Predicting inter-seasonal aboveground grass biomass using Sentinel-2 MSI and machine learning in the Umngeni catchment, KwaZulu-Natal, South Africa

By

Mohamed Ismail Vawda

Student Number: 217008595

Submitted in fulfilment of the academic requirements of

Master of Science

in Environmental Science

Discipline of Geography

School of Agricultural, Earth and Environmental Science

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

February 2023

PREFACE

The research contained in this dissertation was completed by the candidate while based in the Discipline of Geography, School of Agricultural, Earth and Environmental Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The Water Research Commission of South Africa is acknowledged for partial funding through the WRC Project, No. CON2020/2021-00490, "Geospatial modelling of rangelands productivity in water-limited environments of South Africa". The research was also funded by the National Research Foundation of South Africa (Grant Number: MND210517601900).

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

[redacted]

Dr. Romano Lottering
Supervisor

Date: 08/02/2023

[redacted]

Prof. Onisimo Mutanga
Co-supervisor

Date: 08/02/2023

DECLARATION 1: PLAGIARISM

I, Mohamed Ismail Vawda declare that:

(i) the research reported in this thesis, except where otherwise indicated or acknowledged, is my original work;

(ii) this thesis has not been submitted in full or in part for any degree or examination to any other university;

(iii) this thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this thesis 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, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this thesis is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

[redacted]

Mohamed Vawda
Student

Date: 06/02/2023

DECLARATION 2- PUBLICATIONS

Details of contribution to publications that form part of and/or include research presented in this thesis (includes publications in preparation, submitted, in press and published and give details of the contributions of each author to the experimental work and writing of each publication).

Publication 1: Vawda, M.I., Lottering, R., Mutanga, O. and Peerbhay, K. **In preparation**: Comparing the utility of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) in tandem with Sentinel-2 MSI in estimating dry season aboveground grass biomass.

Publication 2: Vawda, M.I., Lottering, R., Mutanga, O. and Peerbhay, K. **In preparation**: Predicting inter-seasonal grass biomass using satellite remote sensing in the Vulindlela area of the Umgeni catchment, KwaZulu-Natal.

The work was done by the first author under the guidance and supervision of the remaining three authors.

Discipline of Geography, School of Agricultural, Earth and Environmental Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville, 3209, South Africa

ABSTRACT

Grasslands contribute significantly to socio-economic growth and ecological wellbeing, especially in South Africa. The productivity of grasslands is of substantial interest to researchers and rangeland managers alike, particularly in the face of climate change. However, grassland productivity fluctuates inter-seasonally. Hence, finding innovative, accurate and low-cost solutions to monitor grasslands is imperative. Recent advancements in remote sensing and machine learning provide a rigorous, cost-effective and timeous information that is useful for vegetation monitoring and management. In addition, contemporary deep learning algorithms create an opportunistic pathway to advance vegetation remote sensing research.

Hence, the focus of this study was to investigate inter-seasonal changes in grassland productivity using aboveground grass biomass as a proxy. This study utilised Sentinel-2 MSI with a Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN), to predict aboveground grass biomass between the dry and wet seasons. Sentinel-2 MSI spectral bands and thirty derived vegetation indices were used as input data to train the models. This study was divided into two overarching objectives. Firstly, the performance of the two neural networks was compared to ascertain which model was more adept at biomass predictions. Thereafter, the better performing algorithm was used to distinguish aboveground grass biomass between the dry and wet seasons.

In comparing the performance of the traditional ANN and the contemporary CNN, findings showed that the deep CNN outperformed the ANN in estimating dry season grass biomass. The deep CNN attained an R^2 of 0,83, an RMSE of 3,36 g/m^2 and an RMSE% of 6,09. Comparatively, the ANN attained an R^2 of 0,75, an RMSE of 5,78 g/m^2 and an RMSE% of 8,90. The sensitivity analysis suggested that the Sentinel-2 blue band, Green Chlorophyll Index (GCl) and Green Normalised Difference Vegetation Index (GNDVI) were the most important variables for model development for both the CNN and ANN. The resulting biomass prediction maps captured the spatial variation in grass biomass as predicted by the two models, with the CNN model producing a more accurate representation of field data.

The deep CNN was utilised to distinguish changes in aboveground biomass between the dry and wet seasons, based on its higher accuracy. The CNN performed commendably in predicting biomass across the two seasons, yielding an R^2 of 0,83, an RMSE of 3,36 g/m^2 and an RMSE% of 6,09 in the dry season and an R^2 of 0,85, an RMSE of 2,41 g/m^2 and an RMSE% of 3,71 in

the wet season. The variables with the highest impact on model development for both seasons were the Sentinel-2 blue band, GCl and GNDVI, as noted previously. Changes in biomass were associated with changes in precipitation and rainfall. Biomass prediction maps portrayed the change in aboveground grass biomass from the dry to the wet season.

This is considered a pilot study as it illustrated the utility of deep learning algorithms for vegetation remote sensing research. Furthermore, it showcased the potential of applying deep CNNs with open-access remotely sensed data in creating a synergistic and intricate technique for geospatial modelling. Such technology can be used to make informed, strategic decisions in managing grasslands at different scales and in different contexts.

ACKNOWLEDGEMENTS

First and foremost, all glory and praise be unto Him for bringing me this far in my career and providing me with an opportunity to contribute to the generation of knowledge. “Verily, after every hardship there is ease”- The Holy Quran 94:6.

This research is dedicated to the late Dr. Riyad Ismail who will always be my guiding light and pushed me to pursue my Masters.

“We stand on the shoulders of giants”- Sir Isaac Newton. I want to express my sincere appreciation and admiration for my supervisors, Dr. Romano Lottering and Prof. Onisimo Mutanga. Your guidance, assistance and overall contribution to this research cannot be quantified and I am eternally grateful for this opportunity. An extended thank you to Dr. Peerbhay, Dr. Sibanda, Prof. Dube, Prof. Mabhaudhi, Prof. Odindi and the WRC project panel for their guidance and good will.

An enormous thank you to my loving wife, Tasmeeyah, for supporting me through the thick and thin of this degree. I cannot forget my dear parents, caring in-laws and extended family who never quiet comprehended what I do, but were there for me nevertheless. This would not have been possible without you all.

Thank you to my research colleagues: Anita, Xolile, Israel, Mthembeni, Adeola and Lwando, who assisted me with field and lab work. To the UKZN Department of Geography, especially Brice Gijsbertsen and Prof. Trevor Hill, for allowing me access to the labs and equipment. To the Grassland Science staff, especially Dr. Michelle Tedder, for assisting me with additional lab equipment.

Finally, I thank the Water Research Commission of South Africa, UKZN and the National Research Foundation for funding this research. Thank you to the South African Weather Services for providing me with climatic data.

CONTENTS

LIST OF ACRONYMS

ANFIS- Adaptive Neuro-Fuzzy Inference System

ANN- Artificial Neural Network

CNN- Convolutional Neural Network

GPS- Global Positioning System

MLR- Multiple Linear Regression

MODIS- Moderate Resolution Imaging Spectroradiometer

MSI- Multispectral Instrument

MSR- Multiple Stepwise Regression

NIR- Near Infrared

NLR- Non-Linear Regression

PLSR- Partial Least Squares Regression

R^2- Co-efficient of Determination

RF- Random Forest

RMSE- Root Mean Square Error

RMSE%- Percentage Root Mean Square Error

RNN- Recurrent Neural Network

SAWS- South African Weather Services

SPLSR- Sparse Partial Least Squares Regression

SVM- Support Vector Machine

SWIR- Shortwave Infrared

UAV- Unmanned Aerial Vehicle

VI- Vegetation Index/Indices

XGBoost- Extreme Gradient Boosting

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

Grasslands, also known as steppes, prairies and meadows in other regions of the world, are defined as areas of land dominated by grasses from the Poaceae family (Boval and Dixon, 2012). Grasses are broadly defined as herbaceous monocots with narrow leaves, a generally well-developed underground root system and a sward-forming canopy (Boval and Dixon, 2012). Grasslands cover approximately 30 to 40% of the earth’s terrestrial area and are found on every continent apart from Antarctica (O'Mara, 2012). Grasslands cover approximately a quarter of South Africa and is the third largest biome present in southern Africa (Mucina and Rutherford, 2006, Van den Hoof et al., 2018). Furthermore, grasslands are natural areas of high economic, social and ecological value in both local and global contexts (Singh et al., 2018).

Grasslands produce grains that are used for manufacturing edibles, they are a food source for livestock and wildlife and consist of numerous plants that are used in the medicinal and pharmaceutical industries (Gxasheka et al., 2017). These industries, and various others, directly benefit from grasslands and hence ensure local economic growth. Grasslands also hold a significant social standing in South Africa. For example, locals residing in rural areas and former homelands heavily depend on communal livestock farming as a livelihood (Gxasheka et al., 2017). Communal grasslands are extensively used as a fodder source for livestock and for other medicinal and edible plants (Gxasheka et al., 2017). Communal rangelands provide households with status, food, income and savings which sustain rural communities (Rasch et al., 2016).

In addition, the ecological role that grasslands play cannot be understated. These are often considered ecosystem services that directly or indirectly benefit human civilisation. Grasslands are integral in the prevention of soil erosion, and the promotion of arable land. They provide suitable habitats and niches for biodiversity to thrive, amongst many other ecological roles (Masenyama et al., 2022). However, an area of contemporary interest is on the role of grasslands in the combat against climate change. Whereas, there is a global notion that the proliferation of forests could increasingly assist in carbon sequestration and regulating atmospheric carbon stocks. Recent studies show that grasslands can play a more decisive role in combatting climate change (O'Mara, 2012). A study by Dass et al. (2018) notes that grasslands are more resilient at withstanding adverse climate changes such as increased drought, heat waves and fires and hence act as more resilient carbon sinks than forests. Grasslands store most of their carbon in the soil which acts as a safer carbon sink than forests that store most of their carbon in aboveground structures (Dass et al., 2018).

Numerous biophysical factors drive and maintain grasslands. Grasslands are influenced by biotic and abiotic factors (Boval and Dixon, 2012). Biotic factors include grazing, soil microbial organisms, pollinators, seed dispersal agents and other vegetation types such as trees and shrubs (Koerner and Collins, 2014). Abiotic factors include precipitation, temperature, altitude, fire, soil moisture and solar radiation. Both biotic and abiotic factors are complex and interact at various spatio-temporal scales to sustain grassland ecosystems (Koerner and Collins, 2014). Koerner and Collins (2014) suggested that precipitation, grazing and fire are the most influential biophysical factors in South African temperate grasslands. These grasslands are generally classified into sweetveld or sourveld grasslands (Dedekind et al., 2020). Dedekind et al. (2020) describe sourveld grasslands as productive and stable grasslands that occur in regions with high rainfall, low altitude, dystrophic soils and have perennial swards. This enables sourveld grasslands to support high livestock production levels (Dedekind et al., 2020). Moyo et al. (2013) characterise sweetveld grasslands as nutritious grasslands that occur in regions with low rainfall, high altitude and fine clay-like soils with poor drainage. Sweetveld grasslands are unsuitable for stocking high rates of livestock due to their higher sensitivity to external factors (Moyo et al., 2013). Some grasslands exhibit a mixture of sweetveld and sourveld and are termed mixed veld grasslands (Ellery et al., 1995).

The stark difference between the sweetveld and sourveld is during the dry winter season. Sourveld grass species tend to withdraw key nutrients such as nitrogen and most minerals from the aboveground sward and store them in the underground roots (Dedekind et al., 2020). This results in dry foliage with low protein and high fibre content, decreasing its nutritive value and palatability (Dedekind et al., 2020). On the contrary, sweetveld grass species remain palatable and nutritious throughout the year, even during the dry winter, since they do not translocate their nutrients to underground reserves (Moyo et al., 2013). This phenomenon is substantially due to precipitation and soil type with higher rainfall in well-drained soils resulting in higher leaching in sourveld grasslands and vice-versa (Ellery et al., 1995, Dedekind et al., 2020).

Essentially, the limiting factor in the sourveld is the quality of forage and the limiting factor in the sweetveld is the quantity of the forage (Ellery et al., 1995).

It is inherently evident that maintaining grasslands and rangelands is vital in maintaining livestock production for socio-economic reasons. Therefore, it is vital to quantify the functioning of grasslands to understand and manage them optimally. Grassland productivity is a term used to assess and gauge the ability of a grassland to support and maintain living organisms, such as wildlife and livestock (Vundla et al., 2020). Grassland productivity can be measured using various parameters or variables. These include grass nutrients (Singh et al., 2018), above-ground biomass (Vundla et al., 2020), grass water content (Sibanda et al., 2021) and leaf area index (Dube et al., 2019). These are all vegetative parameters that can be measured and quantified and provide insight into the productivity of a particular grassland. These parameters enable rangeland managers and scientists to make informed decisions and allow for better planning (Vundla et al., 2020, Sibanda et al., 2021).

To assess historical grassland or rangeland productivity, scientists or managers had to physically plan and execute sampling techniques in the field. However, this is often physically challenging, time-intensive, redundant and expensive process (Ramoelo et al., 2015, Mutanga et al., 2016). The application of remote sensing has significantly advanced over the years and has enabled scientists and professionals alike to study various environmental aspects in croplands, forests and rangelands (Ramoelo et al., 2015, Mutanga et al., 2016). Satellite imagery has become more efficient, highly detailed, relatively inexpensive and easily accessible, particularly within a South African context (Mutanga et al., 2016). Vegetation monitoring has now become a smoother, faster and highly technical activity with an interface between spaceborne satellite data and ground data providing and verifying information (Mutanga et al., 2016). Satellite remote sensing has been used extensively to monitor vegetation condition, both globally and in South Africa, with satellite data proving to be useful and accurate in quantifying and predicting the vegetation variables used to discern such conditions (Mutanga et al., 2016). The inclusion of machine learning in remote sensing studies is now a norm, with machine learning proving to be a useful tool in analysing and interpreting spatial information (Ali et al., 2015).

