* (2017), Gulácsi and Kovács (2018) and Xulu *et al.* (2019), Ahmadi *et al.* (2019). Some of these studies also found that vegetation indices strongly correlate with SPI, but generally the three-month SPI.

However, Mutanga and Skidmore (2004) argued that NDVI hinders the accuracy of classifiers due to NDVI's saturation at densely vegetated areas. Therefore, this study added EVI, NDWI, SR, SAVI and GCHL in anticipation that these indices could counter the limitations of NDVI. The results obtained for this study were then compared with those obtained by Xulu *et al.* (2019) who employed the RF algorithm on NDWI to analyse drought damage; they obtained and OA of 87.7% , which was comparable to our findings. This meant adding vegetation indices did not improve the performance of both algorithms, especially given that RTF generally performs better classifications than RF (Khamar and Eftekhari, 2018).

Classifying based on conditional drought and vegetation indices proved the capability and effectiveness of both algorithms and highlighted the shortcoming of utilising vegetation indices.

Where the algorithms were not as accurate as when they classified based on conditional drought indices. Even though RTF performed better than Kernel – SVM, it can be argued that Kernel – SVM results were also reliable. This expectation is based on the findings by Kavzoglu *et al.* (2015) who found RTF to perform better than Kernel – SVM. The argument for Kernel – SVM is that the drought damage classification maps it produced (Figure 2.8 and 2.9) do not significantly differ from those produced by RTF (Figure 2.10 and 2.11). Also, drawing from the Agri SA (2016) report which stated that 2015 had a drought of very significant magnitude and De Jager (2016) finding that 2015 had the lowest rainfall since the drought of 1904, Kernel – SVM map outputs were able to highlight more drought damage in 2015 compared to 2013 and 2017. Henceforth, Kernel - SVM could be as useful and effective as RTF when employed on conditional drought indices. This creates a niche for machine learning, conditional drought indices and improving vegetation indices application in drought analysis and remote sensing as a discipline.

Moreover, it should be noted that the scale or scope of the study has not been thoroughly investigated in the literature in terms of how performing classifications at a catchment scale affects the accuracy of the algorithm. Most studies tend to analyse droughts at a continental or subcontinental (Yuan *et al.* (2018)), national (Kogan (1995b), Botai *et al.* (2019)) and provincial scale (Botai *et al.* (2016)). However, some studies have studied or analysed droughts at a regional or catchment scale; these include Masupha and Moeletsi (2017), Meshram *et al.* (2018), Masupha and Moeletsi (2018) and Xulu *et al.* (2019). These studies have demonstrated that positive results can be obtained from analysing droughts at a localised scale. However, more research is still required to determine the true potential and accuracy of localised drought analysis, especially for forested catchments.

In summary, this study has provided crucial insight and results for the capability of MLAs to classify droughts at a catchment scale. It is therefore recommended that future studies prioritise machine learning algorithms to analyse drought damage on forest vegetation; however, this does not mean they should be limited to forest vegetation only. In addition, it should not necessarily be only the two algorithms that were used in this study, and researchers should investigate the capability of other machine learning algorithms. Moreover, the results indicated that Landsat data could be a valuable tool for catchment drought analysis and conditional drought indices produce more accurate data compared to vegetation indices. Thus, the objectives that were set for this study were met.

2.5 Conclusion

Intense droughts such as the one experienced between 2015 and 2016 pose a considerable threat to commercial forests and their impacts should be analysed and understood. This study has thus, demonstrated that machine learning could be adapted and utilised for remote sensing drought analysis as RTF and Kernel – SVM produced accurate classifications for drought damaged trees and non-damaged trees. Furthermore, results from this study have also demonstrated that:

1. RTF was more accurate than Kernel – SVM in detecting and mapping drought.
2. Classifications based on conditional drought indices were more accurate than those based on vegetation indices.
3. Landsat imagery can be used for analysing drought damage at a catchment or regional scale due to their high spatial resolution.

However, more research is still required for the application of machine learning and artificial intelligence in remote sensing and drought analysis. More research should also go into vegetation indices, as to why they underperformed compared to conditional drought indices. Moreover, the scope of this study can be extended by analysing drought damage in individual tree species, using different sensors to validate the use of Landsat and by employing the methodology used in this study in a different forested catchment with a different climate and landscape.

3. CHAPTER THREE: COMPARING ROTATION FORESTS AND EXTREME GRADIENT BOOSTING FOR MONITORING DROUGHT DAMAGE ON KWAZULU-NATAL COMMERCIAL FORESTS

Abstract

This study acknowledged that droughts are a recurrent and regular feature in countries such as South Africa where they pose a significant threat on the economy by crippling commercial forests and the society which depends on them. This called for the forecasting and quantification of droughts; however, this is a difficult undertaking given the complexity of droughts. Hence this study explored the utilisation of RTF and XGBoost MLAs to classify drought damage in commercial forests in KZN using information obtained from MODIS derived vegetation and conditional drought indices. The results from this undertaking demonstrated that both algorithms are capable of accurately detecting trees that exhibit drought damage and those that do not, more so when the algorithms were classifying based on information derived from conditional drought indices which yielded an accuracy of 82% for XGBoost and 76% for RTF. However, when the algorithms were classifying using vegetation indices data, the performance of the algorithms decreased resulting in an accuracy of 69% for XGBoost and 72% for RTF. Overall, the results demonstrated that MLAs could be utilised for the classification of drought damage on forest vegetation. Additionally, the study showed that MODIS imagery could be used for MLA classification and the fact that it is freely available in the USGS archives provides an added opportunity for more research to be conducted for the utilisation and improvement of these MLAs.

Keywords: Indices, Rotation Forests, Drought, Extreme gradient boosting, Machine Learning

3.1 Introduction

Droughts are a non-selective natural disaster in that their occurrence can be in both high and low precipitation areas (Wilhite and Glantz, 1985; Wu *et al.*, 2015). However, in semi-arid and arid climates such as Southern Africa droughts are a recurrent and regular feature (Rouault and Richard, 2005; Edossa *et al.*, 2014; Xulu *et al.*, 2018). Droughts can have adverse and sometimes irreversible impacts on society, agriculture and ecology (Edossa *et al.*, 2014). Therefore, given the dependence of Southern African countries on agricultural and ecological products such as commercial forestry for economic gain, droughts pose a major threat to the economy of some of these countries (Wu *et al.*, 2015). This calls for the forecasting and quantification of droughts which requires that the duration, extent and intensity of this phenomenon be investigated simultaneously (Zucchini and Adamson, 1984; Wu *et al.*, 2015). However, this is not an easy undertaking given the complexity of droughts (Mera, 2018).

This complexity arises as a result of droughts creeping nature in that it is difficult to detect their onset or their ending and to determine their severity (Wilhite, 2005). The lack of a definitive definition further compounds this complexity (Wilhite and Glantz, 1985; Mera, 2018). However, there is a consensus in most literature that droughts are defined by a decrease in or a lack of precipitation over an extended period (Wilhite, 2005; Solh and van Ginkel, 2014; Mera, 2018; Xulu *et al.*, 2018; Bayissa *et al.*, 2019). Droughts also are categorised into four types, namely; hydrological, socioeconomic, agricultural and meteorological droughts. Hydrological drought is concerned with the depletion of water supplies in lakes, dams and streams. Meteorological drought relates to decreased rainfall, agricultural drought is related with soil water deficit that leads to decreased agricultural productivity and socioeconomic droughts are concerned with the hindrance of supply of economic goods (Zucchini and Adamson, 1984; Edossa *et al.*, 2014). This study focused on the agricultural drought given that it strongly correlates with the objectives of this study.

The complexity of droughts is also enhanced by the interactions between these different drought types. The commercial forestry industry and therefore the economy of South Africa is arguably susceptible to all these drought types. The susceptibility of commercial forests was highlighted by Sun *et al.* (2018). They stated that as the earth continues to warm, drought induced tree mortality is also increasing and this is recognised as a crucial economic and ecological issue. Droughts directly influence forest vegetation mortality by increasing evaporative water demand through increased temperatures which results in tree water stress and indirectly enhancing tree mortality by making trees more vulnerable to destructive pathogens and insects that thrive under these conditions (Hope *et al.*, 2014; Millar and Stephenson, 2015; Xulu *et al.*, 2018). However, trees during growth have different responses to droughts, therefore some tree species survive a drought while others do not (Sun *et al.*, 2018). Researching and investigating the impacts of droughts on commercial forests will therefore aid the identification of tree species most vulnerable and those that are most resistant to droughts (Martínez-Vilalta *et al.*, 2012). The identification of such trees will, therefore, aid the management and the transition to tree species best adapted to droughts, which would decrease economic losses incurred due to droughts (Millar and Stephenson, 2015).

Multiple indices have been developed to investigate, quantify and classify drought impacts while keeping in consideration the complexities and constraints associated with droughts. As a result, most drought indices are only applicable in a specific region for a specific application (Heim Jr, 2002; Zhuo *et al.*, 2016). However, remote sensing-based drought indices are not constrained by location or application hence they are most suitable and applicable for drought analysis even in the complex forest vegetation. Remote sensing indices determine the state or condition of vegetation through the characterisation of the vegetation area which includes monitoring leaf area coverage, biomass and growth status (Wu *et al.*, 2015). These indices can be derived from MODIS onboard both Terra and Aqua satellites or Landsat sensors. MODIS will be the only sensor utilised to derive drought indices used in this study due to its high temporal resolution with 36 spectral bands that covers the entire Earth surface; however, it possesses a relatively low spatial resolution (Wu *et al.*, 2013; Wu *et al.*, 2015; Jiao *et al.*, 2016).

Determining the impacts of droughts requires the understanding of forest biomass dynamics over time, therefore having access to multiple observations within a year from MODIS can reveal varying forest temporal patterns (Gao *et al.,* 2019). This will allow for the differentiation of natural forest processes such as regrowth from drought-induced processes (Dutrieux *et al.*, 2015). Droughts limit the photosynthetic capacity of the tree and therefore, its growth capacity by reducing its leaf area index (LAI) (Jiao *et al.*, 2016). Vegetation indices such as NDVI and EVI are sensitive to LAI changes which makes them useful for detecting the impacts of droughts on forest vegetation (Tucker, 1979; Huete *et al.*, 2002). In addition to NDVI and EVI, there are multiple indices capable of monitoring droughts such as the NDVI derived VCI (Kogan, 1995a), NDWI (Gao, 1996), VHI and TCI (Kogan, 1995b) amongst others (Wu *et al.*, 2013). However, droughts are a complex phenomenon and they cannot be investigated using one or arguably two indices; multiple indices should be employed to capture different aspects of droughts and therefore determine their impacts accurately (Wu *et al.*, 2013; Zhang *et al.*, 2013; Hao and Singh, 2015).

As a result, this study will utilise three vegetation indices (EVI, NDVI, NDWI) and five conditional drought indices (TCI, VCI, vegetation moisture condition index (VMCI), VHI (Kogan, 2002) and enhanced vegetation condition index (EVCI)). These indices have been employed in various drought studies under varying environmental and climatic conditions (Rojas *et al.*, 2011; Wu *et al.*, 2013; Zambrano *et al.*, 2016). This study, therefore, seeks to extend the application of these indices by utilising them in the classification of drought-impacted trees and non-impacted trees within all commercial forests in KZN. This will be done using RTF and XGBoost. The utilisation of MLAs in this study is bounded on the expanding interest and need to select an accurate ensemble classification algorithm in remote sensing (Colkesen and Kavzoglu, 2017).

The selection of RTF and XGBoost, therefore, takes into consideration the wide utilisation of non-parametric supervised MLAs such as SVMs and RF in remote sensing classifications even though they somewhat produce inferior accuracy results (Colkesen and Kavzoglu, 2017; Georganos *et al.*, 2018). Kuncheva and Rodríguez (2007) investigated the potential and optimality of RTF and they found it to yield higher accuracies than RF, bagging and AdaBoost. Whereas XGBoost developed by Chen and Guestrin (2016) is optimised primarily for large tree structures and yields good classifications and at high execution speeds (Sandino *et al.*, 2018). This is one of the very few studies to utilise XGBoost and to our knowledge, it is one of the few studies to utilise this MLA in the classification of forest vegetation impacted by droughts.

Given the information presented above, this study, therefore, aims to classify the impact of droughts on commercial forests in KZN and compare the accuracy of RTF and XGBoost in meeting this aim. To achieve this, MODIS derived conditional drought and vegetation indices were used to enable the classification and visualisation of drought damaged and non - damaged trees using XGBoost and RTF. The objectives of this study were to expand the utilisation of MLAs in drought analysis, determining the accuracy of MODIS for classifying drought damage on forest trees and determining the accuracy of multiple conditional drought indices compared to vegetation indices. The primary contributions of this study include: (1) expanding the knowledge of drought impact on commercial forests; (2) extending information on the usability of conditional drought indices; (3) extending the use of MLAs on remote sensing data and (4) extending the information on the usability and accuracy of XGBoost in drought applications.

After the introduction in subsection 3.1, the rest of this paper is arranged in the following order; Subsection 3.2 outlines the methodology which consists of the detailed study area description, data acquisition and processing, classification analysis and accuracy assessment. Subsection 3.3 and 3.4 form the last part of this paper, which presents the results and discussion respectively. The concluding remarks are presented in subsection 3.5.

3.2 Methodology

3.2.1 Study area description

Commercial forests within the province of KZN in South Africa which has an area of 94361 km^2 (Figure 3.1) were the focus of this study. KZN is situated on the eastern seaboard of the country and is characterised by a complex landscape due to steep environmental gradients that occur over short distances (Jewitt *et al.*, 2015; Jewitt, 2016). The landscape of KZN ranges from the eastern subtropical climates of the Indian ocean to the western mountainous climates of the Drakensberg escarpments which are over 3000 m above sea level (Jewitt, 2016). KZN is considered the wettest province in South Africa due to its mean annual precipitation of over 800 mm compared to less than 500 mm received by the entire country of South Africa (Jewitt *et al.*, 2015). These conditions arise because of the presence of a warm coastal current that provides humid air to the atmosphere (Jury, 1998; Jewitt *et al.*, 2015). KZN's mean annual temperature ranges between 8 and 23 ℃ (Jewitt, 2016). This makes the province favourable for agriculture, hence the domination of commercial timber plantations, sugar cane, subsistence and commercial crops over the landscape (Jewitt *et al.*, 2015).

However, KZN is susceptible to climate stresses such as droughts which have been recurring over the past decades and likely to persist over coming decades with increased frequencies and durations (Reid and Vogel, 2006; Hlahla and Hill, 2018). Drought stresses are fuelling the need for research into the impacts of droughts, especially in commercial forestry, which is one of the primary economic commodities in the province.