This research focuses on the aboveground biomass parameter, which has extensively been studied and documented in literature. Vegetative biomass studies have been conducted worldwide across various vegetation types such as forests and grasslands using satellite remote sensing. Ali et al. (2016) estimated biomass in Ireland's intensively managed grasslands using MODIS data and multiple linear regression, artificial neural network and adaptive neuro-fuzzy inference system models. In the United States of America, Wang et al. (2019) estimated leaf area index and aboveground biomass in grazing pastures using Sentinel 1, Sentinel-2 and Landsat 8 data combined with multiple linear regression, support vector machine and random forest models. A Mongolian biomass study by Pang et al. (2020) utilised Sentinel-2 imagery and partial least-squares regression and multiple stepwise regression models to estimate grassland productivity and analyse ecosystem condition. In a more recent study, Li et al. (2021) combined Sentinel-2 multispectral data with extreme gradient boosting and random forest algorithms to model aboveground grass biomass in a wetland system in China. The studies above and their main findings are summarised in the Table 1.1:

Table 1.1: A summary of international remote sensing studies of vegetation biomass.

Author	Context	Satellite	Metrics/Indices	Algorithm	Findings
Ali et al. (2017)	Grassland	MODIS	Spectral bands/NDVI, EVI2, SAVI, MSAVI, OSAVI	Multiple linear regression (MLR) Artificial neural network (ANN) Adaptive neuro-fuzzy inference system (ANFIS)	The ANFIS produced the most accurate biomass estimation as compared to the ANN and MLR. The performance of ANFIS and ANN will work better with higher resolution sensors such as Sentinel.
Wang et al. (2019)	Grassland	Sentinel 1 Sentinel-2 Landsat 8	NDVI, EVI, LSWI	MLR Support vector machine (SVM) Random forest (RF)	The integration of satellite data from all three sensors provided the most accurate results, especially for studying seasonal dynamics. The MLR model generally performed better than the SVM and RF models.
Pang et al. (2020)	Grassland	Sentinel-2	Spectral bands/ Simulated spectral bands/ various VIs derived from both spectral and simulated bands	Partial least squares regression (PLSR) Multiple stepwise regression (MSR)	The simulated spectral bands and associated indices produced more accurate biomass predictions than the raw spectral bands and associated indices
Li et al. (2021)	Grassland/ Wetland	Sentinel-2	Spectral bands/ Various VIs/ Various red edge indices/ Gray-level co-occurrence texture matrix	RF Extreme gradient boosting (XGBoost)	The texture matrix moderately improved biomass estimation as compared to VIs and red edge indices.

	However, the use of the texture matrix alone was not useful and should only be used to supplement VIs and red edge indices.

In a more local context, numerous aboveground biomass studies have been conducted in southern Africa. For example, Samimi and Kraus (2004) utilised Landsat 5 and Landsat 7 in biomass estimation studies across Kruger National Park and Madikwe Game Reserve in South Africa and the Gutu District in Zimbabwe. Palmer et al. (2010) estimated green biomass changes for the period 2000 to 2009 using MODIS imagery for grasslands in the Kokstad area of KwaZulu-Natal. In a separate study, Dube and Mutanga (2015) also quantified aboveground biomass using Landsat 8 multispectral data, albeit for forest plantations located in the uMgeni catchment area of KwaZulu-Natal. A study by Sibanda et al. (2017) in Pietermaritzburg utilised WorldView 3 satellite imagery and the sparse partial least squares regression model to estimate grassland biomass in an intensively managed grassland.

Furthermore, Shoko et al. (2018) determined the optimal satellite imagery for estimating aboveground biomass in the Drakensberg. They compared data from Sentinel-2, Landsat 8 and WorldView 2 while utilising the sparse partial least square regression algorithm to determine the most suitable sensor for predicting aboveground biomass. Lastly, Vundla et al. (2020) sought to evaluate the impact of wattle invasions on grass biomass in the Matatiele district of Eastern Cape using Sentinel-2 satellite data and the partial least squares regression model. A summary of the abovementioned southern African studies and their findings are tabulated in Table 1.2:

Table 1.2: A summary of local vegetation biomass studies.

Author	Context	Satellite	Metrics/Indices	Algorithm	Findings
Samimi & Kraus (2004)	Grassland Savannah	Landsat 5 Landsat 7	Spectral bands	Non-linear regression (NLR)	Significant correlations between satellite spectral data and field data for all foliage types.
Palmer et al. (2010)	Grassland	MODIS	NDVI/Spectral bands	NLR	Created a model that predicts the change in biomass over a given period.
Dube & Mutanga (2015)	Forest	Landsat 8	Spectral bands/DVI, GEMI, GNDVI, MSAVI, MSI, NDII, NDVI, NDVIc, OSAVI, RDVI, RSR, SAVI, SAVI2, SR	Stochastic gradient boosting (SGB) RF	Biomass estimates are more accurate when using both spectral data and vegetation indices. Landsat 8 data yields more accurate estimations than Landsat 7.
Sibanda et al. (2017)	Grassland	WorldView 3	Red edge vegetation indices/ Texture	Sparse partial least squares regression (SPLSR)	The combination of red edge indices and texture metrics significantly improves accuracy of aboveground biomass predictions.
Shoko et al. (2018)	Grassland	Sentinel-2 Landsat 8 Worldview 2	Spectral bands/EVI, SAVI, StNDVI, RDVI, SR, MSR	SPLSR	WorldView 2 displayed the best predictive accuracies followed by Sentinel-2 and then Landsat 8 in biomass predictions. Spectral bands within the red edge, SWIR and NIR as well as derived

					indices were the most important in all sensors for assessing biomass.
Vundla et al. (2020)	Grassland	Sentinel-2	Spectral bands/NDVI/SR	PLSR	The most influential metrics in estimating biomass were red edge derived indices, especially the SR.

As alluded to previously, grasslands are driven by various biophysical aspects. This research focuses on the effect of precipitation on grassland productivity, specifically grass biomass. Van den Hoof et al. (2018) stated that grasslands in South Africa are very responsive to variations in precipitation over different spatio-temporal scales. This research attempts to assess the differences in grassland productivity due to inter-seasonal variations in precipitation, with a particular focus on biomass. South Africa has distinct dry seasons and wet seasons that influence grasslands' productivity and functioning (Van den Hoof et al., 2018). However, research on the effects of inter-seasonal fluctuations on grasslands is somewhat limited in the scientific literature (Masenyama et al., 2022).

Ramoelo and Cho (2014) estimated grass biomass in the dry season using reflectance data from RapidEye and Landsat 8 and a random forest algorithm. This study was conducted in the north-eastern region of South Africa, more specifically Kruger National Park, Sabi Sands and Bushbuckridge. Furthermore, a study by Ramoelo et al. (2015) monitored grass nutrients and biomass between wet and dry seasons in Sabi Sands, located in the north-eastern region of South Africa. Ramoelo et al. (2015) utilised WorldView 2 data and random forest modelling to monitor leaf nitrogen content and aboveground biomass as a proxy for rangeland condition. Dingaan and Tsubo (2019) conducted grassland availability studies in Bethlehem, Bloemfontein, Kimberly and Johannesburg. They used MODIS data and regression models to differentiate between green aboveground biomass in the wet season and non-green aboveground biomass in the dry season. The studies alluded to above are summarised in Table 1.3:

Table 1.3: A summary of local grass biomass studies focusing on seasonal variability.

Author	Parameter	Satellite	Metrics/Indices	Algorithm	Findings
Ramoelo & Cho (2014)	Biomass	RapidEye Landsat 8	Spectral bands	RF	Spectral reflectance data can be used to estimate biomass in the dry season accurately. RapidEye data produced better results than Landsat 8.
Ramoelo et al. (2015)	Nitrogen content Biomass	WorldView 2	Spectral bands/ Traditional VIs/ Red edge VIs	RF	The red edge band and derived indices were the most important in assessing both Nitrogen content and biomass for both seasons.
Dingaan & Tsubo (2019)	Biomass	MODIS	Spectral bands/ Traditional VIs	NLR	Green biomass was aptly estimated by NDVI. Non-green biomass was more accurately estimated by NDWI.

In conclusion, grasslands are the focal point for livestock production in South Africa. However, research on the dynamics of this biome is limited and thus hinders the management of grasslands. Inter-seasonal fluctuations and climate-induced changes continue manifesting in grassland ecosystems at varying spatial and temporal scales. It is therefore essential to further research in this arena to promote adaptability and resilience in future. There is an evident research gap in the literature as inter-seasonal grass biomass studies for the KwaZulu-Natal region are scarce. This research aims to fill this gap by predicting aboveground biomass between the wet season and dry seasons in KZN sourveld grasslands. Many of the studies above were conducted in the northern regions of South Africa and have utilised certain satellites and algorithms (See Table 1.3). This research can build upon this foundation and advance grassland studies in the KZN region to benefit rangeland managers by assisting them in planning and decision-making. This research can ensure that optimal livestock production can be maintained inter-seasonally in both a communal and commercial context.

1.2 Aim

The research aimed to predict inter-seasonal aboveground grass biomass using Sentinel-2 MSI and machine learning algorithms in the Umngeni catchment, KwaZulu-Natal.

1.3 Objectives

The broad objectives of this study were to:

1) Compare the performance of traditional Artificial Neural Networks (ANN) and deep Convolutional Neural Networks (CNN) in assessing aboveground biomass using Sentinel-2 data.

2) Predict inter-seasonal (dry and wet season) aboveground grass biomass using Sentinel-2 and deep learning technique (CNN).

1.4 Research Questions

1) Which machine learning technique, between the Artificial Neural Network and the Convolutional Neural Network, performs more aptly at estimating aboveground biomass of grass when paired with Sentinel-2 bands and derived vegetation indices?

2) Can remote sensing and deep learning be used to estimate and assess the difference in grass aboveground biomass between two distinct seasons in South Africa, the dry and wet seasons? And what can this change in biomass be attributed to?

1.5 Thesis Structure

This thesis has four chapters, with chapters 2 and 3 considered independent manuscripts. Hence, these chapters contain their own focused introduction, methods, results and discussion sections. It should be noted that chapters 2 and 3 share similarities and linkages as they ultimately address the same overarching aim of the study.

The chapters are as follows:

Chapter one serves as a general introduction to the thesis, highlighting the importance of grasslands and remote sensing. In addition, previous studies that share similarities with those in this study are briefly outlined. The aim, objectives and research questions are included in this chapter.

Chapter two acts as the first standalone manuscript comparing the performance of two neural networks' in estimating aboveground grass biomass. Sentinel-2 spectral data and derived vegetation indices were used to assess predictive performance of both machine learning algorithms to determine which was more apt to predict biomass.

Chapter three acts as the second manuscript in which the better performing neural network from the previous chapter was utilised to assess and determine the difference in grass biomass between the dry and wet seasons. Once again, Sentinel-2 MSI data and VIs were used to predict grass biomass. Possible explanations of biomass change are suggested in this chapter.

Chapter four serves as a synthesis of the two previous chapters. Significant findings and conclusions are consolidated in this chapter. Furthermore, limitations and recommendations are briefly noted in this chapter.

CHAPTER 2

COMPARING THE UTILITY OF ARTIFICIAL NEURAL NETWORKS (ANN) AND CONVOLUTIONAL NEURAL NETWORKS (CNN) ON SENTINEL-2 MSI TO ESTIMATE DRY SEASON ABOVEGROUND GRASS BIOMASS

Abstract

Grasslands are biomes of significant fiscal, social and environmental value. Grassland or rangeland management often monitors and manages grassland productivity. Productivity is determined by various biophysical parameters, one such being grass aboveground biomass. Advancement in remote sensing has enabled near real-time monitoring of grassland productivity. Furthermore, the increase in sophisticated machine learning algorithms has provided a powerful tool for remote sensing analytics. This study compared the performance of two neural networks, namely: Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), in predicting dry season aboveground biomass using open-access Sentinel-2 MSI data. Sentinel-2 spectral bands and derived vegetation indices were used as input data for the two algorithms. Overall, findings in this study showed that the deep CNN outperformed the ANN in estimating aboveground biomass with an R^2 of 0,83, RMSE of 3,36 g/m^2 and RMSE% of 6,09. In comparison, the ANN produced an R^2 of 0,75, RMSE of 5,78 g/m^2 and RMSE% of 8,90. The sensitivity analysis suggested that the blue band, Green Chlorophyll Index (GCl) and Green Normalised Difference Vegetation Index (GNDVI) were the most significant for model development for both neural networks. This study can be considered a pilot study as it is one of the first to compare different neural network performance using freely available satellite data. This is useful for more rapid biomass estimation and this study exhibits the great potential of deep learning for remote sensing applications.

Keywords: Remote Sensing, Grasslands, Biomass, Artificial Neural Network, Convolutional Neural Network, Sentinel-2.

2.1 Introduction

The study and observation of natural phenomena are increasingly becoming more imperative as the world faces unprecedented environmental change (Ali et al., 2015). The consistent improvements made to airborne and spaceborne platforms and sensors have resulted in a proliferation of remote sensing research (Mutanga et al., 2016). Remote sensing has facilitated earth observations in various facets of the natural world, ranging from weather to vegetation. Vegetation monitoring, in particular, has become an influential research area in remote sensing academia due to the need for continuous and reliable data to assist in decision-making processes (Mutanga et al., 2016). The advent of remote sensing, from simple aerial photographs to current high-resolution imagery, has enabled scientists to study larger spatial and temporal scales (Mutanga et al., 2016). Recently, there has been a significant increase in remote sensing data and ground data in vegetation studies which have established a solid foundation for vegetation monitoring, presently and in the future (Ali et al., 2015).