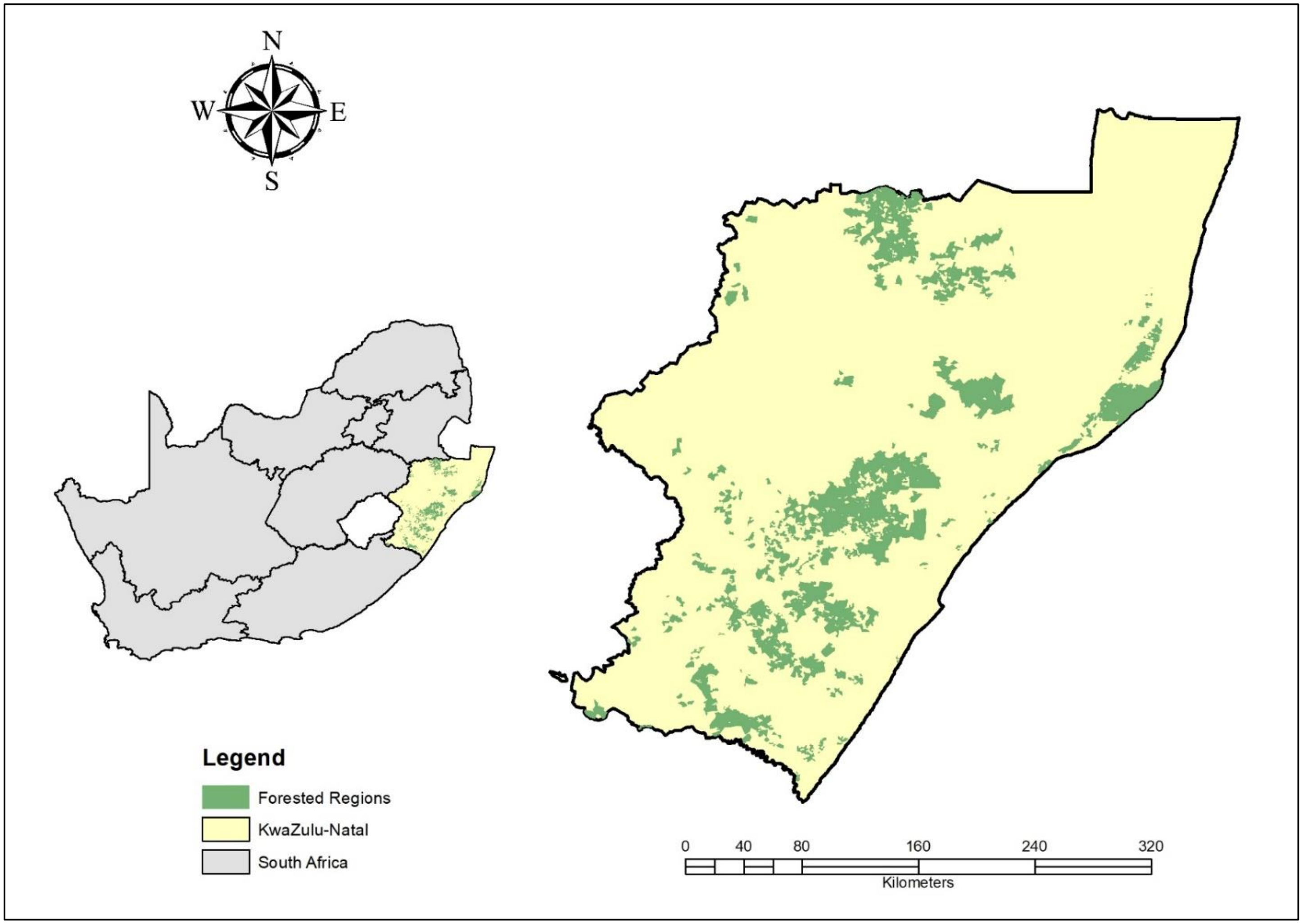


Figure 3.1 The geographical location of KwaZulu-Natal and forested regions within the province.

3.2.2 Data acquisition

Droughts are a creeping process characterised by decreased precipitation that results in increased surface temperatures and soil moisture deficits, thus, imposing stress on vegetation. Monitoring droughts, therefore, requires that parameters derived from precipitation, vegetation and the soil are taken under consideration (Du *et al.*, 2013). MODIS-Terra MOD13Q1 data was acquired from the USGS's Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) website (https://lpdaacsvc.cr.usgs.gov/appeears/). MOD13Q1 scenes were obtained for the year 2015 (January to December); which was due to the intensity of the drought over KZN in this year.

MOD13Q1 has a 16-day revisit cycle with a 250-meter spatial resolution which enables the sensor to analyse vegetation dynamics. The scenes from the sensor were generated using the maximum value compositing method (MVC) which reduces atmospheric and residual cloud effects (Wang *et al.*, 2019). This is significant because South Africa has a considerable cloud cover during the wet months which include January, February and December (Kruger, 2007). MOD13Q1 image band information is provided in Table 3.1 and the products derived from the image and utilised in this study are presented in Table 3.2.

From literature, EVI is considered to be more accurate than NDVI due to NDVI's sensitivity to high biomass conditions and canopy background variations and therefore produces saturated signals whereas EVI has improved sensitivity to such conditions (Huete, 1988; Huete *et al.*, 2002; Chen *et al.*, 2006). This suggests that the conditional drought index derived from EVI is more accurate at analysing drought impacts compared to VCI which is based on NDVI. However, when these indices were analysed using the Pearson correlation coefficient , VCI had a stronger correlation compare to EVCI which had no correlation with drought damage.

Table 3.1 MOD13Q1 image band information provided by USGS.

Science Dataset Name	Description	Units	Data Type	Fill Value	No Data Value	Valid Range	Scale Factor
250m 16 days NDVI	16-day NDVI	NDVI	16-bit signed integer	-3000	N/A	-2000 to 10000	0.0001
250m 16 days EVI	16-day EVI	EVI	16-bit signed integer	-3000	N/A	-2000 to 10000	0.0001
250m 16 days VI Quality	VI quality indicators	Bit Field	16-bit unsigned integer	65535	N/A	0 to 65534	N/A
250m 16 days red reflectance	Surface Reflectance Band 1	N/A	16-bit signed integer	-1000	N/A	0 to 10000	0.0001
250m 16 days NIR reflectance	Surface Reflectance Band 2	N/A	16-bit signed integer	-1000	N/A	0 to 10000	0.0001
250m 16 days blue reflectance	Surface Reflectance Band 3	N/A	16-bit signed integer	-1000	N/A	0 to 10000	0.0001
250m 16 days MIR reflectance	Surface Reflectance Band 7	N/A	16-bit signed integer	-1000	N/A	0 to 10000	0.0001
250m 16 days view zenith angle	View zenith angle of VI Pixel	Degree	16-bit signed integer	-10000	N/A	0 to 18000	0.01
250m 16 days sun zenith angle	Sun zenith angle of VI pixel	Degree	16-bit signed integer	-10000	N/A	0 to 18000	0.01
250m 16 days relative azimuth angle	Relative azimuth angle of VI pixel	Degree	16-bit signed integer	-4000	N/A	-18000 to 18000	0.01
250m 16 days composite day of the year	Day of year VI pixel	Julian day	16-bit signed integer	-1	N/A	1 to 366	N/A
250m 16 days pixel reliability	Quality reliability of VI pixel	Rank	8-bit signed integer	-1	N/A	0 to 3	N/A

Table 3.2 Vegetation and conditional drought indices utilised for this study.

Index	Equation	Label	Reference	Description
Vegetation Indices				
NDVI	$NDVI = \frac{NIR - RED}{NIR + RED}$	10	Tucker (1979)	NDVI is the fraction of radiation that is absorbed by vegetation.
EVI	$EVI = 2.5 \times \left(\frac{NIR - R}{L + NIR + 6R - 7.5B}\right)$	11	Huete *et al.* (2002)	EVI is regarded as the improved NDVI because of its ability to reduce soil and atmospheric noise simultaneously, hence EVI possesses improved vegetation monitoring capabilities.
NDWI	$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}$	12	Gao (1996)	NDWI is defined by Ling *et al.* (2003) as a dimensionless index which determines the availability of water in a vegetation surface.

Conditional Drought Indices				
VCI	$VCI = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}$	13	Kogan (1995b)	When VCI is closer to zero, that means the vegetation is in poor condition and when closer to one that means the vegetation is in good condition.
TCI	$TCI = \frac{LST - LST_{min}}{LST_{max} - LST_{min}}$	14	Kogan (1995b)	TCI assumes that during a drought soil moisture will decrease leading to land surface thermal stress.
VHI	$VHI = 0.5 \times VCI + 0.5 \times TCI$	15	Kogan (2002)	VHI combines temperature and vegetation indices.
VMCI	$VMCI = \frac{NDMI - NDMI_{min}}{NDMI_{max} - NDMI_{min}}$	16		NDMI is used to calculate VMCI.
EVCI	$EVCI = \frac{EVI - EVI_{min}}{EVI_{max} - EVI_{min}}$	17		EVI is used to calculate EVCI, hence it is assumed to be more accurate than VCI.
Note: NIR is near-infrared, SWIR is for short-wave infrared, R is the band designated for red, L is the soil adjustment parameter and LST is Land Surface Temperature				

3.2.3 Reference data

Field visits were conducted by Sappi to monitor and record the condition of forest vegetation in terms of drought damage within the 517 compartments in KZN. Data from these compartments were utilised as reference data for extracting information from the MODIS imagery. R-Studio version 1.2.1335 was used to perform this extraction. The extracted data was then utilised by RTF and XGBoost to perform drought damage classification across all forested regions in KZN.

3.2.4 Statistical analysis

3.2.4.1 Rotation Forests

RTF was proposed by Rodriguez *et al.* (2006) as a typical supervised classifier both as a selector and as a final exported classifier (Kazllarof *et al.*, 2019). This algorithm is based on the decision tree method, which is sensitive to the rotation of the feature axes, making them more accurate compared to other learning methods (Rodriguez *et al.*, 2006). RTF randomly splits the vector feature into disjoint subsets, and then a bootstrap technique is applied with a 75% resampling rate (Kazllarof *et al.*, 2019). An orthogonal transformation is then performed within a PCA feature space (Akar, 2018). Finally, RTF generates multiple classification trees by utilising the data set produced in the PCA feature space (Xia *et al.*, 2013; Gulácsi and Kovács, 2018). Performing rotations of the feature space provide the prospect of greater variety within the ensemble entities which makes RTF better than other decision tree methods which split the class only vertically and horizontally (Kazllarof *et al.*, 2019).

3.2.4.2 Extreme Gradient Boosting

XGBoost is extended from gradient boosting machines (GBMs) which have been applied widely on multiple remote sensing applications (Georganos *et al.*, 2018). GBMs have been suggested to be a powerful machine learning method and capable of performing scene classifications and estimating above ground biomass (Zhang *et al.*, 2015). However, GBMs have also been suggested to be highly susceptible to overfitting due to a lack of a robust regularisation framework. They require more parameters to be optimised compared to SVMs and RF (Georganos *et al.*, 2018). XGBoost is based on boosting which is the combination of all the set of weak learners to develop a strong learner through additive training strategies (Fan *et al.*, 2018). The final classification, therefore, includes the improvements of all the previous modelled trees (Georganos *et al.*, 2018). However, Xia *et al.* (2017) determined that XGBoost is limited by its poor capability to determine the optimal tree structure. The algorithm attempts to overcome this by employing the greedy search technique which is costly when processing data that is at a high dimension and scale (Bergstra *et al.*, 2011; Xia *et al.*, 2017).

3.2.5 Classification algorithm parameter optimisation

It is very rare that a classification algorithm does not possess hyper-parameters (Xia *et al.*, 2017). Hyper-parameters have a significant influence on the accuracy of the algorithm and this, therefore, introduces the need for tuning these parameters to get the best obtainable accuracy (Bergstra *et al.*, 2011). However, hyper-parameter is heavily criticised for the subjective judgement and trial and error method associated with this practice (Bergstra *et al.*, 2011; Georganos *et al.*, 2018). With the improvement in computation, hyper-parameter optimisation scarcely depends on human subjective judgement as improved computation allows for the tuning of parameters using grid search (GS) and automatic optimisation (Georganos *et al.*, 2018).

RTF parameters are *K* which is the number of variable subsets and *L*, which is the number of base classifiers (Rodriguez *et al.*, 2006). For this study, *K* and *L* were obtained using automatic optimisation. GS and automatic optimisation can be utilised to obtain the parameters for XGBoost, however, their feasibility is questionable given the substantial amount of hyper-parameters in XGBoost (Xia *et al.*, 2017). Therefore, for this study, the Bayesian optimisation explained in Bergstra *et al.* (2011) was utilised to optimise XGBoost parameters. Bayesian optimisation is considered an excellent parameter optimisation method capable of excellent results (Georganos *et al.*, 2018).

Therefore, this study employed both XGBoost and RTF with optimal parameters to classify drought-impacted trees and those that were not on all commercial forests within KZN. The study period considers the year 2015 which arguably included the onset and intensification of the 2015 – 2016 drought. To demonstrate drought damage over the year 2015, January, April, July, September and December drought damage classification results were mapped. The *xgboost* and *rotationForest* packages within R-Studio version 1.2.1335 were used to compute the classifications created by XGBoost and RTF. Figure 3.2 presents the overall methodology for this study.

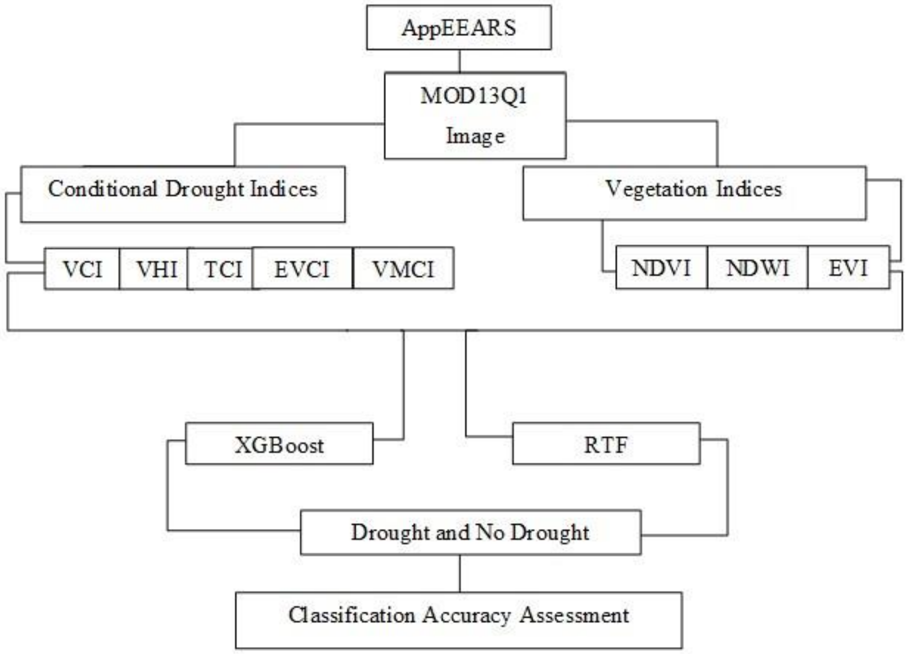


Figure 3.2 The procedure used to obtain results for this study.

3.2.6 Classification accuracy assessment

XGBoost and RTF could be successful at performing the desired classification, however, the accuracy of the classification should be assessed to determine the true potential of these algorithms. The classification assessment will assess the performance of both XGBoost and RTF when classifying based on both vegetation and conditional drought indices.

Therefore, before both algorithms were trained with drought data obtained from the KZN forest compartments, the dataset had two levels and was divided into 70% train and 30% test data. The levels in the data were level 1, which represented the occurrence of drought conditions and level 0 which meant no drought conditions were observed. To ensure that the algorithms do not over fit the sample data, drought damage data was resampled 10-fold and repeated 5 times. Both train and test data were under-sampled as seen in Figure 3.3 and 3.4 where 1 represent no drought and 2 represents the occurrence of drought. XGBoost and RTF were then employed on these data based on the vegetation and conditional drought indices.

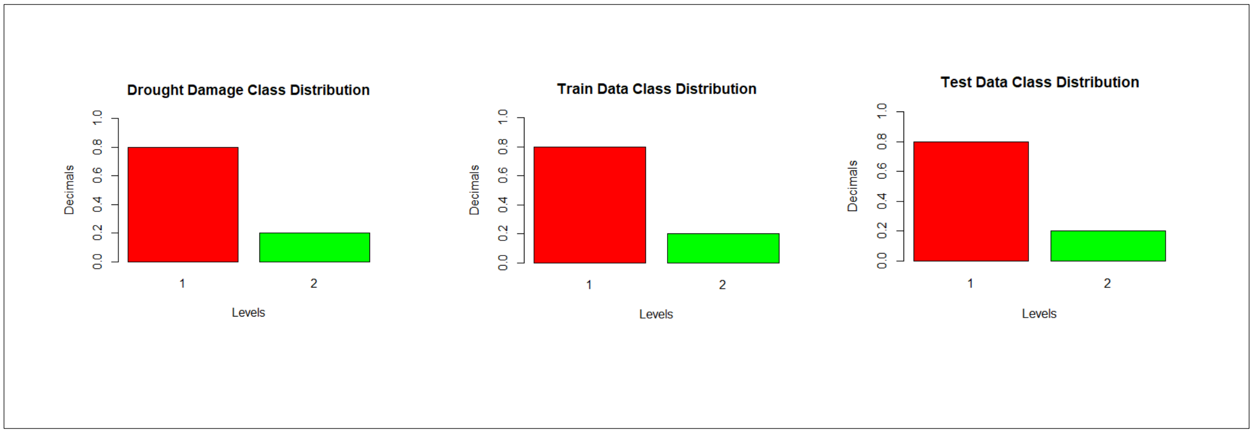


Figure 3.3 Drought damage data before resampling.

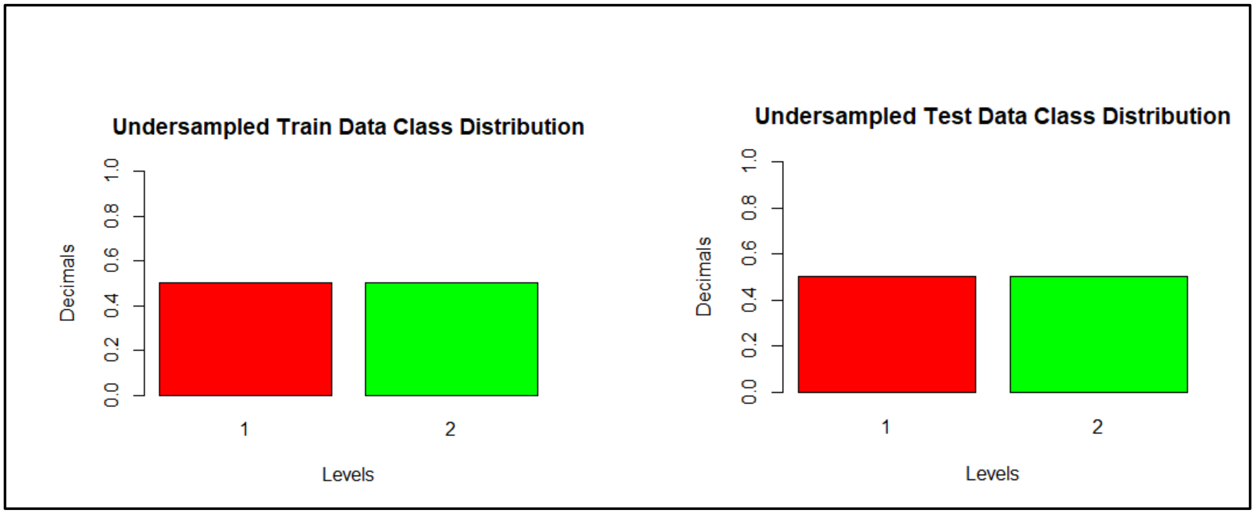


Figure 3.4 Drought damage data after resampling.

The percentage overall accuracy (OA), user and producer accuracy and Cohen's Kappa coefficient from the confusion matrix were specifically used to assess the performance of XGBoost and RTF.

3.3 Results

3.3.1 Classification analysis

This study aimed to determine the potential of XGBoost and RTF to separate trees exhibiting drought damage from those trees that do not show any damage. This meant drought-impacted trees and non-drought impacted trees were the two classes utilised for this study. XGBoost and RTF performed their classifications based on MODIS derived vegetation and conditional drought indices. Both algorithms were successful at performing these classifications, however, at varying accuracies. MLA classifications are generally visualised using decision boundary plots, however, given the large dataset used to train and test RTF and XGBoost it was difficult to determine if more or fewer points were inside or outside the boundary of the class they belong to.

3.3.2 XGBoost and RTF optimal parameters

Optimal parameters for XGBoost are presented in Table 3.3 and were obtained using Bayesian optimisation method. This method presented a combination of parameters that had the highest accuracy possible. These parameters were then used to build the XGBoost classifier based on both conditional drought and vegetation indices.

Table 3.3 XGBoost optimal parameters for both condition and vegetation drought indices obtained using the Bayesian optimisation method.

Parameter	Value
Subsample Ratio (r_s)	0.7991
Learning Rate (τ)	0.2905
Maximum Tree Depth (D_{max})	6
Minimum Child Weight (ω_{mc})	2
Column subsample ratio (r_c)	0.5

Table 3.4 and 3.5 show varying accuracies produced by different values of *K* and *L*. RTF classifiers based on conditional drought and vegetation indices were, therefore, built using a pair of *K* and *L* that produced the highest accuracy which is highlighted in the tables.

Table 3.4 RTF parameters when classifying based on conditional drought indices.

RTF - Conditional Drought Indices		
Parameters		Accuracy (%)
K	*L*	
1	3	76.44
1	6	76.80
1	9	76.93
5	3	76.68
5	6	77.00
5	9	77.28

Table 3.5 RTF parameters when classifying based on vegetation indices.

RTF - Vegetation Indices		
Parameters		Accuracy (%)
K	*L*	
1	3	69.34
1	6	69.75
1	9	69.87
3	3	69.49
3	6	69.62
3	9	69.79

3.3.3 Algorithm variable importance

During classification of trees affected by drought, the algorithms selected the most relevant variables for the classification. They disregarded the irrelevant variables which could have negative implications on the algorithm accuracy (Gregorutti *et al.*, 2017). Table 3.6 shows that when the algorithms were classifying based on conditional drought indices, RTF prioritised VCI where as XGBoost prioritised VHI. When the algorithms were classifying based on vegetation indices, EVI was prioritised by RTF and NDVI was prioritised by XGBoost.

Table 3.6 Overall variable importance for both RTF and XGBoost when classifying based conditional drought and vegetation indices.

RTF				XGBoost			
Conditional Drought Indices		Vegetation Indices		Conditional Drought Indices		Vegetation Indices	
Variable	Overall Importance (%)	Variable (%)	Overall Importance (%)	Variable (%)	Overall Importance (%)	Variable (%)	Overall Importance (%)
VCI	100	EVI	100	VHI	100	NDVI	100
VHI	47.24	NDVI	56.21	TCI	65.88	EVI	90.26
TCI	16.95	NDWI	0	EVCI	7.24	NDWI	0
VMCI	10.79			VMCI	3.68		
EVCI	0			VCI	0		

3.3.4 Classification accuracy assessment

The results in Table 3.7 indicate a somewhat considerable overall difference between the accuracies of algorithms when classifying based on conditional drought indices and when based on vegetation indices. When the algorithms were classifying based on conditional drought indices, they had a better performance than when utilising vegetation indices to perform their classifications. The OA was 76.05% for RTF when classifying based on conditional drought indices. This OA considers the user and producer accuracies for both drought and non-drought classes. When RTF was classifying based on conditional indices the user and producer accuracies for drought were 83% and 65.2%, respectively. For non-drought, user and producer accuracies were 71.6% and 86.7%, respectively. There was a decrease in performance when the algorithms were classifying based on vegetation indices with RTF decreasing from 76.05% to 70.7%. The OA also considers the user and producer accuracies for both drought and non-drought classes. When RTF was classifying based on vegetation indices the user and producer accuracies for drought were 71.7% and 68.3%, respectively. For non-drought, user and producer accuracies were 69.7% and 73.1%, respectively.

For XGBoost when classifying based on conditional drought indices the OA was 81.9%. This OA also considers the user and producer accuracies for both drought and non-drought classes. When XGBoost was classifying based on conditional indices the user and producer accuracies for drought were 84.8% and 77.6%, respectively. For non-drought, user and producer accuracies were 79.6% and 86.7%, respectively. The performance decreased when XGBoost was classifying based on vegetation indices with the OA decreasing from 81.9% to 69.3%. This OA also considers the user and producer accuracies for both drought and non-drought classes. When XGBoost was classifying based on vegetation indices the user and producer accuracies for drought were 70.1% and 67.2%, respectively. For non-drought, user and producer accuracies were 68.5% and 71.4%, respectively.

A considerable difference can also be observed in the Cohen's Kappa index where the algorithms had 63.87% and 52.04% for XGBoost and RTF respectively when classifying based on conditional drought indices. When classifying based on vegetation indices XGBoost and RTF had Cohen's Kappa indices of 38.58% and 41.34%, respectively. However, it should be noted that the criticism surrounding the randomness nature of the Cohen's Kappa index (Fassnacht *et al.*, 2014; Lottering *et al.*, 2020) made so that the conclusion from the results was not directly drawn from this index.

Table 3.7 Results for RTF and XGBoost accuracy assessment.

	RTF				XGBoost			
	Conditional Drought Indices		Vegetation Indices		Conditional Drought Indices		Vegetation Indices	
	No Drought (%)	Drought (%)	No Drought (%)	Drought (%)	No Drought (%)	Drought (%)	No Drought (%)	Drought (%)
User Accuracy	71.6	83	69.7	71.7	79.6	84.8	68.5	70.1
Producer Accuracy	86.7	65.2	73.1	68.3	86.3	77.6	71.4	67.2
Overall Accuracy	76.05		70.7		81.9		69.3	
Cohen's Kappa Coefficient	52		41.3		63.9		38.6	

The results presented by the measures derived from the confusion matrix imply a great potential for machine learning in analysing droughts and their impacts using conditional drought indices and raised more questions on the use of vegetation indices to perform the same task.

3.3.5 Drought damage classification maps

The drought damage maps presented in this subsection are a result of clustering using both RTF and XGBoost which were trained using 2015 vegetation and conditional drought indices data. Figure 3.5 show the drought damage clusters in commercial forested regions within KZN for every three months starting from January 2015 to December of the same year. These were constructed in R-Studio version 1.2.1335 using RTF based on conditional drought indices. From the maps it can be observed that less forested areas were impacted by drought in January compared to April and July, which showed more drought-impacted areas.

There are fewer observable differences in areas affected by drought in July to September; however, there is a remarkable improvement in forest vegetation in December as tiny patches indicate drought impact.

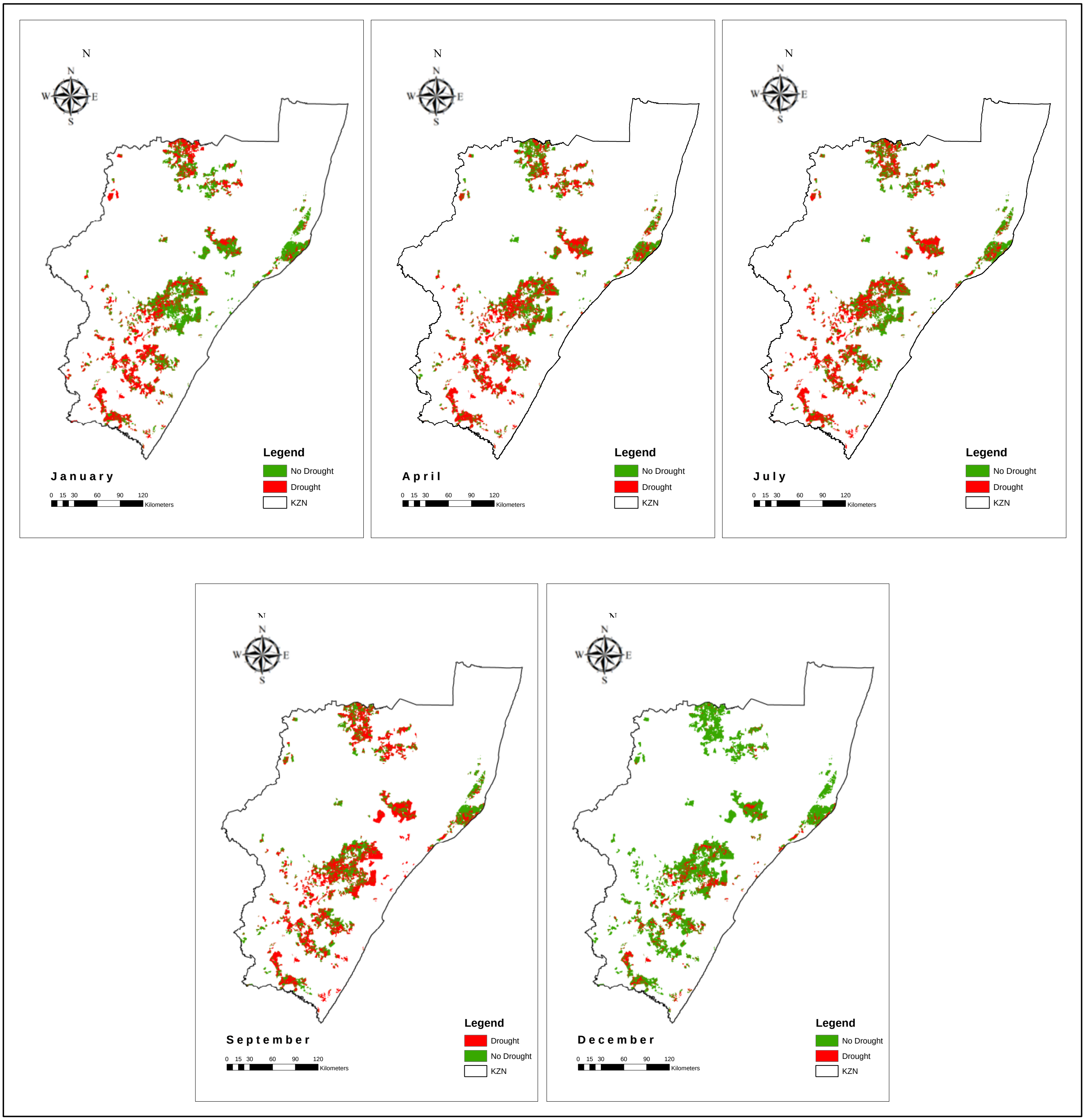


Figure 3.5 Maps showing RTF based on conditional drought indices cluster maps for January, April, July, September and December.