The inundation of remotely sensed data has directed scientists towards finding novel methods for data processing and analysis (Ali et al., 2015). Remotely sensed data has proven to be voluminous, with data being captured at monthly, weekly and even hourly scales (Das et al., 2022). The heterogeneity of remotely sensed data, with a vast array of sensors at varying spatial, temporal and radiometric resolutions, has produced challenges in data processing and analysis (Ali et al., 2015, Das et al., 2022). This challenge has ushered scientists into discovering new methodologies at discerning multi-dimensional data, resulting in a paradigm shift from conventional statistical methods towards machine learning solutions (Das et al., 2022). Artificial intelligence advancement and machine learning technologies have enabled scientists and practitioners to address pressing environmental issues due to their real-time processing of data and strong predictive abilities (Das et al., 2022).

Neural networks, considered a subset of machine learning, are algorithms designed by mimicking of a biological brain's operation (Mas and Flores, 2008). The artificial neural network (ANN), specifically, has been extensively used for remote sensing applications since the 1990s as they provided promising results (Mas and Flores, 2008). Mas and Flores (2008) state that ANNs have been reported to perform much more admirably as compared to traditional statistical methods due to their ability to learn complex patterns, study nonlinear relationships between variables, generalisation abilities and perform various analyses without necessitating the meeting of data assumptions (eg. Normally distributed data). Jensen et al.

(2009) acknowledge that ANNs have been used relatively successfully in remote sensing for biophysical estimation and land classification. However, they have their limitations. These include the complex architectures of ANNs and their demanding computational requirements, the need for large amounts of training data and supervised algorithm training to ensure better accuracy and output (Mas and Flores, 2008, Jensen et al., 2009).

In the last decade, there has been another paradigm shift in machine learning, with the focus now on deep learning approaches (Liu et al., 2019). This is essentially a refinement and improvement on traditional ANNs, to improve predictive accuracy and reduce the complexity of the previous algorithms (Zhu et al., 2017). Zhu et al. (2017) define deep learning as neural networks characterised by more than two deep layers in the neural network structure that extract distinct feature patterns from input data. One such example of a deep neural network is the convolutional neural network (CNN), which has been specially engineered for image processing and analysis (Zhu et al., 2017, Kattenborn et al., 2021). The increased number of layers and interconnections in CNNs have meant that more complex and intricate patterns and relationships can be deciphered, which is particularly useful for vegetation remote sensing studies (Kattenborn et al., 2021). CNNs have an advantage over ANNs because they require less computational time and power and produce higher predictive accuracies. However, they require vast amounts of training data to be able to make such accurate predictions and classifications (Brodrick et al., 2019). Although the use of CNNs in remote sensing is trend-setting, it is currently only in its inception and has to be tested further to reveal its strengths and weaknesses (Kattenborn et al., 2021).

Grasslands are biomes of high socio-economic and conservational value, particularly in southern Africa (Palmer et al., 2010). Grasslands are highly sensitive to environmental change and are often moderated by biophysical factors such as rainfall and grazing (Vundla et al., 2020). Vegetation parameters are used to asses health and condition and can either be physically measured or remotely estimated by remote sensing (Mutanga et al., 2016). Aboveground biomass is one such vegetation measure used to observe and monitor grassland productivity (Ali et al., 2015). Neural networks, especially ANNs, have been used to assess and predict aboveground vegetation biomass for a considerable time (Ali et al., 2015). In most cases, ANNs have outperformed typical Bayesian and iterative statistical methods for estimating biomass, as well as other machine learning algorithms such as support vector machines (Ali et al., 2015). Recent grass biomass studies have gradually incorporated the utilisation of CNNs for biomass predictions, with varying results based on sensor resolutions and platform type (Kattenborn et al., 2021).

There is a substantial lack of grassland biomass studies, in relation to remote sensing, in South Africa, as reported by Masenyama et al. (2022). Furthermore, from a South African context, no research has attempted to investigate the performance of conventional ANNs and contemporary CNNs in estimating aboveground grass biomass. The study of grasslands is imperative in the face of climate change. Hence, the use of machine learning techniques to observe and assess grassland health would be valuable to both researchers and practitioners (Masenyama et al., 2022, Ali et al., 2015). Therefore, the objectives of this study were to: 1) compare the predictive performance of shallow ANNs and deep CNNs in estimating aboveground grass biomass using Sentinel-2 MSI and 2) utilise both neural networks on Sentinel-2 MSI to predict dry season aboveground biomass for Vulindlela communal area in South Africa.

2.2 Methods

2.2.1 Study Area

The Vulindlela area is situated in the greater Umngeni Catchment of the KwaZulu-Natal province, South Africa. The study area is part of the Umgungundlovu district and falls under the uMsunduzi Municipality (Figure 2.1). The local climate can be classified as a subtropical oceanic climate, with cool and dry winters and mild and wet summers (Sibanda et al., 2021). A mean annual rainfall of 979mm with a median annual rainfall of 850mm is received in the study area. Annual maximum and minimum temperatures are approximately 22°C and 10°C respectively (Sibanda et al., 2021). Vegetation growth in the area is limited primarily by climate, with low precipitation, low temperatures and frost being the major factors (Sibanda et al., 2021). Vulindlela has a mean annual potential evaporation that ranges between 1580 and 1620mm, which indicates a possible deficit in relation to mean annual rainfall (Sibanda et al., 2021). The edaphic factors of Vulindlela are characterised by shallow soils with moderate to poor drainage; this presents a potential soil erosion risk if not properly managed (Sibanda et al., 2021).

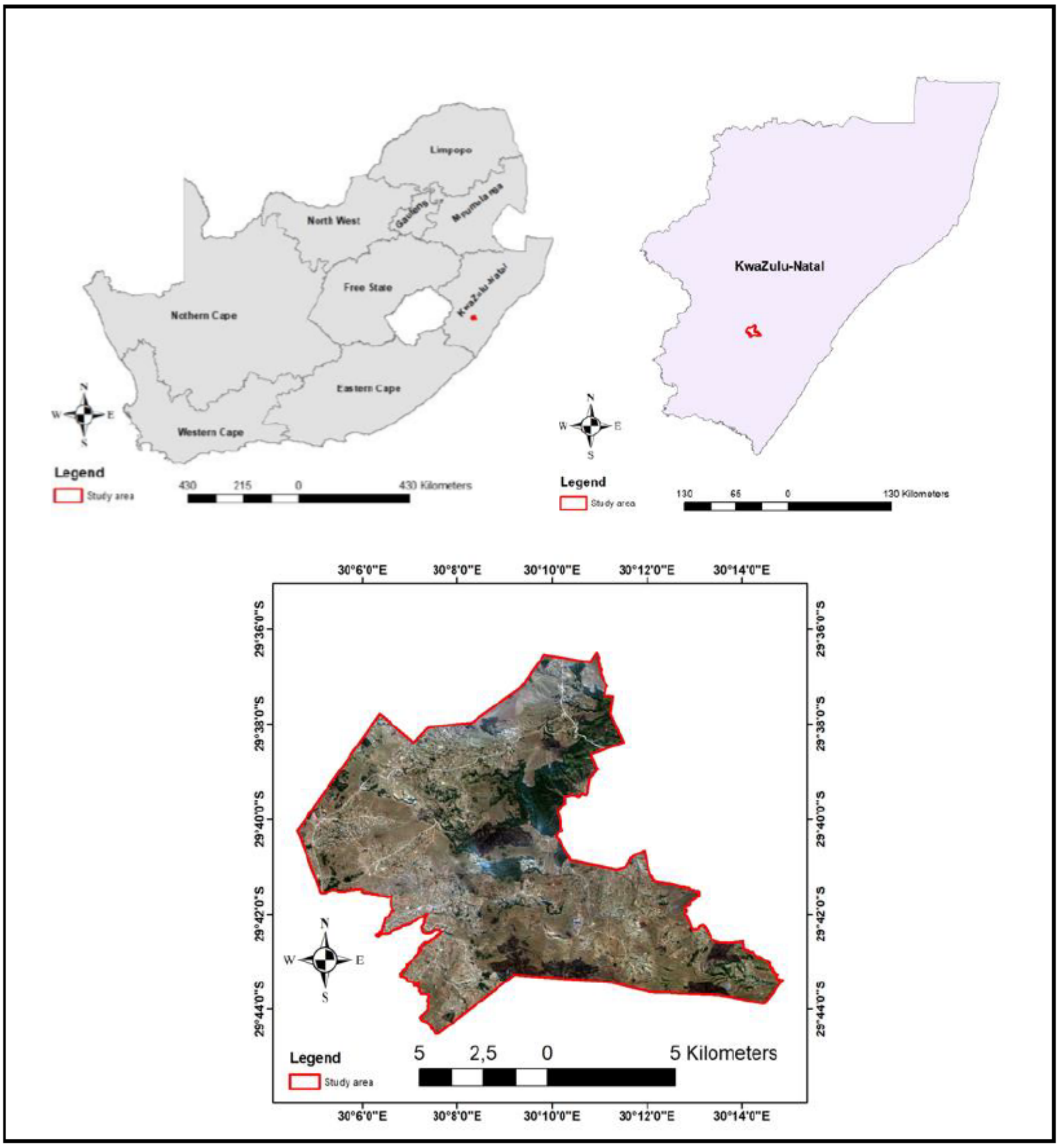


Figure 2.1: Location of Vulindlela area relative to South Africa and KZN.

Grasslands within the study area are characterised as mesic and typically consist of species such as *Themeda triandra,, Eragrostis tenuifolia, Tristachya leucothrix, Paspalum urvillei, Sorghum bicolour, Panicum maximum, Setaria sphacelate, Aristida junciformis* (Figure 2.2) and *Alloteropsis semialata* amongst others (Fynn et al., 2011). According to Scott-Shaw and Morris (2015), the most palatable species to livestock from the abovementioned grasses are *Themeda triandra*, *Tristachya leucothrix* and *Aristida junciformis*, in order of palatability. Masemola et al. (2020) state that the dry season in KwaZulu-Natal usually spans between June and July whereas Roffe et al. (2020) state that the wet season in KwaZulu-Natal typically spans between October and March. Grasslands in the study area are not formally managed using scientific management regimes and are considered communal grasslands (Figure 2.3). These grasslands are managed using indigenous knowledge systems by the traditional authorities. The grasslands are utilised by locals as grazing land for livestock, particularly cattle, sheep and goats. The locals use this as a means of subsistence as well as income generation.

2.2.2 Sentinel-2 MSI satellite imagery

Sentinel-2 is a multispectral imaging sensor operated by the European Space Agency and provides open, freely accessible data. Cloud free Sentinel-2 data covering the study area was acquired from Land Viewer (https://eos.com/products/landviewer/) on the 21st of October 2021. The image downloaded was a Sentinel-2B product which is an orthorectified and atmospherically corrected product. The Sen2Cor algorithm is used within the Sentinel Application Platform environment (SNAP) to perform atmospheric correction and provided the bottom of atmosphere reflectance data (Main-Knorn et al., 2017). The image was captured on 22nd of June 2021 and aligns with the field data collection days. The Sentinel-2 mission acquires 12-bit images with a swath width of 290km and has a temporal resolution of 5-19 days at spatial resolutions of 10, 20 and 60m. The ortho-images have a UTM/WGS84 projection.

The Sentinel-2 MSI consists of 13 spectral bands that covers the visible, NIR and SWIR sections. The three bands with a 60m spatial resolution were excluded from the analysis as they are primarily used for atmospheric monitoring (Shoko et al., 2018). Sentinel-2 spectral bands are summarised in Table 2.1:

Table 2.1: Sentinel-2B spectral bands (https://eos.com/find-satellite/sentinel-2/)

Band Number	Band Name	Central Wavelength (nm)	Bandwidth (nm)	Resolution (m)
1	Coastal aerosol	442.3	20	60
2	Blue	492.1	65	10
3	Green	559	35	10
4	Red	665	30	10
5	Red edge 1	703.8	15	20
6	Red edge 2	739.1	15	20
7	Red edge 3	779.7	20	20
8	NIR	833	115	10
8a	Red edge 8a	864	20	20
9	Water vapour	943.2	20	60
10	SWIR- Cirrus	1376.9	30	60
11	SWIR 1	1610.4	90	20
12	SWIR 2	2185.7	180	20

2.2.3 Field data collection and measurements

Grass biomass samples were collected in the study area between the 21st of June 2021 and 23rd of June 2021. A total of 120 10m x 10m quadrats, spaced approximately 100m apart, were established within Vulindlela using a purposive sampling technique, as conducted by Royimani et al. (2022). A GPS reading was recorded using the Trimble GPS within each plot, which is a highly accurate sub-metre GPS system. Within each plot, two 1m x 1m sub-plots were established, grass clippings were taken, with the dry biomass mass being averaged within each plot (Ma et al., 2019). Grass clippings were cut approximately 5cm from the ground and only tufts within the sub-plot were taken. Only grasses were sampled, other vegetation such as forbs and sedges were discarded. Grass samples were placed into labelled brown paper bags and a calibrated digital scale was used to measure the fresh biomass weight on the day of collection, which is known as wet mass. Grass samples were then placed into an oven at 70°C for 48 hours to remove moisture. The samples were then reweighed after drying to determine dry mass.

a.  b. 

Figure 2.2.: a. *Aristida junciformis* dominated grassland and b. image of a grassland within the study area during the dry season.

2.2.4 Vegetation Indices

Vegetation indices (VIs) were computed using the spectral bands to assess aboveground biomass (Table 2.2). The VIs used in this study were computed using ArcGIS 10.4 software (www.esri.com).

Table 2.2: Various vegetation indices (VIs) used in this study.