Figure 3.6 present similar maps seen in Figure 3.5; however, the drought damage clusters in Figure 3.6 are based on XGBoost trained using conditional drought indices. Contrary to the maps in Figure 3.5, the maps in Figure 3.6 indicate that from January through July 2015, more forested areas were impacted by the drought. This gave a notion that the drought had already commenced before the start of 2015. Drawing from July to September, there are no observable differences in drought damage clusters; however, December presents a considerable improvement as more areas show no drought damage.

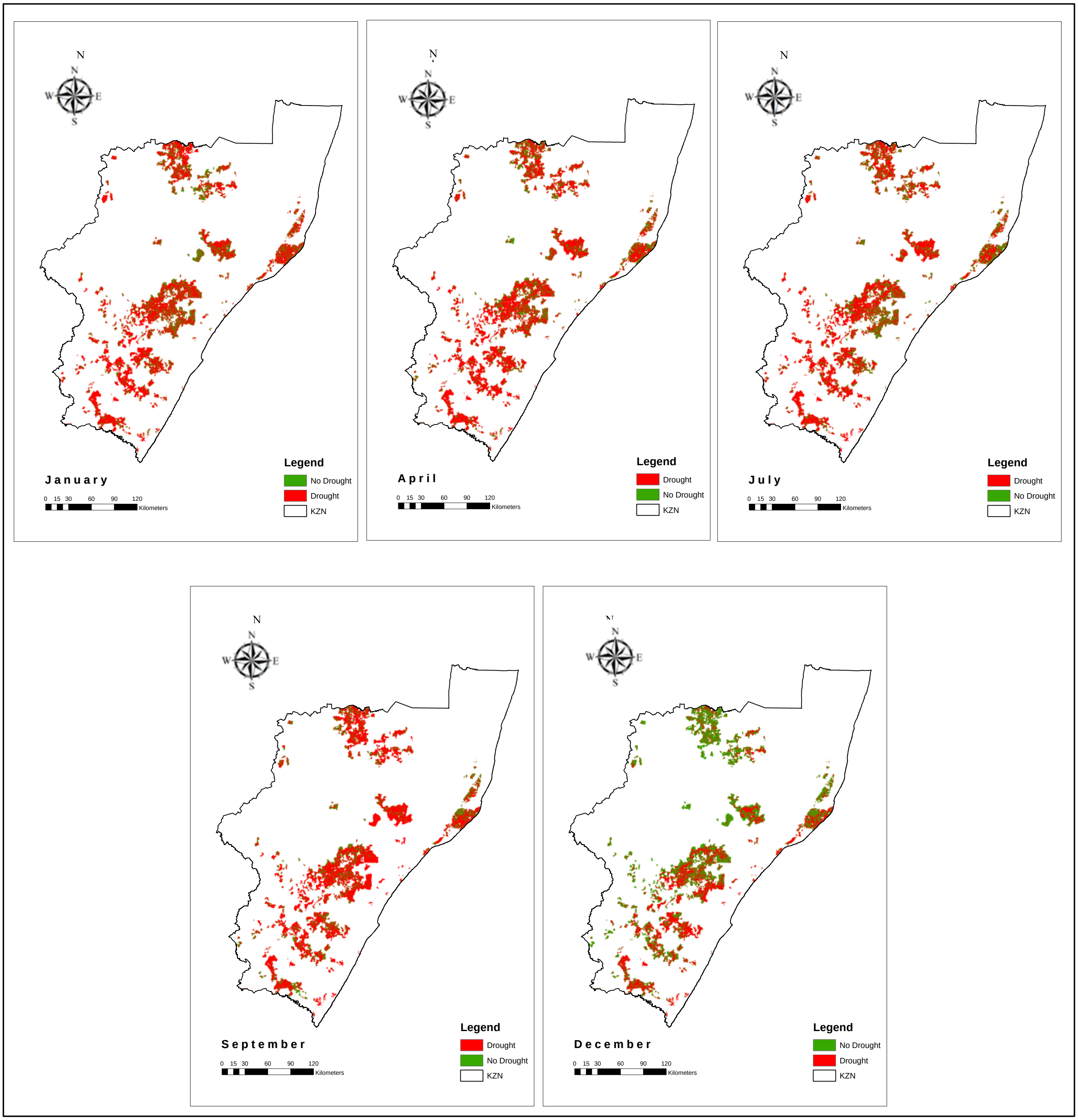


Figure 3.6 XGBoost based on conditional drought indices cluster maps for January, April, July, September and December.

Both algorithm classification based on conditional drought indices indicate that commercial forests of KZN started experiencing drought impacts before the start of 2015. However, there is a considerable improvement by the forest vegetation towards the end of 2015. It can also be observed that XGBoost predicted more drought damage than RTF.

Figure 3.7 presents the drought damage clusters in commercial forested regions within KZN for every three months starting from January 2015 to December of the same year constructed using RTF based on vegetation indices in R-Studio version 1.2.1335. It can be observed that in January there are some patches impacted by drought, these areas increase with time as it can be observed in July where more drought damaged areas were observed than initially in January. However, the trend observed from January to July does not continue to September where RTF predicted a considerable forest vegetation recovery which continues through to December.

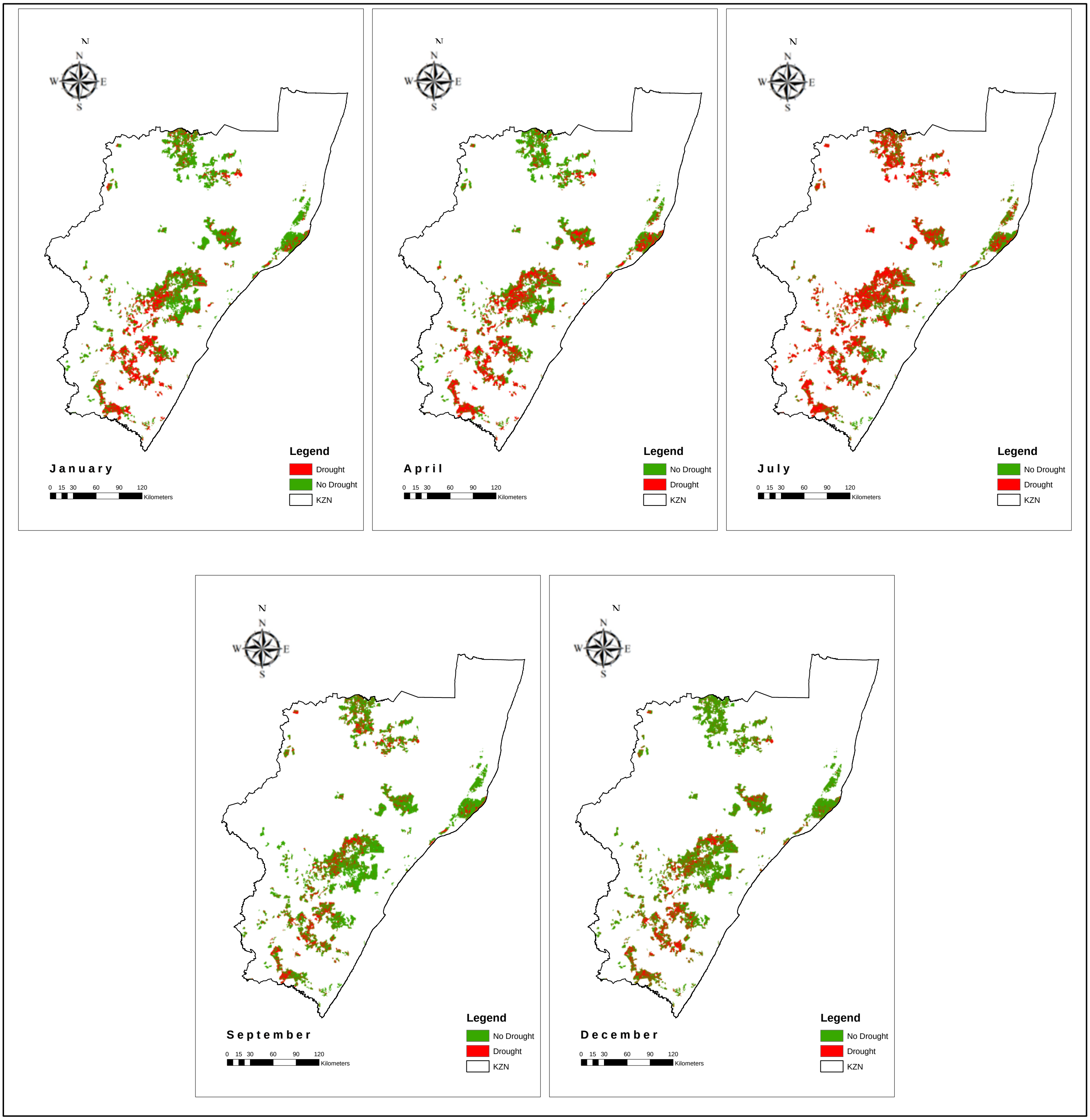


Figure 3.7 Maps showing RTF based on vegetation indices cluster maps for January, April, July, September and December.

The maps for clusters created from RTF trained using conditional drought indices predicted more forested areas to be impacted by droughts in 2015 compared to when the algorithm was trained using vegetation indices.

Figure 3.8 presents the drought damage clusters for commercial forested regions within KZN constructed using XGBoost based on vegetation indices. There are small increments of areas impacted by drought from January to April and to July. The opposite of what was observed from January to July compared to September to December where an incremental recovery of the forested region was observed as areas not impacted by drought increased.

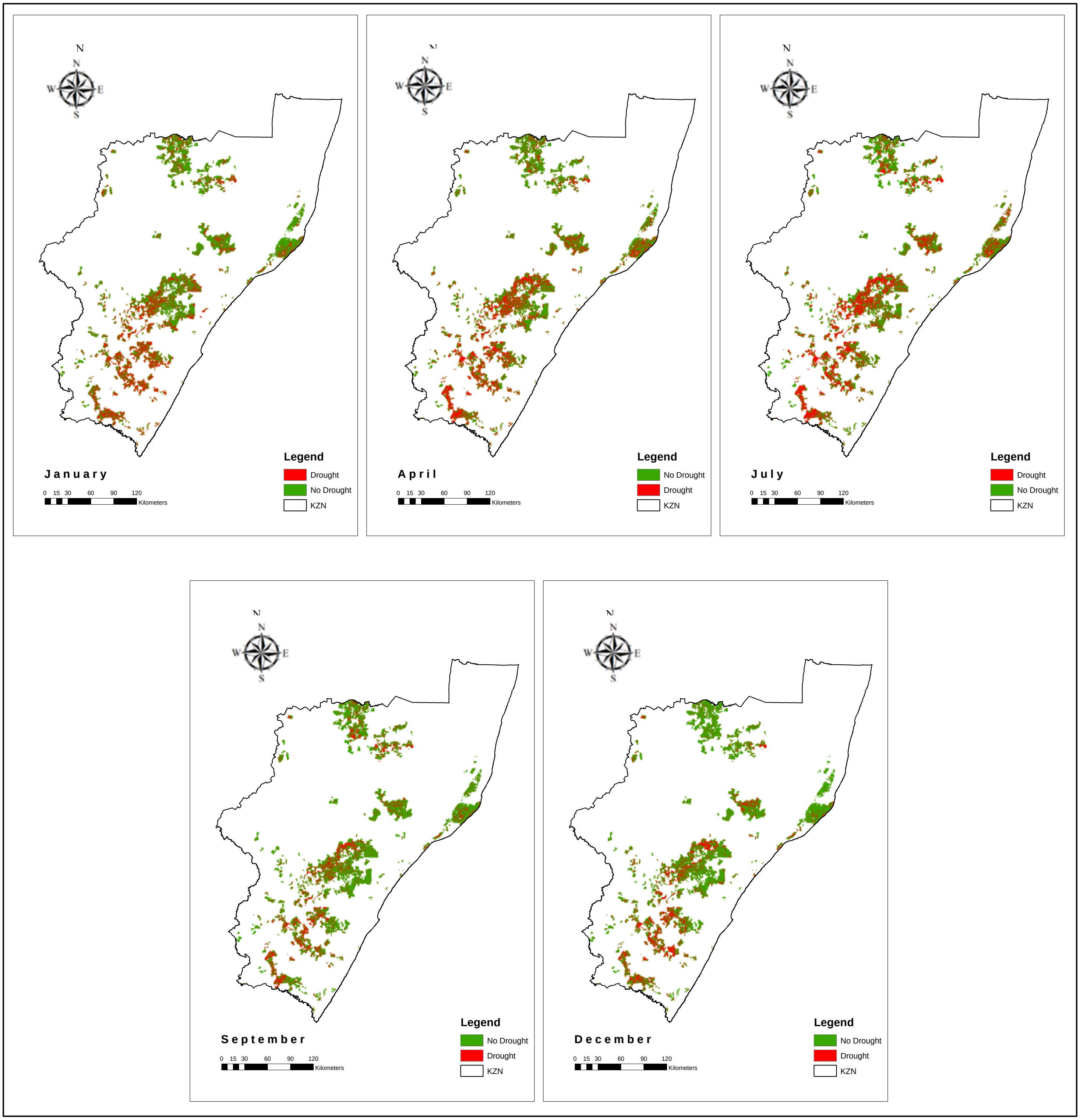


Figure 3.8 Maps showing XGBoost based on vegetation indices cluster maps for January, April, July, September and December.

The overall drought damage maps indicated that before the start of 2015 commercial forest regions in KZN were already being impacted by droughts and as the year continued more forested areas were impacted. However, towards the end of 2015 forested areas were recovering as indicated by the December drought damage clusters. To determine the consistency of the algorithms, drought damage clusters were constructed in R-Studio version 1.2.1335 for January of 2016 using RTF in Figure 3.9 and XGBoost in Figure 3.10, which were trained using 2015 indices information.

The maps for January 2016 showed a considerable difference from December cluster maps in that there was an increase in drought damage than previously observed for the last year. This could be due to the rains that occurred from February to March in 2015, therefore, resulting in the recovery observed in December of 2015 (Monyela, 2017). However, that rainfall period was followed by dry months which could explain the damage observed in January of 2016.

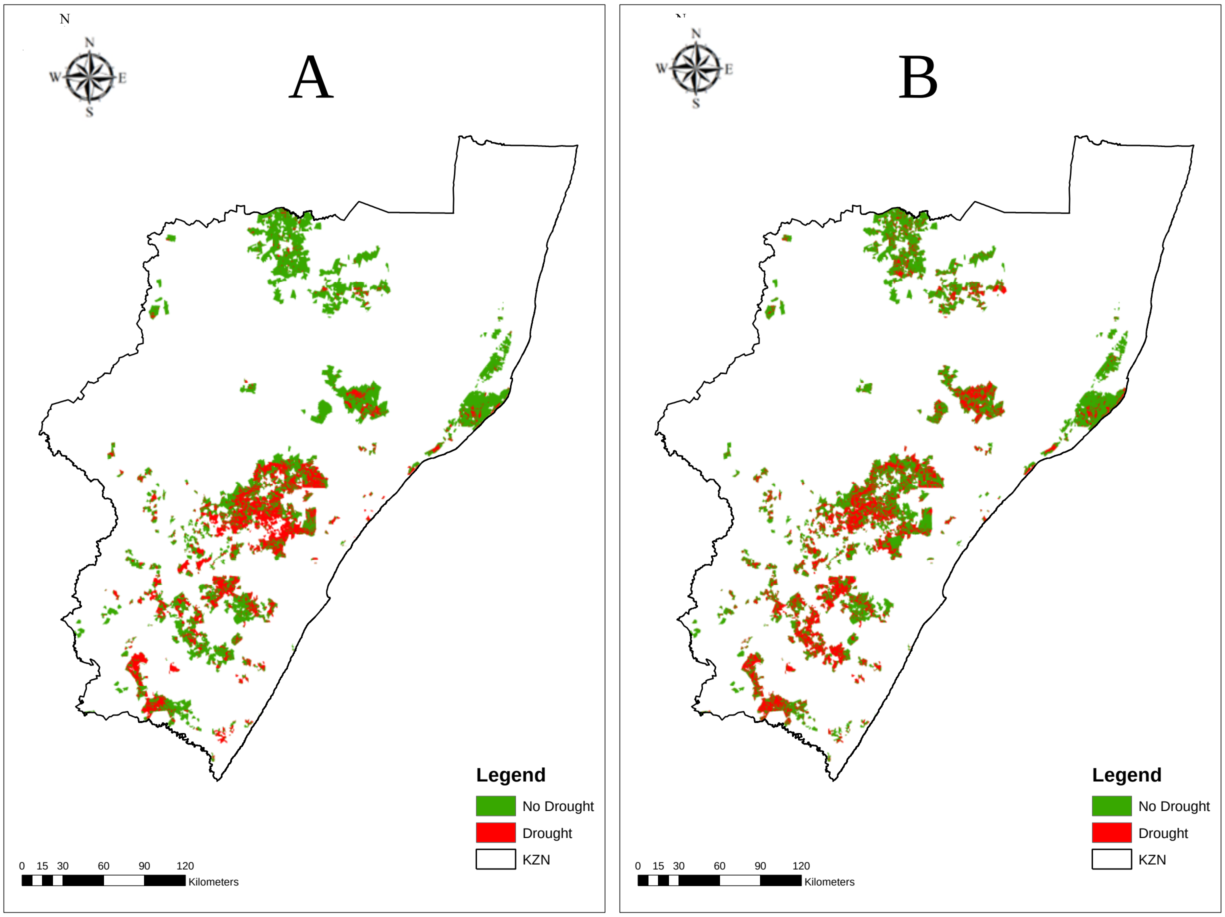


Figure 3.9 Maps showing January 2016 drought damage clusters created using RTF classifications based on conditional drought indices (A) and vegetation indices (B).

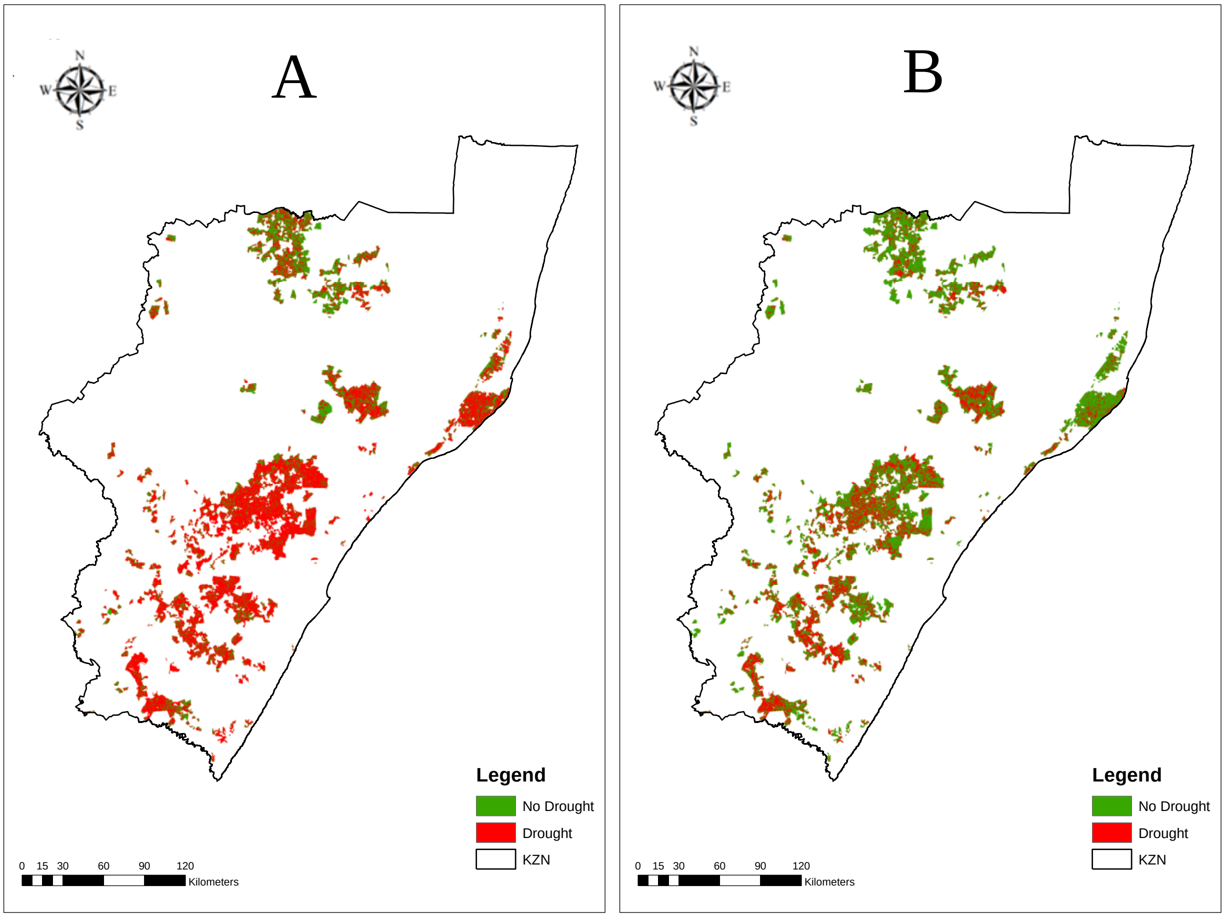


Figure 3.10 Maps showing January 2016 drought damage clusters created using XGBoost classifications based on conditional drought indices (A) and vegetation indices (B).

The overall results give a clear indication that XGBoost and RTF are capable of performing drought damage classifications; however, their performance is dependent on the type of indices used to train them. It can also be concluded that the algorithms have the potential of forecasting drought damage given their consistency drawn from Figure 3.9 and 3.10.

3.4 Discussion

This study was aimed at exploring the potential of RTF and XGBoost to perform drought damage classification for commercial forests in KZN. Even though far from extensive, the results from this study have demonstrated that the performance of RTF and XGBoost depends mainly on the type of index used to train them. This was drawn from the poor performance observed when the algorithms were trained using vegetation indices compared to a somewhat better performance when based on conditional drought indices. Most studies have argued against the use of vegetation indices especially NDVI to perform classifications for densely vegetated areas by demonstrating that these indices saturate in such environments (Gao *et al.*, 2000; Mutanga and Skidmore, 2004; Qiu *et al.*, 2018).

The limitations of NDVI led to its comparison with EVI which is assumed to be a superior index for vegetation monitoring due to its enhanced sensitivity to reduced atmospheric influence and high biomass (Huete *et al.*, 2002; Qiu *et al.*, 2018). Multiple studies have utilised or compared the accuracies of both NDVI and EVI for extracting vegetation information, namely, Lijun *et al.* (2008), Wardlow and Egbert (2010), Testa *et al.* (2018) and Bajocco *et al.* (2019). All the aforementioned studies concluded that EVI is a more convenient index compared to NDVI except for Wardlow and Egbert (2010). They found that both indices are susceptible to saturation during the growth stage of vegetation, therefore concluding that differences in classification based on these indices are negligible. However, in this study the differences were significant for the influence of EVI and NDVI in the accuracy of the classification.

XGBoost with 69.29% accuracy was outperformed by RTF with 70.67% accuracy when both algorithms were classifying based on vegetation indices (Table 3.7). The reason behind this outcome was observed to be the variable prioritised by the algorithm during the classification where XGBoost prioritised NDVI over EVI and NDWI. In contrast RTF prioritised EVI information over the other indices (Table 3.6). This, therefore, implied that XGBoost is hindered by the saturation of the indices and that EVI should be utilised over NDVI when performing forest vegetation classifications. Due to limited literature on the utilisation of XGBoost in drought analysis, these results could not be compared to other studies.

However, even when the EVI was prioritised, the algorithms still underperformed compared to when they classified based on information extracted using conditional drought indices. Vegetation indices have been in most cases utilised to temporary monitor vegetation. However, they cannot perform relative comparisons at pixel location or time period and conditional indices possess this ability (Kogan, 1995a; Peters *et al.*, 2002; Jiao *et al.*, 2016; Mohammad *et al.*, 2018). Zhang *et al.* (2017a) emphasised the fact that conditional drought indices are based on time-series analysis (calculated per pixel over time) makes them more accurate when extracting vegetation information. Kogan (1990) explained that time-series analysis considers the minimum which is determined by the available moisture and other values, including the historical maximum are determined by the weather. Therefore, it is these aspects that make conditional indices more accurate than vegetation indices.

XGBoost outperformed RTF when using conditional drought indices (Table 3.7) with 81.95% and 76.05% classification accuracies respectively. This could have been due to both algorithms prioritising different variables where XGBoost prioritised VHI and RTF prioritised VCI (Table 3.6). The underperformance of VCI in vegetation condition classification has been highlighted in multiple studies including Zhang *et al.* (2017a) and Jiao *et al.* (2016). They argued that VCI is calculated using NDVI which only considers the R and NIR bands. They also argued that NDVI is affected by the difference in spectral bandwidth and atmospheric changes especially in humid regions. Most importantly, Jiao *et al.* (2016) argued that VCI often fails to monitor the leaf status during drought conditions.

On the contrary, VHI is considered to be a superior index to other conditional drought indices because it is computed using both VCI and TCI (Kogan, 1995a; Mohammad *et al.*, 2018). This superiority implies better accuracy, given that it is drawn from the idea that it combines both the vegetation aspect from VCI and the temperature aspect from TCI. Multiple studies have performed drought analysis using VHI and compared its performance to other indices such as VCI, VMCI and TCI and determined that it performs better than those indices, such studies include Bhuiyan *et* *al.* (2006) and Amalo and Hidayat (2017). This, therefore, verifies the good performance obtained by XGBoost.

The algorithms classification results were then utilised to create drought damage clusters which are presented in Figure 3.5 to 3.8. Given the report by Agri SA (2016), which stated the 2015-2016 drought had a very high magnitude and as a result 2015 had the lowest mean annual precipitation since 1904. This led to the expectation that maps produced from the drought clusters should show more drought damage. Although different, maps from both algorithms did show more drought damage throughout the year.

The maps showed that during the beginning of 2015 the forested areas of KZN were already showing drought damage, which meant that the onset of the 2015-2016 drought preluded 2015. The resulting maps also showed an increase in drought damage with time. However, December showed a remarkable forest vegetation recovery which could be attributed to summer rains which are a major influence in the establishment of commercial forests in KZN. When comparing maps produced by RTF and XGBoost without considering the indices used to train them, it was observed that XGBoost classifications showed more drought damage than maps produce by RTF classifications. Given the expectation based on the Agri SA (2016), it can, therefore be argued that RTF was understating the true drought damage while maps based on XGBoost are true. This also applies to indices where classifications based on vegetation indices yielded less drought damage compared to more damage when classifying based on conditional drought indices.

Overall, the results presented in this study clearly show the potential of both XGBoost and RTF to classify drought damage in forest vegetation. However, very few studies have utilised these algorithms in drought analysis and in forested environments, which makes it nearly impossible to draw comparisons for the verification of the obtained results. Nonetheless, the obtained accuracy range of 69% to 82% could be improved upon especially for XGBoost. Xia *et al.* (2017) suggested that by carefully tuning the parameters of XGBoost its accuracy could be improved. This study has also demonstrated the superiority of conditional drought indices over vegetation indices which was also the main finding in such studies as Zhang *et al.* (2017a) and Mohammad *et al.* (2018). This therefore signifies the need to focus drought research towards conditional drought indices.

Moreover, it is suggested that future research should focus on all types of vegetation to determine where these algorithms provide optimal classification accuracies.

3.5 Conclusion

This study aimed to utilise RTF and XGBoost to classify drought damage in commercial forests in KZN using vegetation and conditional drought indices and in meeting this aim it was identified that vegetation indices, especially, NDVI hinder the performance of these indices. It was therefore concluded that MLAs perform better when using information derived from conditional drought indices. Additionally, the results demonstrated that:

1. The ability of MLAs to perform drought damage classifications.
2. XGBoost was more superior in detecting and mapping drought when compared to RFT.
3. Conditional drought indices performed better than vegetation indices in detecting drought in commercial forest plantations.
4. The onset of the 2015-2016 drought was before 2015.
5. Variable prioritisation has a significant influence on the classification accuracy of the algorithm.
6. MODIS imagery can be used by MLAs for classifications and the fact that it is freely available in the USGS archives provides an opportunity for more research to be conducted for the utilisation and improvement of these algorithms.

Although the results obtained in this study are promising, more research is still required to improve these results and to determine how good MLAs can be for drought analysis. Also, research into MLAs should not be limited to forest vegetation and drought damage, it should also be aimed at multiple environmental facets including pressing issues such as climate change.