Vegetation Index	Abbreviation	Formula	Reference
	Broadband VIs		
Enhanced Vegetation Index	EVI	$2.5(\frac{NIR - R}{1 + NIR + 6R - 7.5 \times B})$	(Huete et al., 2002)
Soil adjusted vegetation index	SAVI	$\frac{(NIR - R) \times (1 + L)}{(NIR + R + L)}$	(Huete, 1988)
Normalised difference vegetation index	NDVI	$\frac{(NIR - R)}{(NIR + R)}$	(Huete, 1988)
Renormalised difference vegetation index	RDVI	$\frac{(NIR - R)}{Sqrt\ (NIR + R)}$	(Roujean and Breon, 1995)
Simple ratio	SR	$\frac{NIR}{R}$	(Chen, 1996)
Modified simple ratio	MSR	$\frac{(NIR \div R - 1)}{Sqrt\ (NIR \div R) + 1}$	(Chen, 1996)
Green normalised difference vegetation index	GNDVI	$\frac{(NIR - Green)}{(NIR + Green)}$	(Fernández-Manso et al., 2016)
Green-blue normalised difference vegetation index	GBNDVI	$\frac{NIR - (G + B)}{NIR + (G + B)}$	(Santoso et al., 2011)
Chlorophyll green index	CGM	$\frac{NIR}{G} - 1$	(Gitelson and Merzlyak, 1997)
Red-green ratio	RGR	$\frac{Red}{Green}$	(Gamon and Surfus, 1999)
Atmospherically resistance vegetation index	ARVI	$\frac{(NIR - Red)}{(NIR + Blue)}$	(Kaufman and Tanre, 1996)
Transformed difference vegetation index	TDVI	$\sqrt{0.5 + \frac{(NIR - Red)}{(NIR + Red)}}$	(Bannari et al., 2002)
Difference vegetation index	DVI	$NIR - Red$	(Tucker, 1979)
	Red edge VIs		
Red edge 1 NDVI	NDVIRE1	$\frac{(NIR - RE1)}{(NIR + RE1)}$	
Red edge 2 NDVI	NDVIRE2	$\frac{(NIR - RE2)}{(NIR + RE2)}$	
Red edge 3 NDVI	NDVIRE3	$\frac{(NIR - RE3)}{(NIR + RE3)}$	
Red edge 8a NDVI	NDVIRE8a	$\frac{(NIR - RE8a)}{(NIR + RE8a)}$	

Red edge 1 SR	SRRE1	$\frac{NIR}{RE1}$	(Shoko et al., 2018)
Red edge 2 SR	SRRE2	$\frac{NIR}{RE2}$	
Red edge 3 SR	SRRE3	$\frac{NIR}{RE3}$	
Red edge 8a SR	SRRE8a	$\frac{NIR}{RE8a}$	
Normalised difference red edge 1	NDRE1	$\frac{RE1 - Red}{RE1 + Red}$	(Guerini Filho et al., 2020)
Normalised difference red edge 2	NDRE2	$\frac{RE2 - Red}{RE2 + Red}$	
Normalised difference red edge 3	NDRE3	$\frac{RE3 - Red}{RE3 + Red}$	
Normalised difference red edge 8a	NDRE8a	$\frac{RE8a - Red}{RE8a + Red}$	
Anthocyanin reflectance index	ARI	$\frac{1}{Green} - \frac{1}{RE1}$	(Kobayashi et al., 2020)
Red edge chlorophyll index	RECl	$\frac{RE3}{RE1} - 1$	(Clevers and Gitelson, 2013)
Green chlorophyll index	GCl	$\frac{RE3}{Green} - 1$	(Clevers and Gitelson, 2013)
Plant senescence reflective index	PSRI	$\frac{Red - Blue}{RE1}$	(Guerini Filho et al., 2020)
Browning reflective index	BRI	$\frac{\frac{1}{Green} - \frac{1}{RE1}}{NIR}$	(Kobayashi et al., 2020)

2.2.5 Statistical analysis and machine learning

2.2.5.1 Artificial Neural Network (ANN)

This study utilised an artificial neural network (ANN) to predict aboveground biomass using a Sentinel-2 multispectral dataset. The ANN is a machine learning algorithm based on the human brain's computational mechanisms (Mas and Flores, 2008). ANNs can be trained to recognise patterns, perform complex computations and develop self-organising abilities (Mas and Flores, 2008). ANNs are typically comprised of multiple layers (Figure 2.3): an input, output and one or more hidden layers (Yang et al., 2018). A more significant number of layers is associated with greater complexity of the model (Yang et al., 2018). In terms of remote sensing applications, ANNs have been utilised extensively and have proven to provide more reliable results than conventional statistical methods (Mas and Flores, 2008). Regarding aboveground biomass studies, Deb et al. (2017) and Yang et al. (2018) used ANNs to predict aboveground grass biomass.

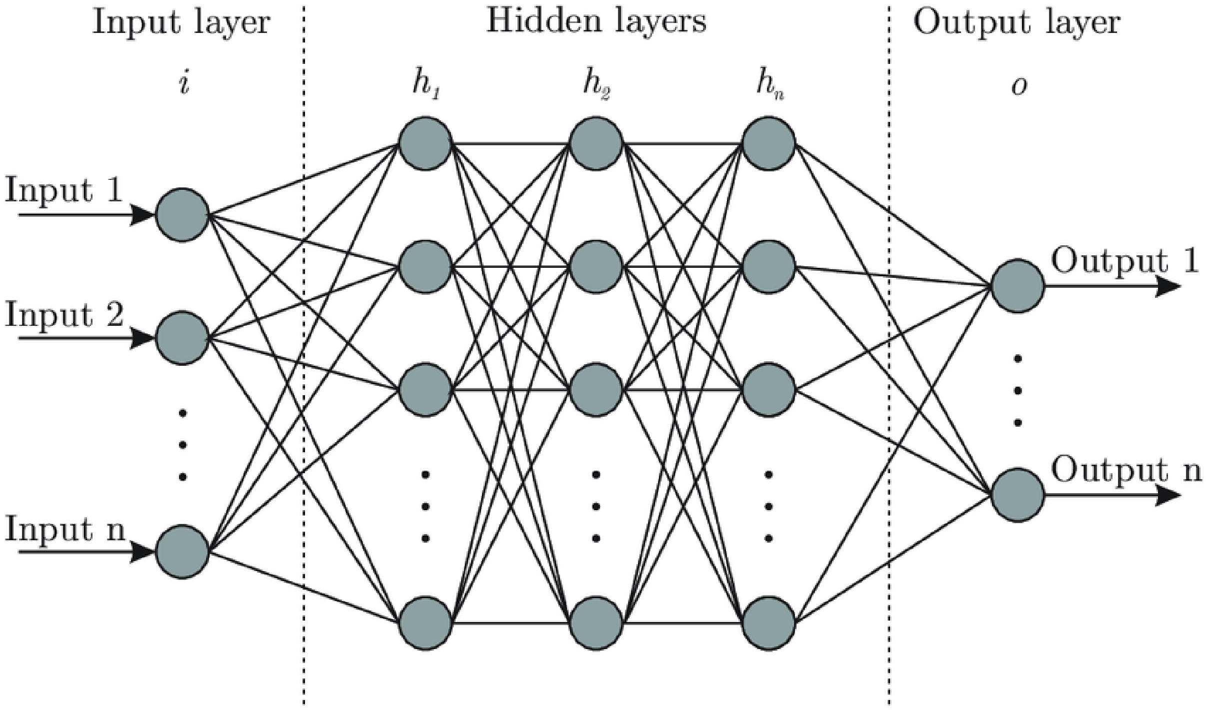


Figure 2.3: The general architecture of an ANN (Jensen et al., 2009)

2.2.5.2 Convolutional Neural Network (CNN)

CNNs are an advancement to typical ANNs and have been developed explicitly for analysing visual imagery (Pires de Lima and Marfurt, 2020). CNNs have increasingly become a valuable and powerful tool in the remote sensing field, especially with image classification (Pires de Lima and Marfurt, 2020). Unlike ANNs that use weights or neurons to “learn” the data, CNNs use multiple layers cast on images to analyse them (Figure 2.4) (Pires de Lima and Marfurt, 2020). ANNs are more suited to concrete datasets, whereas CNNs are more suited for visual datasets. CNNs also provide a more automated approach to deep learning as it can detect important patterns and features in images with minimal human supervision (Ma et al., 2019). For example, Ma et al. (2019) successfully utilised a deep CNN to estimate aboveground biomass for wheat.

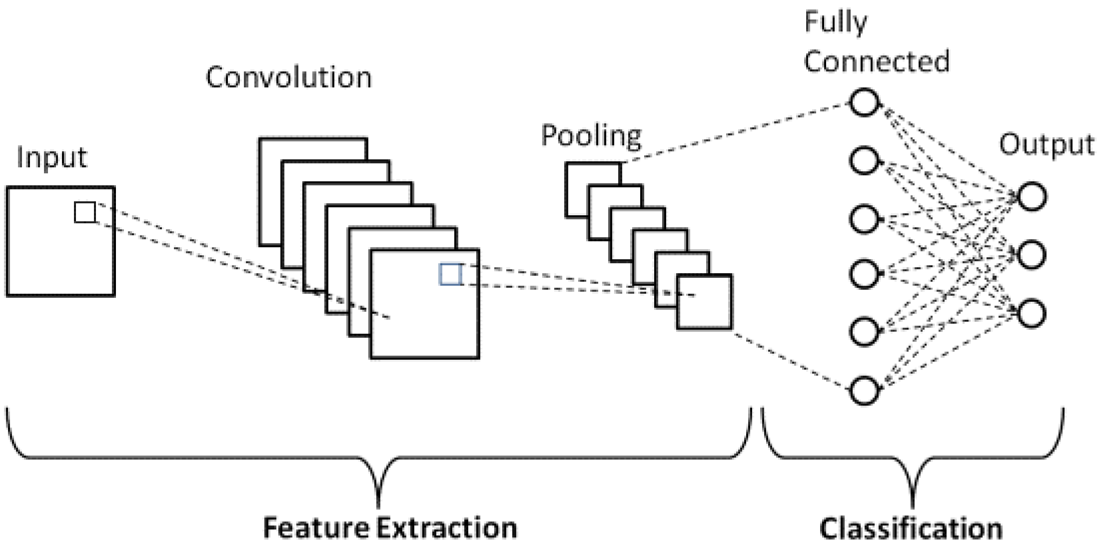


Figure 2.4: The general architecture of a CNN (Liu et al., 2019)

The ANN and CNN models were run to determine the relationship between VIs and spectral data with aboveground biomass. The model successfully determined the relationship between the variables by manually changing the number of nodes in the hidden layer. Table 2.3 lists the parameters used to train the different models.

Table 2.3: Hyper-parameters used to train the ANN and CNN models

Model	**Hyper-parameters**	**Value**
ANN	Number of hidden layers	4
	Number of epochs	50
	Learning rate	0.001
	Activation Function	Sigmoid
CNN	Kernel number	32, 64,128, 256, 512
	Size	1*2
	Stride	2
	Number of epochs	30
	Learning rate	0.001
	Activation Function	ReLu

2.2.6 Accuracy Assessment

The models were run a maximum of five times with random initial weights. Model performance was analysed using the coefficient of determination (R^2), root mean square error (RMSE) and RMSE% assessments. The coefficient of determination is a statistical measure of the accuracy of a regression by comparing actual versus predicted data points (Schreiber et al., 2022). The value of R^2 ranges from 0 to 1 with a higher value insinuating a higher accuracy (Schreiber et al., 2022). The equation for R^2 is found below (Li et al., 2021):

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(y_j - y)^2}{\sum_{j=1}^{n}(y_j - Y)^2} \quad (1)$$

Where y_j and y represents measured and estimated biomass values, respectively; Y is the average measured biomass over all samples and n denotes the number of samples (Li et al., 2021).

According to Shoko et al. (2018), the RMSE measures the difference between actual and predicted values, in this instance, actual and predicted biomass values. The RMSE was calculated using the following formula as documented by Shoko et al. (2018):

$$RMSE = \sqrt{\frac{\sum_{i-1}^{n}(measured\ value - predicted\ value)^2}{n}} \quad (2)$$

where *measured value* is the measured biomass in the field, *predicted value* is the predicted biomass by the model and *i* is the predictor variable included. The RMSE% was calculated using the following formula as expressed by Shoko et al. (2018):

$$RMSE\% = \frac{\sqrt{\frac{1}{n}}\sum_{1=n}^{n}(y_i - Y_i)^2}{y} \quad (3)$$

Where n is the number of measured values, y_i is the measured value, Y_i is the estimated value and y is the average of the measured aboveground biomass (Shoko et al., 2018).

Models yielding the highest R^2 and lowest RMSE/RMSE% between predicted and measured biomass levels, based on an independent test dataset (i.e., 30% of the dataset) were retained for predicting biomass levels. An aboveground biomass distribution map was computed using the ANN and CNN models with spectral data and VIs. A sensitivity analysis was also conducted to determine which variables were most important in model development for the ANN and

CNN. All statistical analyses were conducted utilising R statistical software package version 3.1.3. The methodology used in this study is illustrated in Figure 2.5.

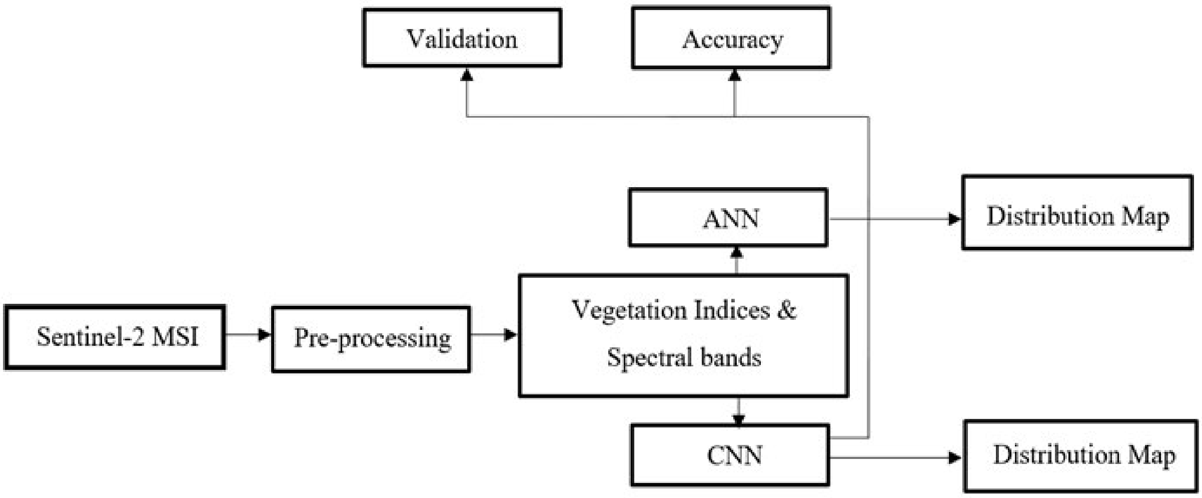


Figure 2.5: Flowchart of the methodology undertaken in this study.

2.3 Results

2.3.1 Descriptive Statistics

Table 2.4: Descriptive statistics of the observed biomass (g/m^2) over the dry season

Period	***n***	**Mean**	**Std. Dev**	**Min.**	**Max.**	**Range**
Dry	120	47.82	23.38	8.2	123.8	115.6

Observed grass biomass (g/m^2) during the dry season across 120 sample plots had an average of 47,82 g/m^2 with a standard deviation of 23,38 g/m^2. The highest biomass recorded was 123,8 g/m^2 whereas 8,2 g/m^2 was the lowest (Table 2.4).

2.3.2 ANN vs CNN

Figure 2.6 show both models' training and validation process with their set hyperparameters. The x-axis represents the number of epochs and the y-axis represents the root mean square error in biomass. An epoch is essentially one cycle of the forward- and back-propagation phases. The CNN model was more adept at learning than the ANN, as the CNN required 30 epochs to minimise error whereas the ANN required 50 epochs to minimise error. The error remained more or less constant after the 30th epoch in the CNN and the 50th epoch in the ANN. Determining a suitable number of epochs is essential to preventing under- or overfitting of models (Ali et al., 2015, Ali et al., 2016)

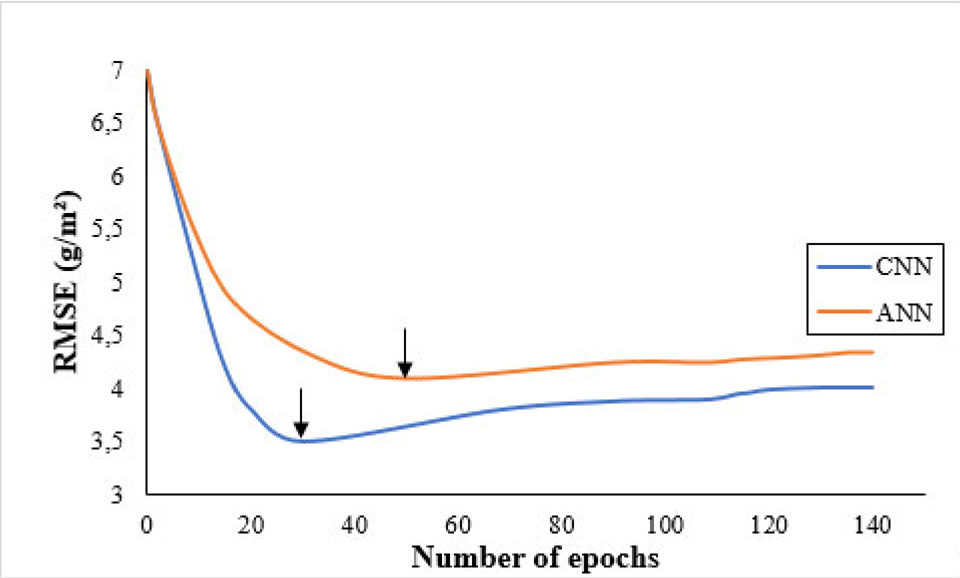


Figure 2.6: Number of epochs for each model. The arrows indicate that for the CNN and ANN models the number of epochs that gave the lowest error was 30 and 50, respectively

In assessing the predictive performance of both the ANN and CNN machine learning algorithms in estimating aboveground biomass during the dry season, the ANN produced an R^2 value of 0,75 with an RMSE of 5,78 g/m^2 and a RMSE% of 8,90 (Figure 2.7a). In comparison, the CNN produced an R^2 value of 0,83 with an RMSE of 3,36 g/m^2 and a RMSE% of 6,09 (Figure 2.7b).

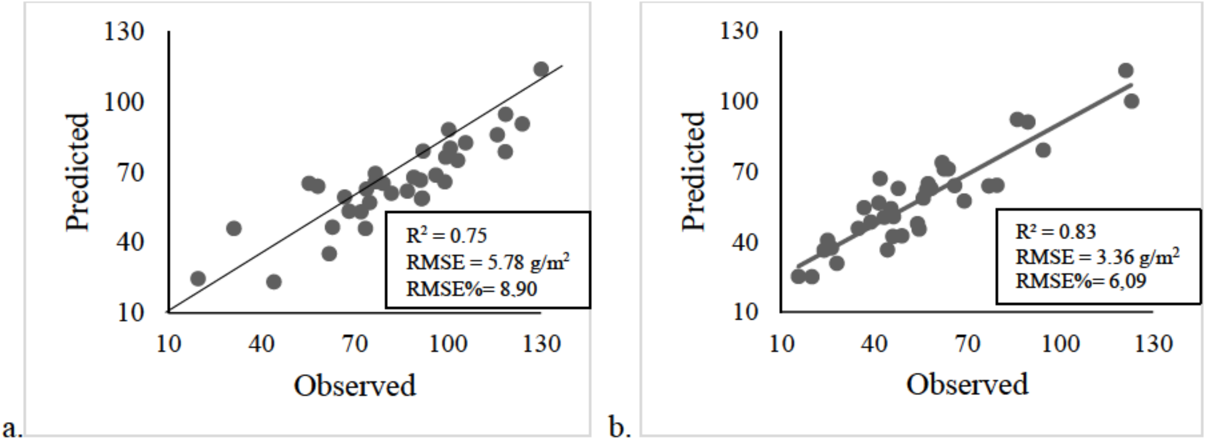


Figure 2.7: Scatterplots showing biomass over the dry season for a. ANN and b. CNN

Figure 2.8 illustrates the spatial distribution of aboveground grass biomass in the study area during the dry season, as predicted by both the ANN and CNN. It can be observed that the predictive map conjured by the CNN has slightly more accurate aboveground biomass representation as compared to the ANN, especially in the peripheral areas of the study site.

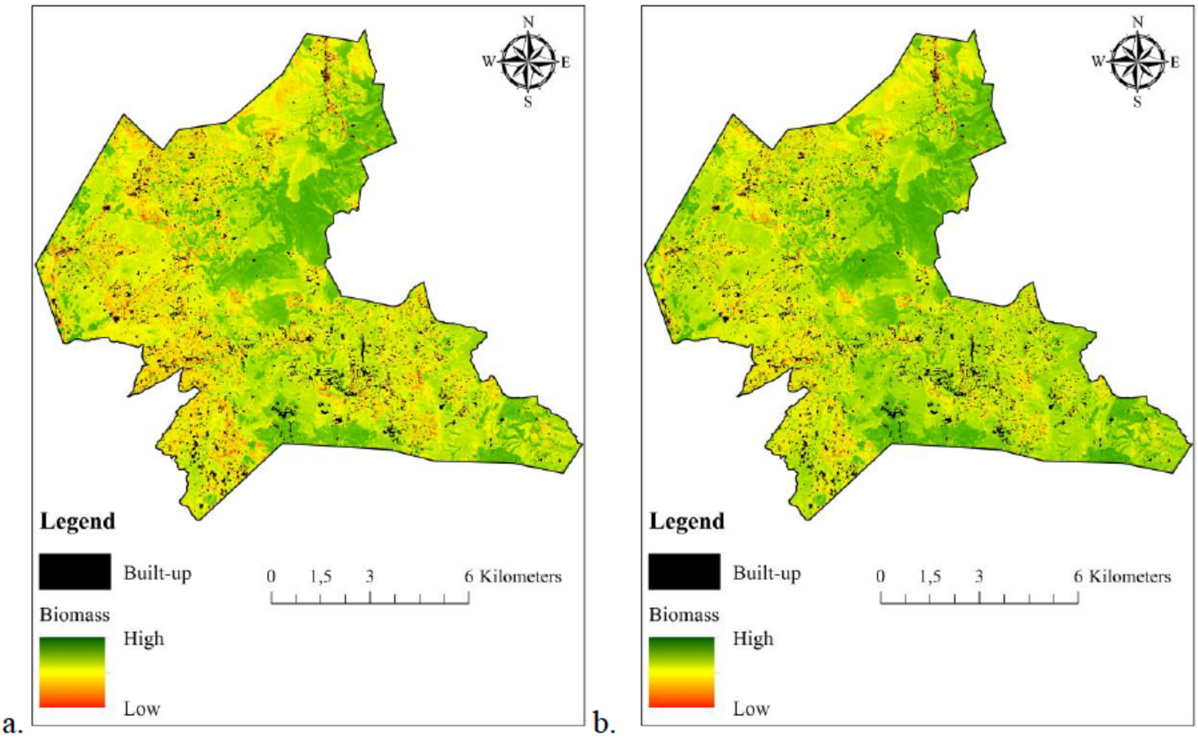


Figure 2.8: Biomass (g/m^2) over the dry season for a. ANN and b. CNN

2.3.3 Sensitivity Analysis

In terms of a sensitivity analysis (Figure 2.9), whereby the importance of spectral bands and VIs are determined and ranked in relation to dry season biomass estimation for each model, the blue band (B02) from Sentinel-2 MSI was the most important band for the ANN, followed by the GCI and the GNDVI. In comparison, the GNDVI was the most important variable for estimating biomass for the CNN, followed closely by the GCI and the blue spectral band (B02) from Sentinel-2. The GBNDVI was the least significant variable in biomass estimation for both models. Only variables with an average impact of >0,1 were included in the models.

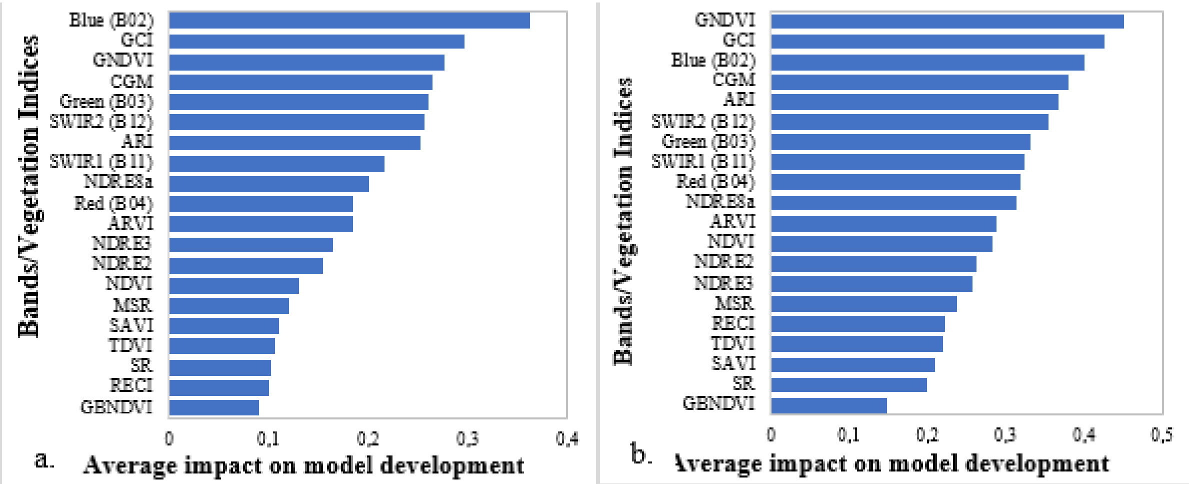


Figure 2.9: Ranking the importance of variables for developing the a. ANN and b. CNN models for biomass detection

2.4 Discussion

This study investigated the utilisation of two neural networks in predicting aboveground grass biomass and compared their respective performance. The advancement in machine learning has provided scientists with numerous opportunities to test their performance in real-world applications, such as in remote sensing and vegetation monitoring (Ali et al., 2015). Comparing the accuracies of two different neural networks helps reveal the relationships between biomass and remote sensing variables (Dong et al., 2020). To date, machine learning algorithms have proven to be much more complex and dynamic than traditional statistical modelling, allowing for more complex modelling of biophysical parameters and more decisive findings and correlations (Das et al., 2022). This study specifically demonstrated the refinement in neural networks, as the contemporary convolutional neural network (R^2=0,83) outperformed the conventional artificial neural network (R^2=0,75) in aboveground grass biomass predictions. This indicates the onset of deep learning approaches in remote sensing applications (Zhu et al., 2017).