4. CHAPTER FOUR: SUMMARY OF STUDY FINDINGS

This study aimed to analyse the impacts of droughts on commercial forests in KZN while exploring the effectiveness of MLAs at performing such analysis. These algorithms were employed at both catchment and provincial scales, using both Landsat and MODIS data. The obtained results therefore conclusively demonstrated the capabilities of MLAs and highlighted advantages they bring when analysing drought impacts. This study is one of the very few studies to utilise XGBoost and RTF for the classification of drought damage on commercial forests. The study also highlighted the advantages of conditional drought indices over conventional vegetation indices when analysing drought damage on commercial forests. Therefore, in this chapter objectives set in chapter one will be reviewed against the findings.

One of the objectives in this study was to investigate and improve the utilisation of conditional and vegetation indices in forest drought analysis. The key finding relating to this objective in both chapter two and chapter three was the high accuracy demonstrated by the three MLAs when performing classifications based on conditional drought indices compared to when using information derived from vegetation indices. These findings highlighted the limitations of vegetation indices, especially, NDVI and recommended the use of EVI instead. However, even when EVI was prioritised by the algorithms they still underperformed compared to when using conditional indices to perform classifications.

The high accuracy of conditional drought indices over vegetation indices observed in the study is due to conditional drought indices being calculated per pixel over time makes which makes it more accurate when extracting vegetation information. This, therefore, indicates that for ML algorithms to perform at their best they require more accurate information, which for vegetation drought analysis is best obtained from conditional drought indices.

Most drought studies are not focused on drought analysis at catchment level which creates a gap in drought research. Therefore, this study set an objective to close that gap by utilising Landsat imagery to classify drought damage at the Sappi Shafton catchment. This was detailed in chapter two and the results demonstrated that catchment drought analysis is a viable method of determining drought damage on commercial forests. This was drawn from the high accuracy of Kernel – SVM and RTF when classifying drought damage on commercial forests.

When comparing these results (catchment scale and Landsat imagery) with those in chapter three (provincial scale and MODIS imagery), the results from the latter produced more accurate classifications. This outcome is more important because in chapter three XGBoost which is said to be more accurate than RTF and Kernel – SVM was used and OA results demonstrated that it underperformed compared to RTF and Kernel – SVM results in chapter two. To demonstrate this conclusion, Figure 4.1 presents the drought damage classification using RTF based on both MODIS and Landsat information at the catchment scale. It can be observed that there are more details on the Landsat map compared to the MODIS map, which is due to higher resolution of the Landsat imagery. It can also be observed that MODIS predicted more forested areas to be impacted by droughts.

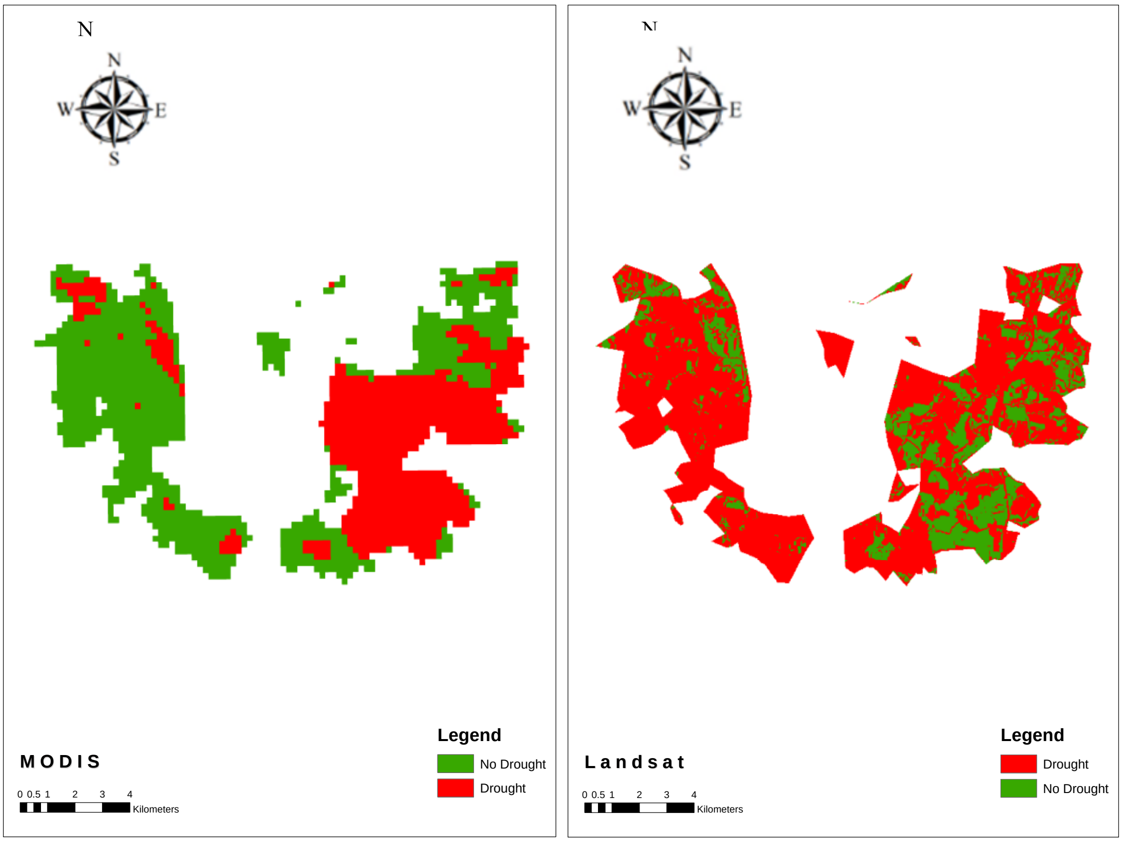


Figure 4.1 Drought damage on Shafton commercial forest mapped using RTF based on MODIS and Landsat information.

This meant high spatial resolution drought analysis at the catchment scale using Landsat imagery should be preferred over low spatial resolution drought analysis.

The results from the study also demonstrated the high potential of MLAs in improving drought analysis for forested regions which was also one of the objectives. As aforementioned, these algorithms showed great potential, especially when classifying drought damage at a catchment scale using conditional indices information over vegetation indices. Therefore, calling for more incorporation of MLAs in drought analysis and other remote sensing applications. Another reason for the use of MLAs is that their accuracy can be improved over time by gaining experience through being employed in multiple datasets (Jean *et al.*, 2016; Luo, 2016). This is also beneficial given the prospect of increased data availability as exemplified by the free availability of Landsat data through USGS.

However, it should be conceded that MLAs are highly parameter dependent and that might have limited the findings in this study. That is because for this study RTF and Kernel – SVM parameters were selected using automatic parameterisation to avoid the labour-intensive manual parameterisation. It is argued that there is still room for improving automatic parameterisation in MLAs which will result in improved efficiency and accuracy of these algorithms. With more improvements required in the algorithm's automatic parameterisation, therefore creates room for error in the results obtained in this study.

5. CONCLUSION

Droughts have a devastating impact on developing countries given their dependence on drought susceptible agricultural resources and commercial forestry. Therefore, analysing droughts is necessary for understanding this phenomenon and the development of ways to mitigate its impacts. As aforementioned, droughts are complicated and require a holistic consideration of all its facets which include extent, severity and duration. This study, therefore, explored the potential of MLAs to perform this function with the exception of the severity aspect. The result from the study demonstrated that MLAs are capable of classifying and visualising the extent of drought damage on commercial forests at the catchment scale and provincial scale.

This conclusion is based on the findings from this thesis and answers the research question established in the first chapter of the thesis:

Can Landsat be used to analyse the impacts of droughts on commercial forests at a catchment scale?

- RTF and Kernel - SVM produced high accuracies when classifying drought damage on commercial forests within the Sappi Shafton plantation using Landsat images. However, this accuracy varies according to the type of indices used to perform the classification where conditional drought indices produced higher accuracies compared to vegetation indices.

How does the drought damage classification conducted at catchment scale compare with those conducted at provincial scale?

- Drought damage classification conducted at catchment scale produced higher overall accuracies and more detailed maps in terms of delineating areas with trees impacted by droughts compared to provincial drought damage classification.

Can MLAs improve on results achieved using conventional classification methods?

- MLAs have an inherent ability to learn and improve over time and provide a robust solution to drought analysis and this study provides an uncomplicated way of utilising them. However, there was a significant limitation encountered which is that the accuracy of MLAs decreased when computing drought damage based on vegetation indices. The mentioned decrease is not inherent in the algorithms, instead it is associated with the limitations of the vegetation indices themselves. Therefore, it was concluded that conditional drought indices information should be preferred for drought classification over vegetation indices and MLAs should be preferred over conventional classification methods.

Overall, this study provided insight into three of MLA, namely, RTF, Kernel – SVM and XGBoost. This is mainly for XGBoost, which is rarely used in drought investigations. Researchers should utilise this study as a base for future drought analysis using MLAs.

6. FUTURE RESEARCH RECOMMENDATIONS

Future research should focus on improving methods of parameter selection for MLAs to produce more efficient and accurate classifications. This study utilised three MLAs and given the large pool of supervised and unsupervised algorithms to select from, and future studies should utilise more of these algorithms to determine their efficiency compared to the ones used in this study. Also, MLAs drought damage classifications can be compared to other proposed classification such as the linear combined index (LCI) and linear spectral unmixing.

This study focused on forested study areas; future research can, therefore shift, the focus towards different land covers such as grasslands or areas with sparse vegetation where there could be less saturation of vegetation indices. It was also noted in the study that different tree species respond differently to drought therefore, these differences should be investigated thoroughly. Future research can also focus mainly on high-resolution remote sensing imagery to perform drought analysis at the catchment scale to obtain more details as demonstrated in this study. Droughts require a holistic consideration, therefore developing a method in addition to ML algorithms that can investigate the extent, severity and duration of a drought simultaneously is essential.

7. REFERENCES

Afanador, NL, Smolinska, A, Tran, TN and Blanchet, L. 2016. Unsupervised random forest: a tutorial with case studies. *journal of Chemometrics* 30 (5): 232-241.

Agri SA. 2016. A Raindrop in the Drought. *Report to the Multi-stakeholder Task Team on the Drought*

Ahmadi, S, Azarnivand, H, Khosravi, H, Dehghana, P and Behrang Manesh, M. 2019. Assessment the effect of drought and land use change on vegetation using Landsat data. *Desert* 24 (1): 23-31.

Akar, Ö. 2018. The Rotation Forest algorithm and object-based classification method for land use mapping through UAV images. *Geocarto international* 33 (5): 538-553.

Albaugh, JM, Dye, PJ and King, JS. 2013. Eucalyptus and water use in South Africa. *International Journal of Forestry Research* 2013

Allen, CD, Macalady, AK, Chenchouni, H, Bachelet, D, McDowell, N, Vennetier, M, Kitzberger, T, Rigling, A, Breshears, DD and Hogg, ET. 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest ecology and management* 259 (4): 660-684.

Amalo, LF and Hidayat, R.2017. Comparison between remote-sensing-based drought indices in east java. *IOP Conference Series: Earth and Environmental Science*, 012009. IOP Publishing,

Arganda-Carreras, I, Kaynig, V, Rueden, C, Eliceiri, KW, Schindelin, J, Cardona, A and Sebastian Seung, H. 2017. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* 33 (15): 2424-2426.

Asner, GP. 2001. Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing* 22 (18): 3855-3862.

Asner, GP and Alencar, A. 2010. Drought impacts on the Amazon forest: the remote sensing perspective. *New Phytologist* 187 (3): 569-578.

Bahta, YT, Jordaan, A and Muyambo, F. 2016. Communal farmers' perception of drought in South Africa: Policy implication for drought risk reduction. *International Journal of Disaster Risk Reduction* 20 39-50.

Bajocco, S, Ferrara, C, Alivernini, A, Bascietto, M and Ricotta, C. 2019. Remotely-sensed phenology of Italian forests: Going beyond the species. *International Journal of Applied Earth Observation and Geoinformation* 74 314-321.

Baniya, B, Tang, Q, Xu, X, Haile, GG and Chhipi-Shrestha, G. 2019. Spatial and Temporal Variation of Drought Based on Satellite Derived Vegetation Condition Index in Nepal from 1982–2015. *Sensors* 19 (2): 430.

Baudoin, M-A, Vogel, C, Nortje, K and Naik, M. 2017. Living with drought in South Africa: lessons learnt from the recent El Niño drought period. *International Journal of Disaster Risk Reduction* 23 128-137.

Bayissa, YA, Tadesse, T, Svoboda, M, Wardlow, B, Poulsen, C, Swigart, J and Van Andel, SJ. 2019. Developing a satellite-based combined drought indicator to monitor agricultural drought: a case study for Ethiopia. *GIScience & Remote Sensing* 56 (5): 718-748.

Belgiu, M and Drăguţ, L. 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96 67-75.

Bello, OM and Aina, YA. 2014. Satellite Remote Sensing as a Tool in Disaster Management and Sustainable Development: Towards a Synergistic Approach. *Procedia - Social and Behavioral Sciences* 120 365-373.

Bergstra, JS, Bardenet, R, Bengio, Y and Kégl, B.2011. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 2546-2554.

Bhuiyan, C, Singh, RP and Kogan, FN. 2006. Monitoring drought dynamics in the Aravalli region (India) using different indices based on ground and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation* 8 (4): 289-302.

Blaser, R and Fryzlewicz, P. 2016. Random rotation ensembles. *The Journal of Machine Learning Research* 17 (1): 126-151.

Botai, C, Botai, J, Dlamini, L, Zwane, N and Phaduli, E. 2016. Characteristics of droughts in South Africa: A case study of free state and north west provinces. *Water* 8 (10): 439.

Botai, JO, Botai, CM, De Wit, JP, Muthoni, M and Adeola, AM. 2019. Analysis of Drought Progression Physiognomies in South Africa. *Water* 11 (2): 299.

Breshears, DD, Cobb, NS, Rich, PM, Price, KP, Allen, CD, Balice, RG, Romme, WH, Kastens, JH, Floyd, ML, Belnap, J, Anderson, JJ, Myers, OB and Meyer, CW. 2005. Regional vegetation die-off in response to global-change-type drought. *Proceedings of the National Academy of Sciences of the United States of America* 102 (42): 15144-15148.

Brunton, SL, Noack, BR and Koumoutsakos, PJARoFM. 2020. Machine learning for fluid mechanics. 52 477-508.

Bulcock, H and Jewitt, G. 2012. Field data collection and analysis of canopy and litter interception in commercial forest plantations in the KwaZulu-Natal Midlands, South Africa. *Hydrology and Earth System Sciences* 16 (10): 3717-3728.

Caccamo, G, Chisholm, L, Bradstock, RA and Puotinen, M. 2011. Assessing the sensitivity of MODIS to monitor drought in high biomass ecosystems. *Remote Sensing of Environment* 115 (10): 2626-2639.

Chandrasekar, P, Qian, K, Shahriar, H and Bhattacharya, P.2017. Improving the Prediction Accuracy of Decision Tree Mining with Data Preprocessing. *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, 481-484.