Numerous recent studies have conducted a comparative analysis between different machine learning algorithms in remote sensing applications, particularly for vegetation monitoring. The ANN is one of the oldest machine learning algorithms and used extensively for grassland biomass retrieval (Ali et al., 2015). A study by Xie et al. (2009) compared the performance of ANN to a multiple linear regression (MLR) in estimating aboveground biomass of grasslands in Mongolia. The study used Landsat ETM+ (NDVI, bands 1,3,4,5,7) data and the results showed that the ANN (R^2= 0.817, NRMSE= 42,36%) outperformed the multiple linear regression (R^2= 0,591, NRMSE= 53,2%). Similarly, Yang et al. (2018) found that the ANN (R^2= 0,75-0,85) outperforms MLR (R^2=0,4-0,64) in estimating grass biomass. Their study utilised the normalised difference vegetative index (NDVI), enhanced vegetation index (EVI), modified soil adjusted vegetation index (MSAVI), soil adjusted vegetation index (SAVI) and optimised soil adjusted vegetation index (OSAVI) derived from MODIS data. Xie et al. (2009) utilised a single date image whereas Yang et al. (2018) utilised a multi-temporal time series. This proved that machine learning techniques are an improvement to typical regression analyses, even at different spatio-temporal scales (Ali et al., 2015).

Masenyama et al. (2022) has stated that the average R^2 value for remote sensing-grassland productivity studies range between 0,65 (65%) and 0,75 (75%). In comparison, the performance of both the ANN and CNN in this study are commendable with model accuracy of 75% for the ANN and 83% for the CNN. Furthermore, studies by Dong et al. (2020) and Schreiber et al. (2022) found that CNNs outperformed ANNs in aboveground biomass estimation from remotely sensed data. Dong et al. (2020) compared the performance of CNNs against three other machine learning algorithms, namely random forest, support vector regression and ANN, in estimating aboveground biomass of bamboo. The WorldView-2 platform was used in this study, with spectral bands and vegetation indices as input data. Overall, the CNN produced better results than the ANN with an R^2 of 0,94 and RMSE of 23,1% whereas the ANN could only achieve an R^2 of 0,86 and RMSE of 36,1%. The random forest and support vector regression obtained slightly better accuracy than the CNN. However, it must be noted that the CNN had limited input variables in this study, with only spectral bands being used as input data compared to the other two algorithms that had spectral, VI and texture data (Dong et al., 2020).

Similarly, Schreiber et al. (2022) compared the performance of ANNs and CNNs in predicting aboveground biomass of wheat using UAV-based imagery (RGB imagery with 2,14 cm^2 pixel size). Their findings show that the CNN reached an R^2 of 0,9065 whereas the ANN reached an R^2 of 0,9056. In this case, the ANN was slightly outperformed by the CNN. However, Schreiber et al. (2022) also acknowledge that the homogeneity of wheat cultivation could be a slight advantage to the ANN. In contrast, more heterogenous study environments could see the accuracy of ANNs diminish and the accuracy of CNNs flourish. Furthermore, they also note that the use of hyperspectral data and vegetation indices could greatly improve the accuracy of CNNs, which was absent in their study.

Kattenborn et al. (2021) stated that since CNNs have been specialised for image analysis and processing, they are highly suitable for remote sensing applications. CNNs have proven to be extremely useful in extracting biophysical parameters of vegetation from remotely sensed data, such as species composition and biomass (Kattenborn et al., 2021). Deep learning approaches, which include CNNs, are gradually replacing shallow learning techniques such as ANNs, as they analyse, interpret and predict spatial data much more effectively (Zhu et al., 2017, Pires de Lima and Marfurt, 2020, Kattenborn et al., 2021). There has been a steady influx of biomass estimation studies utilising CNNs and remotely sensed data in the academic and research circles.

Ma et al. (2019) utilised a deep CNN in tandem with very high-resolution RGB digital imagery to estimate aboveground biomass of wheat. The CNN had a high coefficient of determination (R^2= 0,808) and a low NRMSE (NRMSE= 24,95%) in predicting wheat biomass (Ma et al., 2019). Karila et al. (2022) utilised a drone with RGB and hyperspectral capabilities to estimate grass sward quality and quantity. They compared the performance of multiple deep neural networks, a CNN included, to the random forest method. Overall, their findings show that the CNN model (NRMSE= 21%) performed better than the random forest model in estimating aboveground grass biomass, with the CNN being the most consistent with hyperspectral data as compared to only RGB data. Varela et al. (2022) predicted various key traits, one of them being aboveground biomass, of Miscanthus grass using UAV imagery and two CNNs. The best R^2 achieved by the 2D CNN, which was multispectral input from a single image, was 0,59 with an RMSE of 180 g. In contrast, the 3D CNN, which was multispectral and multi-temporal, produced a slightly higher R^2 of 0,69 and an RMSE of 149 g.

There have been numerous biomass estimation studies for grasslands using remote sensing and machine learning in a southern African context. However, none of these attempted to utilise CNNs for biomass prediction (Masenyama et al., 2022). For example, Ramoelo and Cho (2014) estimated dry season aboveground grass biomass using the random forest algorithm and by comparing Landsat 8 and RapidEye data. They only utilised band reflectance data to estimate biomass, stating that VIs are not always plausible for biomass estimation during the dry season since the vegetation lacks “greenness” (Ramoelo and Cho, 2014). RapidEye yielded better results with random forest, with an R^2 of 0,86, RMSE of 13,42 g/m^2 and RRMSE of 10,61% whereas Landsat 8 yielded an R^2 of 0,81, RMSE of 15,79 g/m^2 and RRMSE of 12,49%.

Shoko et al. (2018) utilised the sparse partial least square regression (SPLSR) to estimate grass biomass using three different satellites, namely: Sentinel-2 MSI, Landsat 8 OLI and WorldView-2 in the Drakensberg. They utilised seven spectral bands from Landsat 8, ten from Sentinel-2, eight from WorldView 2 and various VIs. Their findings showed that WorldView 2 derived variables yielded the best predictive accuracies (R^2 between 0,71 and 0,83; RMSE between 6,92% and 9,84%), followed by Sentinel-2 (R^2 between 0,6 and 0,79; RMSE between 7,66% and 14,66%) and lastly Landsat 8 (R^2 between 0,52 and 0,71; RMSE between 9,07% and 19,88%). Vundla et al. (2020) assessed aboveground biomass of grasslands in the Eastern Cape using Sentinel-2 MSI and the partial least squares regression (PLSR) algorithm. They utilised the visible, red-edge and shortwave infrared bands and NDVI and simple ratio (SR) as input data for the PLSR. Their results show that the PLSR performed well in estimating grass biomass, with an R^2 of 0,83 and an RMSE of 19,11 g/m^2.

The sensitivity analysis for both models were conducted to determine which spectral bands and VIs were most important in estimating aboveground grass biomass. This is discerned by examining the correlation between aboveground biomass and spectral/VI values (Li et al., 2021). For both models, the vegetation indices and spectral bands proved to be relatively accurate proxies for estimating aboveground grass biomass, a finding that also concurs with Pang et al. (2020). This contrasts the suggestions of Ramoelo and Cho (2014) that vegetation indices may not be suitable for dry season grass biomass estimation due to grass senescence, with only spectral data yielding better results during the dry season. For both the ANN and CNN models, five bands (blue, green, red, SWIR2, SWIR1), nine VIs (GNDVI, CGM, ARVI, NDVI, MSR, SAVI, TDVI, SR, GBNDVI) and six red edge VIs (GCI, ARI, NDRE8a, NDRE3, NDRE2, RECI) were considered important for model development. Other studies also showed that utilising both spectral data and VIs improved on biomass predictions as opposed to using them independently (Shoko et al., 2018, Yang et al., 2018, Pang et al., 2020, Li et al., 2021).

By comparing the results of the sensitivity analysis in this study to other similar biomass studies, it is evident that direct comparisons cannot be established due to the diversity in platform and machine learning algorithms used. Shoko et al. (2018), using Sentinel-2 MSI and SPLSR, found that the SWIR1, SWIR2, green, red and red edge 1 were the most important sensor variables, whereas NDRE1 and NDVI were the most important VIs in predicting grass biomass. Parallels can be drawn between the spectral bands for this study and Shoko et al. (2018); however, this study provides substantially more VIs of significance. Vundla et al. (2020), using Sentinel-2 MSI and PLSR, discovered that simple ratio VIs had the highest importance whilst NDVI had the lowest importance in assessing grass biomass. Their findings on the simple ratio VIs contradict the findings in this study, with the simple ratio VI being of less importance in both the ANN and CNN models. However, NDVI was shown to have a reduced significance in both studies.

NDVI, a widely used VI in remote sensing, was shown to have a moderate impact on biomass estimation in this study for both models. This concurs with Deb et al. (2017) who also found that other VIs produced better biomass estimates than NDVI when paired with neural networks. Ramoelo and Cho (2014) suggest that NDVI is susceptible to grass senescence during the dry season and hence will tend to underestimate biomass. This is due to NDVI essentially measuring vegetation “greenness”, which is primarily absent from grasses during the dry season (Ramoelo and Cho, 2014). Deb et al. (2017) stated that NDVI is often subjected to variations in atmospheric conditions, soil elements, plant phenology and external disturbances which hinder its efficacy in estimating biomass. The findings in this study also show a lesser impact of NDVI on biomass estimates than other VIs.

No known studies to date have utilised Sentinel-2 data with CNNs to assess aboveground biomass. Hence, comparing sensitivity analysis results for the CNN is not plausible. Findings in this study come closest to findings by Li et al. (2021) who predicted aboveground grass biomass using Sentinel-2 MSI in tandem with RF and XGBoost algorithms. They found that the GNDVI and GCI were the most important variables for developing the RF model. GNDVI and GCI were also high impacting variables in this study for both the ANN and CNN models.

According to Dusseux et al. (2022), GNDVI has been well documented and used concerning vegetation biomass. GCI, and other red edge based VIs, have also proven to be particularly useful in biomass studies due to their important relationship with chlorophyll content and nutrients present in plant cells (Dusseux et al., 2022, Vundla et al., 2020). Ramoelo and Cho (2014) found that the blue, green and red edge spectral data were important for predicting aboveground grass biomass during the dry season, albeit using RapidEye and RF. This study also shows the importance of blue and green spectra; however, shortwave infrared was deemed more significant in this study than red edge spectra.

Compared to the abovementioned studies, the performance of both NNs in this study is relatively commendable. However, it must be acknowledged that all machine learning algorithms used in remote sensing studies are not comprehensive tools for classifying or predicting biophysical attributes (Dong et al., 2020). They all have their own benefits, strengths and limitations based on numerous factors such as sensor type, spatial resolution, temporal resolution and spectral resolution (Das et al., 2022). Dong et al. (2020) state that both CNN and ANN are highly sensitive to the architecture and parameter settings; hence, these aspects must be geared appropriately to avoid poor model performance. Model performance can either be too low, whereby the model's predictive ability is poor, or too high, whereby the model begins to overfit the data (Kattenborn et al., 2021). Neural networks are typically known for their tendency to overfit data; hence, preventative solutions must be implemented during data processing to mitigate this (Ali et al., 2015).

Furthermore, both types of neural networks used in this study require high computational power and are time-consuming (Dong et al., 2020). Other machine learning techniques such as random forest are much more compatible with smaller sample size or input data than neural networks, and these factors must be accommodated (Ali et al., 2015). It must also be noted that CNNs perform better with multi-temporal spatial data (3D CNNs) when compared to single date imagery (2D CNNs) (Varela et al., 2022). The utilisation of CNNs in remote sensing applications is still gaining momentum. Hence, there is need for future research to optimise the algorithm for vegetation remote sensing (Kattenborn et al., 2021). Much of the focus of CNNs in vegetation remote sensing has been on object identification and classification. However, semantic segmentation applications, such as biomass and LAI predictions, must be explored further (Kattenborn et al., 2021). It is unlikely that CNNs will replace ANNs altogether in the remote sensing field as they both provide advantages and disadvantages, and the practicality of each is case-specific (Ali et al., 2015). However, CNNs have great potential for grassland biomass studies in future.

2.5 Conclusion

This study compared two neural networks' performance in estimating grass's aboveground biomass, using Sentinel-2 space-borne spectral data and derived vegetation indices. Findings in this study suggest that the deep learning neural network, CNN, outperforms the traditional ANN. However, both algorithms performed satisfactorily in predicting grass biomass. This study can be considered pilot scale research, particularly in a southern African context, as no known research has attempted to compare the performance of two different neural networks in grassland monitoring. Although each algorithm has pros and cons, with large training datasets and computational time being a common disadvantage, this pioneering research establishes great potential for the utilisation of CNNs in remote sensing research in the future. Future research can improve upon this research by incorporating larger training datasets, utilising multi-temporal and higher resolution data to enhance the performance of CNNs in biophysical remote sensing studies. The primary objective of this study was to determine which neural network would better predict grass biomass using open-access and freely available satellite data. The CNN model developed in this study can be considered effective for accurate estimation of biomass in grassland monitoring, and is evident in the advancement in applied deep learning.

2.6 Link to next chapter

The chapter above showcased the ability of the CNN to detect and map aboveground biomass more effectively than the ANN over the dry season. Therefore, the next chapter will utilise the CNN to compare aboveground biomass over the wet and dry season.

CHAPTER THREE

PREDICTING INTER-SEASONAL GRASS BIOMASS UTILISING SATELLITE REMOTE SENSING IN THE VULINDLELA AREA OF THE UMGENI CATCHMENT, KWAZULU-NATAL

Abstract

The maintenance of optimal grassland productivity is an imperative objective of many rangeland managers. This is especially for subsistence farmers who depend on communal grasslands for their livelihoods. However, grassland productivity fluctuates between dry and wet seasons, primarily due to changes in temperature and precipitation. Biomass is often used as a proxy for grassland productivity and is also influenced by seasonal changes. This study aimed to distinguish changes in grass biomass from the dry season to the wet season using Sentinel-2 MSI data and deep Convolutional Neural Networks (CNNs). Remote sensing in tandem with machine learning has proven to be cheap and rapid for vegetation monitoring over traditional field measurements. The advancement in deep learning, particularly, has prompted novel research for vegetation monitoring applications. In this study, Sentinel-2 MSI spectral and VI data were used as a proxy for developing a 2-D CNN model to predict inter-seasonal aboveground grass biomass. A sensitivity analysis was also conducted to discern which input variables were the most influential for model development. Overall, the CNN performed satisfactory in estimating aboveground biomass during both dry and seasons. The model produced an R^2 of 0,83, an RMSE of 3,36 g/m^2 and an RMSE% of 6,09 in the dry season and an R^2 of 0,85, an RMSE of 2,41 g/m^2 and an RMSE% of 3,71 in the wet season. The blue band, GCl and GNDVI proved to be the most important input variables for model development in terms of the sensitivity analysis. The findings in this study also suggested that grass biomass was substantially influenced by changes in rainfall and temperature. This study exhibits the great potential in utilising deep learning for grassland monitoring.

Keywords: Remote Sensing, Grasslands, Biomass, Convolutional Neural Network, Sentinel-2, Dry Season, Wet Season.

3.1 Introduction

Grasslands cover approximately 26% of the total land area with most grasslands situated in tropical and sub-tropical developing countries around the world (Boval and Dixon, 2012). Grasslands are defined by Mucina and Rutherford (2006) as grass dominated biomes, with the majority of grasses being C4 plants in low to mid-altitudes and C3 plants being more prominent in higher altitudes. Woody species are controlled by frost, fire or grazing in grasslands allowing grasses to dominate (Mucina and Rutherford, 2006). Grasslands are environmentally, socially and economically valuable biomes worldwide as they serve as water catchments, biodiversity reserves, carbon sinks, recreational areas and agricultural practices (Boval and Dixon, 2012). The grassland biome is a major biome in South Africa and exists mainly in the eastern parts of the country from altitudes of near sea level to around 2800m above sea level (Mucina and Rutherford, 2006).

In both commercial and rural contexts, grasslands are used extensively in South Africa as a fodder for livestock (Richardson et al., 2010), where grasslands are commonly termed rangelands (Richardson et al., 2010). South Africa consists primarily two types of grasslands based on environmental factors such as precipitation and altitude: the sourveld and sweetveld grasslands (O'Connor et al., 2011). The sourveld grasslands have higher fibre content and tend to withdraw nutrients from the leaves during winter or dry periods whereas the sweetveld grasses have lower fibre content and maintain a consistent conglomerate of nutrients during winter or dry periods (Ellery et al., 1995). A significant portion of South African grasslands are intensively managed for optimal foraging efficiency in livestock production ensuring that livestock remains in peak condition to maintain productivity (O'Connor et al., 2011).

The livestock and wildlife sectors in South Africa are heavily dependent on the grassland biome for maintaining productivity and functionality (Palmer et al., 2010). Livestock production is essential for meeting the demand for high quality meat and dairy in South Africa (O'Connor et al., 2011). However, livestock is not only economically important, but also socially important, as many rural communities depend on livestock for a sustainable livelihood (Palmer et al., 2010, Richardson et al., 2010). This is particularly so in a South African context whereby indigenous people depend heavily on livestock for food, income, social status and overall wellbeing (Rasch et al., 2016). Grasslands are a cheap source of stock feed for many rural communities in which natural disasters and socio-economic challenges are prevalent (Sibanda et al., 2017). Hence, rangelands must be managed efficiently to ensure that livestock production remains viable and beneficial.

For grazers like livestock, biomass is the primary indicator of the quantity of fodder available to livestock for consumption (Ramoelo and Cho, 2014). Biomass varies throughout the year based on seasonal fluctuations in precipitation and temperature, namely the wet season in the summer months and the dry season in the winter months (Ramoelo and Cho, 2014). Livestock are often limited by fodder availability during the dry season. However, in the KZN sourveld grasslands, livestock may be limited by quantity and quality during the dry season (Ramoelo et al., 2015). Rust and Rust (2013) emphasise that climate change poses a significant threat to rangelands and livestock production as increased variability in climatic conditions provides rangeland managers with a challenge to predict future fodder availability. Other threats to grassland productivity include infrastructural development, crop farming and overgrazing (Sibanda et al., 2017). On the contrary, O'Mara (2012) suggests that grasslands may play a significant role in food security and carbon sequestration in the future, despite the many threats. Therefore, predicting fluxes in seasonal fodder biomass is essential to inform planning and management strategies in grasslands (Ramoelo and Cho, 2014).

The technological advancements in remote sensing enable scientists to successfully predict and estimate biomass in both natural and agricultural contexts (Ramoelo and Cho, 2014). Remote sensing allows managers to monitor the quantity and quality of fodder throughout the year to inform decision making and maintain rangeland productivity (Ramoelo et al., 2015). Ramoelo et al. (2015) state that conventional methods of predicting biomass are time- and energy-intensive. Mutanga et al. (2016) have documented how remote sensing has been applied to vegetation monitoring in South Africa with readily available and easily accessible satellite data. Remote sensing, also termed as earth observation, has been used extensively to facilitate biomass monitoring at various spatio-temporal scales with satisfactory results of accuracy and precision (Sibanda et al., 2017). However, earth observation is a complex process and it often produces varying degrees of success based on the different methodologies used as well as the diverse biophysical and environmental traits in vegetation (Sibanda et al., 2017).

The number of biomass estimation studies using remote sensing data is ever-growing. However, such studies are still limited in data-scarce countries such as southern Africa (Sibanda et al., 2017). Most historical biomass studies have focused on forests and used Landsat data due to its availability (Samimi and Kraus, 2004, Sibanda et al., 2017). However, Landsat data has its limitations with restricted spatial and radiometric capabilities (Shoko et al., 2018). Recent studies have gradually progressed to more high-resolution data, such as Sentinel-2 (Shoko et al., 2018), WorldView-2 (Shoko et al., 2018) and WorldView-3 (Sibanda et al., 2017). These satellites provide higher spatial and spectral resolution with faster revisit times, allowing for much more refined biomass monitoring and estimation studies (Shoko et al., 2018).

Furthermore, the complexity and multi-dimensionality of remote sensing data have proved to be a challenge in data processing and analysis, especially when using traditional statistical methods (Ali et al., 2015). Scientists have successfully applied machine learning to remote sensing studies over the years for classification, object identification or prediction of biophysical variables (Mas and Flores, 2008). However, the ongoing improvement and refinement of machine learning techniques resulted in numerous algorithms that can be applied to remote sensing (Ali et al., 2015). Currently, there is a growing shift towards deep learning approaches, which have the potential to yield much more accurate results in remote sensing studies (Zhu et al., 2017). One such example of a deep learning technique is convolutional neural networks (CNNs) which are specifically geared for imagery (Kattenborn et al., 2021). The utility of CNNs for vegetation biomass has been investigated before by Ma et al. (2019), Dong et al. (2020) and Varela et al. (2022). However, the practicality of CNNs for assessing aboveground grass biomass has yet to be determined.

The advancement in multispectral scanners such as the introduction of Sentinel-2 provides great opportunities to build on and improve biomass studies in southern Africa (Sibanda et al., 2017). There is a lack of studies with regards to inter-seasonal changes in grasslands (Masenyama et al., 2022), particularly at larger spatial scales in both South Africa (Dingaan and Tsubo, 2019) and the rest of Africa (Hunter et al., 2020). In this regard, this study aimed at fulfilling this research gap in literature and assist rangeland managers and rural communities in making more informed decisions. This study aimed to predict inter-seasonal (dry season and wet season) fodder quantity in the Vulindlela area of the Umgeni catchment using high resolution satellite imagery. The objectives of this study were:

- To compare aboveground grass biomass from wet and dry seasons and relate potential changes to changes in biophysical conditions.
- Utilise Sentinel-2 bands and derived vegetation indices in tandem with deep learning CNN to estimate aboveground biomass inter-seasonally.

- Explore potential measures that can be used to improve grassland management.

3.2 Methods

3.2.1 Study Area

Vulindlela is located within the Umgungundlovu district and is considered a part of the greater Umgeni river catchment in the KwaZulu-Natal province. Sibanda et al. (2021) describe the local climate as sub-tropical, typical of cool, dry winters and mild, wet summers. The area receives a mean annual rainfall of approximately 980mm and a median annual rainfall of around 850mm. Recorded annual maximum and minimum temperatures are 22°C and 10°C respectively (Sibanda et al., 2021). Soil factors within the area are shallow with moderate to poor drainage (Sibanda et al., 2021). Climatic factors such as temperature and precipitation are the main driving factors for vegetation in this area (Sibanda et al., 2021).

Fynn et al. (2011) state that grasslands within the study area are categorised as mesic grasslands and are usually dominated by a few species, depending on grassland condition. The grasslands were initially dominated by *Themeda triandra* grass (Royimani et al., 2022). However, due to anthropogenic transformation, the grasslands are now characterised by species such as *Aristida junciformis*, *Panicum maximum* and *Paspalum urvillei*, amongst others (Royimani et al., 2022). The study area experiences the dry season in June/July (Masemola et al., 2020) whereas the wet season usually stretches from October to March (Roffe et al., 2020). Local communities utilise the communal grasslands as rangelands for their livestock and cultural purposes. Livestock is a significant source of income for the locals; hence, rangeland productivity affects them directly.

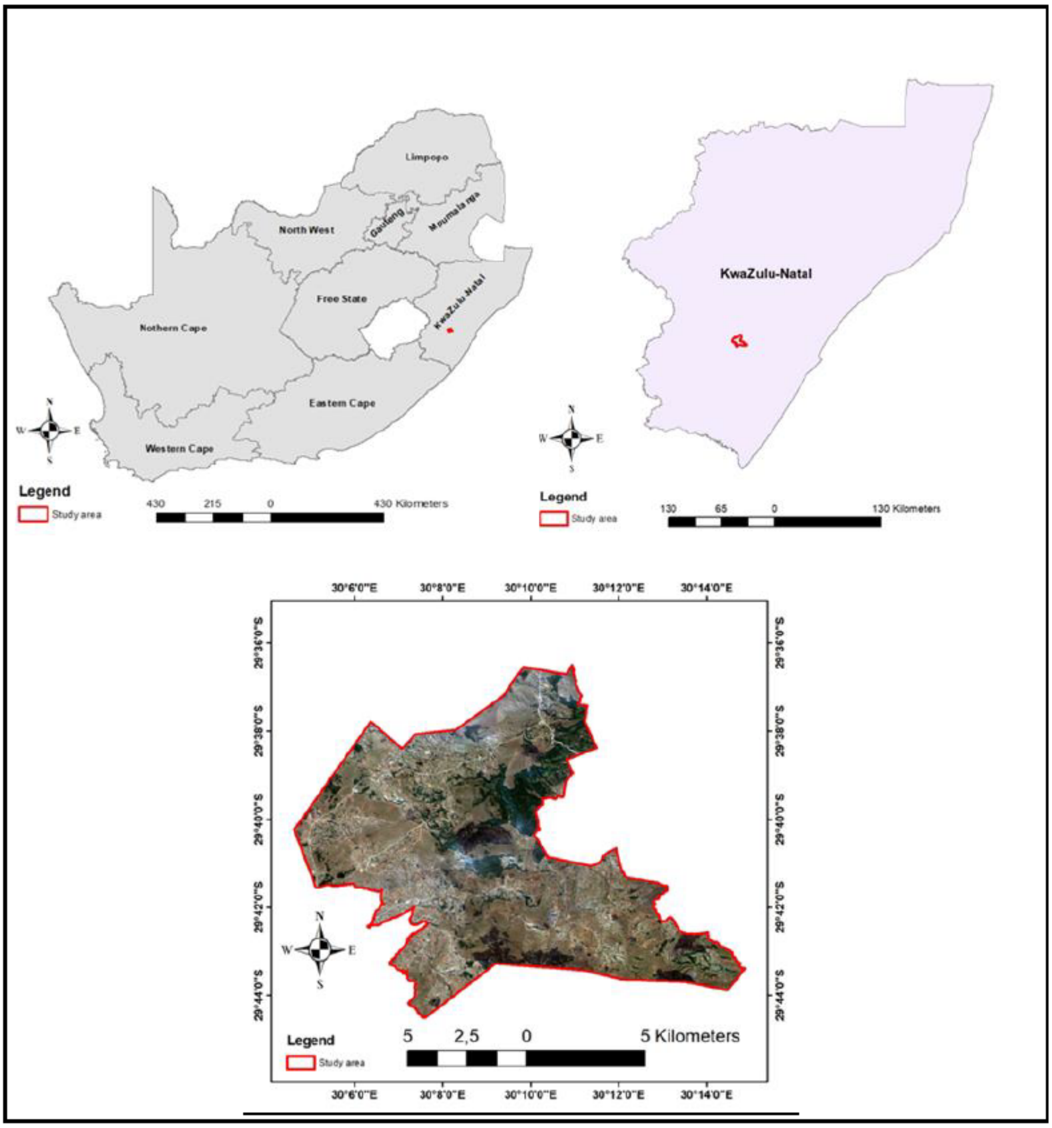


Figure 3.1: Location of Vulindlela in KwaZulu-Natal Province, South Africa

3.2.2 Sentinel-2 MSI satellite imagery

Two Sentinel-2 Multi-Spectral Instrument (MSI) seasonal scenes were freely acquired from Land Viewer (https://eos.com/products/landviewer/). The images were downloaded on the 21st of October 2021 and 14th of April 2022 respectively. Both images were downloaded as cloud-free Sentinel-2B products which are orthorectified and atmospherically corrected products pre-processed in the Sentinel Application Platform (SNAP) using the Sen2Cor algorithm (Main-Knorn et al., 2017). The dry season image was captured on the 22nd of June 2021 whereas the wet season image was captured on the 29th of March 2022. The image acquisition dates- therefore, align with the field data collection periods. Sentinel-2 MSI acquires 12-bit images with a swath width of 290km, a revisit time of 5-19 days and at spatial resolutions ranging from 10 to 60m. Sentinel-2 MSI has been highly recommended for grassland monitoring mainly due to extensive coverage, high spatial and temporal resolutions and the ability to capture data in the Red Edge section of the electromagnetic spectrum (Royimani et al., 2022).