Chen, P-Y, Fedosejevs, G, Tiscareño-LóPez, M and Arnold, JG. 2006. Assessment of MODIS-EVI, MODIS-NDVI and VEGETATION-NDVI Composite Data Using Agricultural Measurements: An Example at Corn Fields in Western Mexico. *Environmental Monitoring and Assessment* 119 (1): 69-82.

Chen, T and Guestrin, C.2016. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794. ACM,

Clulow, A, Everson, C and Gush, M. 2011. The long-term impact of Acacia mearnsii trees on evaporation, streamflow and ground water resources. *Water Research Commission Report No. TT505/11, WRC, Pretoria, South Africa*

Colkesen, I and Kavzoglu, T. 2017. Ensemble-based canonical correlation forest (CCF) for land use and land cover classification using sentinel-2 and Landsat OLI imagery. *Remote Sensing Letters* 8 (11): 1082-1091.

Cortes, C and Mohri, M.2005. Confidence intervals for the area under the ROC curve. *Advances in neural information processing systems*, 305-312.

da Costa, ACL, Galbraith, D, Almeida, S, Portela, BTT, da Costa, M, de Athaydes Silva Junior, J, Braga, AP, de Gonçalves, PHL, de Oliveira, AA, Fisher, R, Phillips, OL, Metcalfe, DB, Levy, P and Meir, P. 2010. Effect of 7 yr of experimental drought on vegetation

dynamics and biomass storage of an eastern Amazonian rainforest. *New Phytologist* 187 (3): 579-591.
Dale, VH, Joyce, LA, McNulty, S, Neilson, RP, Ayres, MP, Flannigan, MD, Hanson, PJ, Irland, LC, Lugo, AE and Peterson, CJ. 2001. Climate change and forest disturbances: climate change can affect forests by altering the frequency, intensity, duration, and timing of fire, drought, introduced species, insect and pathogen outbreaks, hurricanes, windstorms, ice storms, or landslides. *BioScience* 51 (9): 723-734.
De Jager, E. 2016. SA rainfall in 2015 the lowest on record - SAWS. [Internet]. Politics Web. Available from: https://www.politicsweb.co.za/documents/sa-rainfall-in-2015-the-lowest-on-record--saws. [Accessed: 16 August 2019].
De Sy, V, Herold, M, Achard, F, Asner, GP, Held, A, Kellndorfer, J and Verbesselt, J. 2012. Synergies of multiple remote sensing data sources for REDD+ monitoring. *Current Opinion in Environmental Sustainability* 4 (6): 696-706.
Du, L, Tian, Q, Yu, T, Meng, Q, Jancso, T, Udvardy, P and Huang, Y. 2013. A comprehensive drought monitoring method integrating MODIS and TRMM data. *International Journal of Applied Earth Observation and Geoinformation* 23 245-253.
Du, P, Samat, A, Waske, B, Liu, S and Li, Z. 2015. Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing* 105 38-53.
Dube, T and Mutanga, O. 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing* 101 36-46.
Dube, T, Mutanga, O and Riyad, I. 2016. Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs. *Tropical Ecology* 57 (3): 393-405.
Duncan, GE. 2019. Modelling the impact of land cover and forestry change on the hydrological characteristics of the Two Streams sub-catchment, Natal Midlands. Unpublished thesis,
Dutrieux, LP, Verbesselt, J, Kooistra, L and Herold, M. 2015. Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia. *ISPRS Journal of Photogrammetry and Remote Sensing* 107 112-125.
Dutta, D, Kundu, A, Patel, NR, Saha, SK and Siddiqui, AR. 2015. Assessment of agricultural drought in Rajasthan (India) using remote sensing derived Vegetation Condition Index (VCI) and Standardized Precipitation Index (SPI). *The Egyptian Journal of Remote Sensing and Space Science* 18 (1): 53-63.
Edossa, DC, Woyessa, YE and Welderufael, WA. 2014. Analysis of droughts in the central region of South Africa and their association with SST anomalies. *International Journal of Atmospheric Sciences* 2014
Fan, J, Wang, X, Wu, L, Zhou, H, Zhang, F, Yu, X, Lu, X, Xiang, YJEc and management. 2018. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. 164 102-111.
Fassnacht, FE, Neumann, C, Förster, M, Buddenbaum, H, Ghosh, A, Clasen, A, Joshi, PK and Koch, B. 2014. Comparison of Feature Reduction Algorithms for Classifying Tree Species With Hyperspectral Data on Three Central European Test Sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (6): 2547-2561.
Fils, SCN, Mimba, ME, Dzana, JG, Etouna, J, Mounoumeck, PV and Hakdaoui, M. 2018. TM/ETM+/LDCM Images for Studying Land Surface Temperature (LST) Interplay

with Impervious Surfaces Changes over Time Within the Douala Metropolis, Cameroon. *Journal of the Indian Society of Remote Sensing* 46 (1): 131-143.

Führer, E. 2000. Forest functions, ecosystem stability and management. *Forest Ecology and management* 132 (1): 29-38.

Ganey, JL and Vojta, SC. 2011. Tree mortality in drought-stressed mixed-conifer and ponderosa pine forests, Arizona, USA. *Forest Ecology and Management* 261 (1): 162-168.

Gao, B-C. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment* 58 (3): 257-266.

Gao, X, Huete, AR, Ni, W and Miura, T. 2000. Optical–Biophysical Relationships of Vegetation Spectra without Background Contamination. *Remote Sensing of Environment* 74 (3): 609-620.

Gao, Y, Quevedo, A, Szantoi, Z and Skutsch, M. 2019. Monitoring forest disturbance using time-series MODIS NDVI in Michoacán, Mexico. *Geocarto International* 1-17.

Gazol, A, Camarero, JJ, Anderegg, WRL and Vicente-Serrano, SM. 2017. Impacts of droughts on the growth resilience of Northern Hemisphere forests. *Global Ecology and Biogeography* 26 (2): 166-176.

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

Ghimire, B, Rogan, J, Galiano, VR, Panday, P and Neeti, N. 2012. An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. *GIScience & Remote Sensing* 49 (5): 623-643.

Gregorutti, B, Michel, B and Saint-Pierre, P. 2017. Correlation and variable importance in random forests. *Statistics and Computing* 27 (3): 659-678.

Gu, Y, Hunt, E, Wardlow, B, Basara, JB, Brown, JF and Verdin, JP. 2008. Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. *Geophysical Research Letters* 35 (22):

Gulácsi, A and Kovács, F. 2018. Drought monitoring of forest vegetation using MODIS-based normalized difference drought index in Hungary. *Hungarian Geographical Bulletin* 67 (1):

Han, N, Du, H, Zhou, G, Xu, X, Ge, H, Liu, L, Gao, G and Sun, S. 2015. Exploring the synergistic use of multi-scale image object metrics for land-use/land-cover mapping using an object-based approach. *International Journal of Remote Sensing* 36 (13): 3544-3562.

Han, P, Shang, S, Sun, A, Zhao, P, Zheng, K and Kalnis, P.2019. AUC-MF: Point of Interest Recommendation with AUC Maximization. *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 1558-1561.

Hans, WSA. 1948. The Present Climatic Fluctuation. *The Geographical Journal* 112 (4/6): 165-193.

Hao, Z and AghaKouchak, A. 2013. Multivariate standardized drought index: a parametric multi-index model. *Advances in Water Resources* 57 12-18.

Hao, Z, Hao, F, Singh, VP, Xia, Y, Ouyang, W and Shen, X. 2016. A theoretical drought classification method for the multivariate drought index based on distribution properties of standardized drought indices. *Advances in Water Resources* 92 240-247.

Hao, Z and Singh, VP. 2015. Drought characterization from a multivariate perspective: A review. *Journal of Hydrology* 527 668-678.

Hayes, M, Svoboda, M, Wall, N and Widhalm, M. 2011. The Lincoln declaration on drought indices: universal meteorological drought index recommended. *Bulletin of the American Meteorological Society* 92 (4): 485-488.
Heim Jr, RR. 2002. A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society* 83 (8): 1149-1166.
Hitayezu, P, Zegeye, EW and Ortmann, GF. 2014. Some Aspects of Agricultural Vulnerability to Climate Change in the KwaZulu-Natal Midlands, South Africa:A Systematic Review. *Journal of Human Ecology* 48 (3): 347-356.
Hlahla, S and Hill, TR. 2018. Responses to Climate Variability in Urban Poor Communities in Pietermaritzburg, KwaZulu-Natal, South Africa. *SAGE Open* 8 (3): 2158244018800914.
Hobbs, N. 2018. Kernel SVM for Image Classification.
Hope, A, Fouad, G and Granovskaya, Y. 2014. Evaluating drought response of Southern Cape Indigenous Forests, South Africa, using MODIS data. *International journal of remote sensing* 35 (13): 4852-4864.
Hosonuma, N, Herold, M, De Sy, V, De Fries, RS, Brockhaus, M, Verchot, L, Angelsen, A and Romijn, E. 2012. An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters* 7 (4): 044009.
Huete, A, Didan, K, Miura, T, Rodriguez, EP, Gao, X and Ferreira, LG. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment* 83 (1-2): 195-213.
Huete, A, Justice, C and Van Leeuwen, W. 1999. MODIS vegetation index (MOD13). *Algorithm theoretical basis document* 3 213.
Huete, AR. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25 (3): 295-309.
Jain, AK. 2010. Data clustering: 50 years beyond K-means. *Pattern recognition letters* 31 (8): 651-666.
Jean, N, Burke, M, Xie, M, Davis, WM, Lobell, DB and Ermon, S. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353 (6301): 790.
Jewitt, D. 2016. Mapping landscape beta diversity of plants across KwaZulu-Natal, South Africa, for aiding conservation planning. *Biodiversity and Conservation*
Jewitt, D, Ps, G, Bfn, E, Tg, OC and Witkowski, E. 2015. Systematic land-cover change in KwaZulu-Natal, South Africa: Implications for biodiversity. *South African Journal of Science* 111
Jiao, W, Zhang, L, Chang, Q, Fu, D, Cen, Y and Tong, Q. 2016. Evaluating an enhanced vegetation condition index (VCI) based on VIUPD for drought monitoring in the continental United States. *Remote Sensing* 8 (3): 224.
Jordan, MI and Mitchell, TM. 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349 (6245): 255-260.
Jury, MR. 1998. Statistical Analysis and Prediction of KwaZulu-Natal Climate. *Theoretical and Applied Climatology* 60 (1): 1-10.
Kavzoglu, T, Colkesen, I and Yomralioglu, T. 2015. Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution WorldView-2 image. *Remote Sensing Letters* 6 (11): 834-843.
Kazllarof, V, Karlos, S and Kotsiantis, S. 2019. Active learning Rotation Forest for multiclass classification. *Computational Intelligence*
Khamar, M and Eftekhari, M. 2018. Multi-Manifold based Rotation Forest for classification. *Applied Soft Computing* 68 626-635.

Kirilenko, AP and Sedjo, RA. 2007. Climate change impacts on forestry. *Proceedings of the National Academy of Sciences* 104 (50): 19697-19702.
Kogan, F. 1990. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of Remote Sensing* 11 (8): 1405-1419.
Kogan, F. 2002. World droughts in the new millennium from AVHRR-based vegetation health indices. *Eos, Transactions American Geophysical Union* 83 (48): 557-563.
Kogan, F and Sullivan, J. 1993. Development of global drought-watch system using NOAA/AVHRR data. *Advances in Space Research* 13 (5): 219-222.
Kogan, FN. 1995a. Application of vegetation index and brightness temperature for drought detection. *Advances in Space Research* 15 (11): 91-100.
Kogan, FN. 1995b. Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. *Bulletin of the American Meteorological Society* 76 (5): 655-668.
Kogan, FN. 1997. Global drought watch from space. *Bulletin of the American Meteorological Society* 78 (4): 621-636.
Kruger, A. 2007. Trends in cloud cover from 1960 to 2005 over South Africa. *Water SA* 33 (5):
Kuncheva, LI and Rodríguez, JJ.2007. An experimental study on rotation forest ensembles. *International workshop on multiple classifier systems*, 459-468. Springer,
Lambin, EF. 1999. Monitoring forest degradation in tropical regions by remote sensing: some methodological issues. *Global ecology and biogeography* 8 (3-4): 191-198.
Lijun, Z, Zengxiang, Z, Tingting, D and Xiao, W. 2008. Application of MODIS/NDVI and MODIS EVI to extracting the information of cultivated land and comparison analysis. *Transactions of the Chinese Society of Agricultural Engineering* 2008 (3):
Ling, CX, Huang, J and Zhang, H.2003. AUC: a statistically consistent and more discriminating measure than accuracy. *Ijcai*, 519-524.
Liu, HQ and Huete, A. 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE transactions on geoscience and remote sensing* 33 (2): 457-465.
Lottering, RT, Govender, M, Peerbhay, K and Lottering, S. 2020. Comparing partial least squares (PLS) discriminant analysis and sparse PLS discriminant analysis in detecting and mapping Solanum mauritianum in commercial forest plantations using image texture. *ISPRS Journal of Photogrammetry and Remote Sensing* 159 271-280.
Luo, G. 2016. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Network Modeling Analysis in Health Informatics and Bioinformatics* 5 (1): 18.
Ma, L, Li, M, Ma, X, Cheng, L, Du, P and Liu, Y. 2017. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 130 277-293.
Martin-Benito, D, Anchukaitis, KJ, Evans, MN, Del Río, M, Beeckman, H and Cañellas, I. 2017. Effects of Drought on Xylem Anatomy and Water-Use Efficiency of Two Co-Occurring Pine Species. *Forests* 8 (9): 332.
Martínez-Vilalta, J, Lloret, F and Breshears, DD. 2012. Drought-induced forest decline: causes, scope and implications. *Biology Letters* 8 (5): 689-691.
Masupha, TE and Moeletsi, ME. 2017. Use of standardized precipitation evapotranspiration index to investigate drought relative to maize, in the Luvuvhu River catchment area, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C* 102 1-9.

Masupha, TE and Moeletsi, ME. 2018. Analysis of potential future droughts limiting maize production, in the Luvuvhu River catchment area, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C* 105 44-51.
Matsushita, B, Yang, W, Chen, J, Onda, Y and Qiu, G. 2007. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. *Sensors* 7 (11): 2636-2651.
Maxwell, AE, Warner, TA and Fang, F. 2018. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing* 39 (9): 2784-2817.
Melendez-Pastor, I, Navarro-Pedreño, J, Gómez, I and Koch, M. 2010. Detecting drought induced environmental changes in a Mediterranean wetland by remote sensing. *Applied Geography* 30 (2): 254-262.
Melgani, F and Bruzzone, L. 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42 (8): 1778-1790.
Mera, GA. 2018. Drought and its impacts in Ethiopia. *Weather and climate extremes* 22 24-35.
Meshram, SG, Gautam, R and Kahya, E. 2018. Drought analysis in the Tons River Basin, India during 1969-2008. *Theoretical and Applied Climatology* 132 (3): 939-951.
Millar, CI and Stephenson, NL. 2015. Temperate forest health in an era of emerging megadisturbance. *Science* 349 (6250): 823.
Mladenova, IE, Jackson, TJ, Njoku, E, Bindlish, R, Chan, S, Cosh, MH, Holmes, TRH, de Jeu, RAM, Jones, L, Kimball, J, Paloscia, S and Santi, E. 2014. Remote monitoring of soil moisture using passive microwave-based techniques — Theoretical basis and overview of selected algorithms for AMSR-E. *Remote Sensing of Environment* 144 197-213.
Mohammad, K, Jang, M-W, Hwang, S and Jang, T. 2018. Evaluating the Agricultural Drought for Pre-Kharif Season in Bangladesh using MODIS Vegetation Health Index. *Journal of The Korean Society of Agricultural Engineers* 60 (6): 55-63.
Monyela, BM. 2017. A two-year long drought in summer 2014/2015 and 2015/2016 over South Africa. Unpublished thesis, University of Cape Town,
Mountrakis, G, Im, J and Ogole, C. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3): 247-259.
Mutanga, O and Skidmore, AK. 2004. Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrus ciliaris) measured under controlled laboratory conditions. *International journal of applied earth observation and geoinformation* 5 (2): 87-96.
Naidoo, S, Davis, C and Van Garderen, EA. 2013. Forests, rangelands and climate change in southern Africa. *Forests and climate change working paper* 12
Nanda, AM, Seminar , BK, Dodi, N and Akhiruddin, M. 2018. A comparison study of kernel functions in the support vector machine and its application for termite detection. *Information* 9 (1): 5.
Nepstad, DC, Tohver, IM, Ray, D, Moutinho, P and Cardinot, G. 2007. Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology* 88 (9): 2259-2269.
Nicholson, SE. 1986. Climate, drought, and famine in Africa.
Nouri, H, Beecham, S, Anderson, S and Nagler, P. 2014. High spatial resolution WorldView-2 imagery for mapping NDVI and its relationship to temporal urban landscape evapotranspiration factors. *Remote sensing* 6 (1): 580-602.

Novillo, CJ, Arrogante-Funes, P and Romero-Calcerrada, R. 2019. Recent NDVI Trends in Mainland Spain: Land-Cover and Phytoclimatic-Type Implications. *ISPRS International Journal of Geo-Information* 8 (1): 43.

Olsson, L. 1993. On the Causes of Famine: Drought, Desertification and Market Failure in the Sudan. *Ambio* 22 (6): 395-403.

Orhan, O, Ekercin, S and Dadaser-Celik, F. 2014. Use of landsat land surface temperature and vegetation indices for monitoring drought in the Salt Lake Basin Area, Turkey. *The Scientific World Journal* 2014

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R and Dubourg, V. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12 (Oct): 2825-2830.

Peters, AJ, Walter-Shea, EA, Ji, L, Vina, A, Hayes, M and Svoboda, MD. 2002. Drought monitoring with NDVI-based standardized vegetation index. *Photogrammetric engineering and remote sensing* 68 (1): 71-75.

Poursanidis, D and Chrysoulakis, N. 2017. Remote Sensing, natural hazards and the contribution of ESA Sentinels missions. *Remote Sensing Applications: Society and Environment* 6 25-38.

Qiu, J, Yang, J, Wang, Y and Su, H. 2018. A comparison of NDVI and EVI in the DisTrad model for thermal sub-pixel mapping in densely vegetated areas: a case study in Southern China. *International Journal of Remote Sensing* 39 (8): 2105-2118.

Reid, P and Vogel, C. 2006. Living and responding to multiple stressors in South Africa—Glimpses from KwaZulu-Natal. *Global Environmental Change* 16 (2): 195-206.

Richard, Y, Fauchereau, N, Poccard, I, Rouault, M and Trzaska, S. 2001. 20th century droughts in southern Africa: spatial and temporal variability, teleconnections with oceanic and atmospheric conditions. *International Journal of Climatology* 21 (7): 873-885.

Rodriguez, JJ, Kuncheva, LI and Alonso, CJ. 2006. Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence* 28 (10): 1619-1630.

Rojas, O, Vrieling, A and Rembold, F. 2011. Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. *Remote sensing of Environment* 115 (2): 343-352.

Roli, F and Fumera, G. 2001. *Support vector machines for remote sensing image classification*. SPIE,

Rosenqvist, Å, Milne, A, Lucas, R, Imhoff, M and Dobson, C. 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science & Policy* 6 (5): 441-455.

Rouault, M and Richard, Y. 2003. Intensity and spatial extension of drought in South Africa at different time scales. *water SA* 29 (4): 489-500.

Rouault, M and Richard, Y. 2005. Intensity and spatial extent of droughts in southern Africa. *Geophysical Research Letters* 32 (15):

SA Forestry. 2016. Challenges of forestry in Zululand – ICFR Zululand Field Day [Internet]. SA Forestry Online. Available from: http://saforestryonline.co.za/articles/2844/. [Accessed: 16 August 2019].

Saatchi, S, Asefi-Najafabady, S, Malhi, Y, Aragão, LE, Anderson, LO, Myneni, RB and Nemani, R. 2013. Persistent effects of a severe drought on Amazonian forest canopy. *Proceedings of the National Academy of Sciences* 110 (2): 565-570.

Sahu, AS. 2014. *Identification and mapping of the water-logged areas in Purba Medinipur part of Keleghai river basin, India: RS and GIS methods*.

Sandino, J, Gonzalez, F, Mengersen, K and Gaston, KJ. 2018. UAVs and machine learning revolutionising invasive grass and vegetation surveys in remote arid lands. *Sensors* 18 (2): 605.
Sarkar, S, Vinay, S, Pateshwari, V and Maiti, J.2016. Study of optimized SVM for incident prediction of a steel plant in India. *2016 IEEE Annual India Conference (INDICON)*, 1-6.
Schleeweis, K, Goward, SN, Huang, C, Dwyer, JL, Dungan, JL, Lindsey, MA, Michaelis, A, Rishmawi, K and Masek, JG. 2016. Selection and quality assessment of Landsat data for the North American forest dynamics forest history maps of the US. *International Journal of Digital Earth* 9 (10): 963-980.
Schönau, APG and Schulze, RE. 1984. Climatic and Altitudinal Criteria for Commercial Afforestation with Special Reference to Natal. *South African Forestry Journal* 130 (1): 10-18.
Schroeder, TA, Cohen, WB, Song, C, Canty, MJ and Yang, Z. 2006. Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. *Remote sensing of environment* 103 (1): 16-26.
Scott, DF, Le Maitre, DC and Fairbanks, D. 1998. Forestry and streamflow reductions in South Africa: A reference system for assessing extent and distribution.
Shackleton, RT, Shackleton, CM and Kull, CA. 2019. The role of invasive alien species in shaping local livelihoods and human well-being: A review. *Journal of Environmental Management* 229 145-157.
Shiferaw, H, Bewket, W and Eckert, S. 2019. Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem. *Ecology and evolution* 9 (5): 2562-2574.
Solh, M and van Ginkel, M. 2014. Drought preparedness and drought mitigation in the developing world′ s drylands. *Weather and Climate Extremes* 3 62-66.
Sun, S, Qiu, L, He, C, Li, C, Zhang, J and Meng, P. 2018. Drought-Affected Populus simonii Carr. Show Lower Growth and Long-Term Increases in Intrinsic Water-Use Efficiency Prior to Tree Mortality. *Forests* 9 (9): 564.
Tenny, S and Hoffman, MR. 2019. Prevalence. [Internet]. StatPearls. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430867/. [Accessed: 30 August 2019].
Testa, S, Soudani, K, Boschetti, L and Borgogno Mondino, E. 2018. MODIS-derived EVI, NDVI and WDRVI time series to estimate phenological metrics in French deciduous forests. *International Journal of Applied Earth Observation and Geoinformation* 64 132-144.
Tewari, DD. 2001. Is commercial forestry sustainable in South Africa? The changing institutional and policy needs. *Forest Policy and Economics* 2 (3-4): 333-353.
Tucker, CJ. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment* 8 (2): 127-150.
Tuswa, N, Bugan, RDH, Mapeto, T, Jovanovic, N, Gush, M, Kapangaziwiri, E, Dzikiti, S, Kanyerere, T and Xu, Y. 2019. The impacts of commercial plantation forests on groundwater recharge: A case study from George (Western Cape, South Africa). *Physics and Chemistry of the Earth, Parts A/B/C* 112 187-199.
Van Westen, C and Soeters, R. 2002. Remote sensing and geographic information systems for natural disaster management. *Environmental modelling with GIS and remote sensing* 200-226.
Vapnik, V. 2013. *The nature of statistical learning theory*. Springer science & business media,

Vogel, C. 1994. (Mis) management for droughts in South Africa: past, present and future. *South African Journal of Science* 90 (1): 4-5.
Wang, L, You, Z-H, Xia, S-X, Chen, X, Yan, X, Zhou, Y and Liu, F. 2018. An improved efficient rotation forest algorithm to predict the interactions among proteins. *Soft Computing* 22 (10): 3373-3381.
Wang, M, Gu, Q, Liu, G, Shen, J and Tang, X. 2019. Hydrological Condition Constrains Vegetation Dynamics for Wintering Waterfowl in China's East Dongting Lake Wetland. *Sustainability* 11 (18): 4936.
Warburton, ML and Schulze, RE. 2008. Potential impacts of climate change on the climatically suitable growth areas of Pinus and Eucalyptus: results from a sensitivity study in South Africa. *Southern Forests: a Journal of Forest Science* 70 (1): 27-36.
Wardlow, BD and Egbert, SL. 2010. A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas. *International Journal of Remote Sensing* 31 (3): 805-830.
Wei, XA, van Noordwijk, M, Creed, IF, Gush, M, Ellison, D, Blanco, JA, Bishop, K, McNulty, SG, Archer, E and Tobella, AB. 2018. Forest landscape hydrology in a 'New Normal'era of climate and land use change. *In: Forest and Water on a Changing Planet: Vulnerability, Adaptation and Governance Opportunities Assessment Report (IF Creed and M. van Noordwijk, eds.). IUFRO World Series* 38 81-99.
Wilhite, D. 2005. Drought and water crisis: science, technology, and management issues. p 406.
Wilhite, DA and Glantz, MH. 1985. Understanding: the drought phenomenon: the role of definitions. *Water international* 10 (3): 111-120.
Woodcock, CE, Allen, R, Anderson, M, Belward, A, Bindschadler, R, Cohen, W, Gao, F, Goward, SN, Helder, D and Helmer, E. 2008. Free access to Landsat imagery. *Science* 320 (5879): 1011-1011.
Wu, D, Qu, JJ and Hao, X. 2015. Agricultural drought monitoring using MODIS-based drought indices over the USA Corn Belt. *International Journal of Remote Sensing* 36 (21): 5403-5425.
Wu, J, Zhou, L, Liu, M, Zhang, J, Leng, S and Diao, C. 2013. Establishing and assessing the Integrated Surface Drought Index (ISDI) for agricultural drought monitoring in mid-eastern China. *International Journal of Applied Earth Observation and Geoinformation* 23 397-410.
Wulder, MA, Masek, JG, Cohen, WB, Loveland, TR and Woodcock, CE. 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment* 122 2-10.
Xia, J, Du, P, He, X and Chanussot, J. 2013. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters* 11 (1): 239-243.
Xia, Y, Liu, C, Li, Y and Liu, NJESwA. 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. 78 225-241.
Xulu, S, Peerbhay, K, Gebreslasie, M and Ismail, R. 2018. Drought influence on forest plantations in Zululand, South Africa, using MODIS time series and climate data. *Forests* 9 (9): 528.
Xulu, S, Peerbhay, K, Gebreslasie, M and Ismail, R. 2019. Unsupervised Clustering of Forest Response to Drought Stress in Zululand Region, South Africa. *Forests* 10 (7): 531.
Ye, S, Pontius, RG and Rakshit, R. 2018. A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS Journal of Photogrammetry and Remote Sensing* 141 137-147.

Yin, S and Yin, J. 2016. Tuning kernel parameters for SVM based on expected square distance ratio. 370 (C %J Inf. Sci.): 92–102.
Young, NE, Anderson, RS, Chignell, SM, Vorster, AG, Lawrence, R and Evangelista, PH. 2017. A survival guide to Landsat preprocessing. *Ecology* 98 (4): 920-932.
Yuan, X, Wang, L and Wood, EF. 2018. Anthropogenic intensification of southern African flash droughts as exemplified by the 2015/16 season. *Bulletin of the American Meteorological Society* 99 (1): S86-S90.
Zambrano, F, Lillo-Saavedra, M, Verbist, K and Lagos, O. 2016. Sixteen years of agricultural drought assessment of the BioBío region in Chile using a 250 m resolution Vegetation Condition Index (VCI). *Remote Sensing* 8 (6): 530.
Zargar, A, Sadiq, R, Naser, B and Khan, FI. 2011. A review of drought indices. *Environmental Reviews* 19 (NA): 333-349.
Zerrouki, N and Bouchaffra, D.2014. Pixel-based or Object-based: Which approach is more appropriate for remote sensing image classification? *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 864-869.
Zhang, A and Jia, G. 2013. Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data. *Remote Sensing of Environment* 134 12-23.
Zhang, F, Du, B and Zhang, L. 2015. Scene classification via a gradient boosting random convolutional network framework. *IEEE Transactions on Geoscience and Remote Sensing* 54 (3): 1793-1802.
Zhang, L, Jiao, W, Zhang, H, Huang, C and Tong, Q. 2017a. Studying drought phenomena in the Continental United States in 2011 and 2012 using various drought indices. *Remote Sensing of Environment* 190 96-106.
Zhang, X, Chen, N, Li, J, Chen, Z and Niyogi, D. 2017b. Multi-sensor integrated framework and index for agricultural drought monitoring. *Remote sensing of environment* 188 141-163.
Zhang, Y, Peng, C, Li, W, Fang, X, Zhang, T, Zhu, Q, Chen, H and Zhao, P. 2013. Monitoring and estimating drought-induced impacts on forest structure, growth, function, and ecosystem services using remote-sensing data: recent progress and future challenges. *Environmental Reviews* 21 (2): 103-115.
Zhu, Z, Fu, Y, Woodcock, CE, Olofsson, P, Vogelmann, JE, Holden, C, Wang, M, Dai, S and Yu, Y. 2016. Including land cover change in analysis of greenness trends using all available Landsat 5, 7, and 8 images: A case study from Guangzhou, China (2000–2014). *Remote Sensing of Environment* 185 243-257.
Zhu, Z and Woodcock, CE. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment* 118 83-94.
Zhuo, W, Huang, J, Zhang, X, Sun, H, Zhu, D, Su, W, Zhang, C and Liu, Z.2016. Comparison of five drought indices for agricultural drought monitoring and impacts on winter wheat yields analysis. *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1-6. IEEE,
Zucchini, W and Adamson, PT. 1984. The occurrence and severity of droughts in South Africa.