3.2.3 Field data collection and measurements

Dry season data collection was conducted between the 21st of June 2021 and 23rd of June 2021. Wet season data collection was conducted between the 28th of March 2022 and 1st of April 2022. The sampling strategy remained uniform between the two seasons. In each data collection period, a total of 120 plots of 10m x 10m in size at a distance of 100m apart were established using the purposive sampling technique (Royimani et al., 2022). Within each plot, a GPS reading was recorded using a Trimble GPS which outputs co-ordinates at a sub-metre level. Within each plot, two sub-plots of 1m x 1m were sampled for aboveground biomass, with the mean dry biomass being recorded for each plot (Ma et al., 2019). Ma et al. (2019) state that 1m x 1m quadrats are suitable for heterogenous grasslands, such as natural grasslands. Grass clippings were initially weighed using a calibrated scale and wet mass was recorded. The samples were then oven-dried for 48 hours at 70°C and were thereafter reweighed to determine dry mass.

a.  b. 

Figure 3.2: Images of the study area in the wet season where a. has not been grazed and b. has been grazed by livestock (March 2022)

a. b.

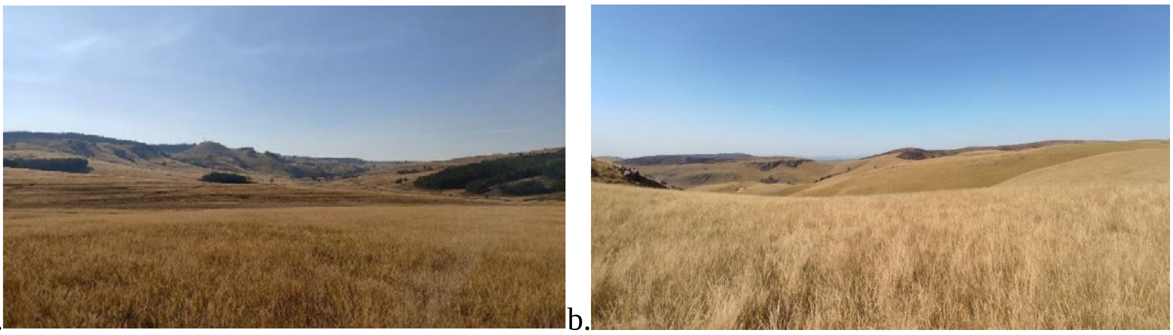


Figure 3.3: Images of the study area in the dry season where a. has been grazed and b. has not been grazed (July 2021)

3.2.4 Sentinel-2 spectral bands and variables

Sentinel-2 MSI provides spectral data in 13 bands that range from visible to shortwave infrared. Blue, green, red and NIR have a spatial resolution of 10m, whereas Red Edge and SWIR bands have a spatial resolution of 20m. Coastal aerosol, SWIR Cirrus and water vapour have a spatial resolution of 60m. However, this study excluded these as they are mainly used for atmospheric monitoring (Shoko et al., 2018).

Numerous vegetation indices (VIs) were computed from the Sentinel-2 spectral data. All VIs were calculated in ArcGIS 10.4 (www.esri.com). A detailed description of all VIs and their associated formulas are given in Table 3.1 below.

Table 3.1: Vegetation indices used in this study derived from Sentinel-2 spectral data

Vegetation Index	**Abbreviation**	**Formula**	**Reference**
	Broadband VIs		
Enhanced Vegetation Index	EVI	$2.5(\frac{NIR - R}{1 + NIR + 6R - 7.5 \times B})$	(Huete et al., 2002)
Soil adjusted vegetation index	SAVI	$\frac{(NIR - R) \times (1 + L)}{(NIR + R + L)}$	(Huete, 1988)
Normalised difference vegetation index	NDVI	$\frac{(NIR - R)}{(NIR + R)}$	(Huete, 1988)
Renormalised difference vegetation index	RDVI	$\frac{(NIR - R)}{Sqrt\ (NIR + R)}$	(Roujean and Breon, 1995)
Simple ratio	SR	$\frac{NIR}{R}$	(Chen, 1996)
Modified simple ratio	MSR	$\frac{(NIR \div R - 1)}{Sqrt\ (NIR \div R) + 1}$	(Chen, 1996)
Green normalised difference vegetation index	GNDVI	$\frac{(NIR - Green)}{(NIR + Green)}$	(Fernández-Manso et al., 2016)
Green-blue normalised difference vegetation index	GBNDVI	$\frac{NIR - (G + B)}{NIR + (G + B)}$	(Santoso et al., 2011)
Chlorophyll green index	CGM	$\frac{NIR}{G} - 1$	(Gitelson and Merzlyak, 1997)
Red-green ratio	RGR	$\frac{Red}{Green}$	(Gamon and Surfus, 1999)
Atmospherically resistance vegetation index	ARVI	$\frac{(NIR - Red)}{(NIR + Blue)}$	(Kaufman and Tanre, 1996)
Transformed difference vegetation index	TDVI	$\sqrt{0.5 + \frac{(NIR - Red)}{(NIR + Red)}}$	(Bannari et al., 2002)
Difference vegetation index	DVI	$NIR - Red$	(Tucker, 1979)
	Red edge VIs		
Red edge 1 NDVI	NDVIRE1	$\frac{(NIR - RE1)}{(NIR + RE1)}$	(Shoko et al., 2018)
Red edge 2 NDVI	NDVIRE2	$\frac{(NIR - RE2)}{(NIR + RE2)}$	
Red edge 3 NDVI	NDVIRE3	$\frac{(NIR - RE3)}{(NIR + RE3)}$	
Red edge 8a NDVI	NDVIRE8a	$\frac{(NIR - RE8a)}{(NIR + RE8a)}$	
Red edge 1 SR	SRRE1	$\frac{NIR}{RE1}$	
Red edge 2 SR	SRRE2	$\frac{NIR}{RE2}$	
Red edge 3 SR	SRRE3	$\frac{NIR}{RE3}$	
Red edge 8a SR	SRRE8a	$\frac{NIR}{RE8a}$	
Normalised difference red edge 1	NDRE1	$\frac{RE1 - Red}{RE1 + Red}$	(Guerini Filho et al., 2020)
Normalised difference red edge 2	NDRE2	$\frac{RE2 - Red}{RE2 + Red}$	
Normalised difference red edge 3	NDRE3	$\frac{RE3 - Red}{RE3 + Red}$	

Normalised difference red edge 8a	NDRE8a	$\frac{RE8a - Red}{RE8a + Red}$	
Anthocyanin reflectance index	ARI	$\frac{1}{Green} - \frac{1}{RE1}$	(Kobayashi et al., 2020)
Red edge chlorophyll index	RECl	$\frac{RE3}{RE1} - 1$	(Clevers and Gitelson, 2013)
Green chlorophyll index	GCl	$\frac{RE3}{Green} - 1$	(Clevers and Gitelson, 2013)
Plant senescence reflective index	PSRI	$\frac{Red - Blue}{RE1}$	(Guerini Filho et al., 2020)
Browning reflective index	BRI	$\frac{\frac{1}{Green} - \frac{1}{RE1}}{NIR}$	(Kobayashi et al., 2020)

3.2.5 Statistical analysis and machine learning

Convolutional Neural Networks are an emerging class of machine learning algorithms that have been used to interpret geospatial information in primarily two ways: object detection and semantic segmentation (Brodrick et al., 2019). Object detection is characterised as identifying key components in an image and semantic segmentation is classifying each pixel individually in an image (Brodrick et al., 2019). CNNs are a subset of deep learning models and are viewed as an advancement to typical ANNs (Brodrick et al., 2019). In the application of remote sensing for vegetation monitoring, input data in spectral indices and texture metrics are the cornerstone of modelling (Kattenborn et al., 2021). However, these predictors are endless and it is not easy to define the most appropriate predictors for vegetation analysis as they are influenced by the biochemical and structural properties of plants and other environmental factors (Kattenborn et al., 2021). With deep learning, the CNN can learn and decipher which input variables are the best for analysis based on learning spatial features present in the data (Kattenborn et al., 2021).

CNNs are made up of neurons that are organised in layers, with three main layers: input, hidden and output layers (Kattenborn et al., 2021). Neurons within the same layer and between different layers are connected by weights and biases (Kattenborn et al., 2021). CNNs contain at least one convolutional layer within the hidden layers. These convolutional layers exploit patterns in the data using filters by convolving, which is the sliding of the filter over the layer and calculating the dot-product of the filter and layer values (Kattenborn et al., 2021). The product of convolving is called a feature map. The feature maps are simplified in a pooling layer which assists in data reduction, simpler model parameters, lower computational load and a reduction in overfitting (Kattenborn et al., 2021). Biomass prediction would be performed using a semantic segmentation variation of a CNN. Encoding layers within the convolutional layer cluster and aggregate information from the entire dataset. Decoding layers follow encoding layers which are responsible for increasing spatial resolution and decreasing convolution depth. In simple terms, this allows the CNN to make pixel-by-pixel predictions at the same spatial resolution as the input data. This would ensure that model predictions and ground truthing data can be compared directly (Brodrick et al., 2019). A typical structure of a CNN, known as CNN architecture, is depicted below in Figure 3.4:

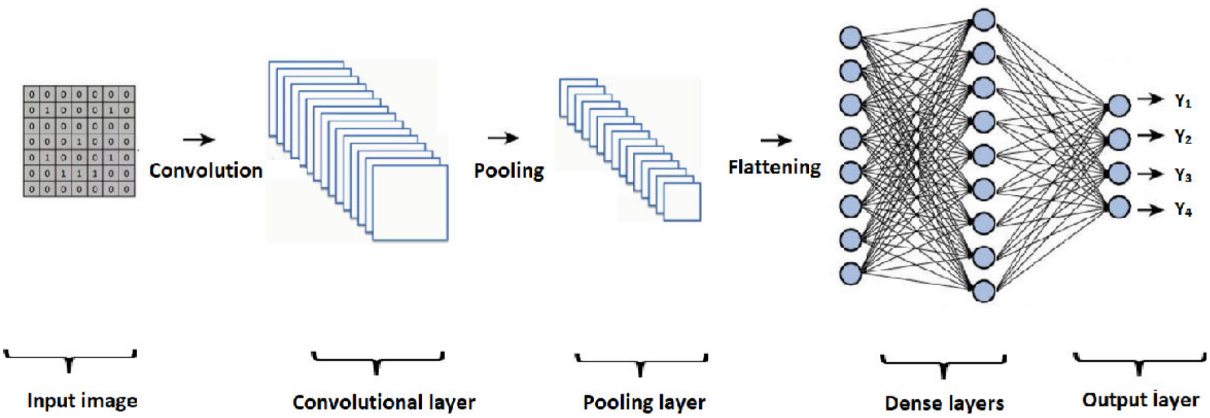


Figure 3.4: A general structure of a CNN (Kattenborn et al., 2021)

Dong et al. (2020) give the formula of convolution as:

$$map_{l,j}^{x,y} = f\left[\sum_m \sum_{h=0}^{Hi-1} \sum_{w=0}^{Wi-1} k_{l,j.m}^{h,w}\, map_{(l-1),m}^{(x+h),(y+w)} + b_{l,j}\right] \quad (4)$$

where $k_{l,j.m}^{h,w}$represents the value at the position (h,w) of the kernel connected to the *m*th feature map in the *(l-1)*th layer, H_i and W_i are the height and width of the kernel, $b_{l,j}$ is the bias of the *j*th feature map in the *l*th layer and *f* is the activation function (Dong et al., 2020). The CNN model was constructed and run using R statistical software version 3.1.3. The hyper-parameters of the CNN in this study can be found in Table 3.2.

Table 3.2: Hyper-parameters used to train the CNN model.

Model	Hyper-parameters	Value
CNN - Dry	Kernel number	32, 64,128, 256, 512
	Size	1*2
	Stride	2
	Number of epochs	30
	Learning rate	0.001
	Activation Function	ReLu
CNN - Wet	Kernel number	32, 64,128, 256, 512
	Size	1*2
	Stride	2
	Number of epochs	30
	Learning rate	0.001
	Activation Function	ReLu

3.2.6 Accuracy Assessment

Accuracy assessments are essential for understanding model performance and determining model practicality. Three standardised error metrics were used to assess model performance: coefficient of determination (R^2), root mean square error (RMSE) and root mean square error percentage (RMSE%). Schreiber et al. (2022) define R^2 as a statistical measure of accuracy by comparing observed versus predicted data points. R^2 values to range from 0 to 1 with a higher value translating into higher model accuracy and vice-versa. The equation for R^2 is found below (Li et al., 2021):

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(y_j - y)^2}{\sum_{j=1}^{n}(y_j - Y)^2} \quad (5)$$

Where y_j represents measured biomass, y is estimated biomass, Y is mean biomass across all samples and n is the sample number (Li et al., 2021).

The RMSE measures the difference between actual and predicted values and is calculated by:

$$RMSE = \sqrt{\frac{\sum_{i-1}^{n}(measured\ value - predicted\ value)^2}{n}} \quad (6)$$

as documented by Shoko et al. (2018). The *measured value* and *predicted value* is the actual biomass in the field and the predicted biomass, respectively.

The RMSE% provides a magnitude of error concerning actual values and can be expressed by the following formula (Shoko et al., 2018):

$$RMSE\% = \frac{\sqrt{\frac{1}{n}\sum_{1=n}^{n}(y_i - Y_i)^2}}{y} \quad (7)$$

Where n is the number of samples, y_i and Y_i are measured and predicted values, respectively; and y is the average of the measured values.

Following the general training/test rule, 70% of the dataset was used to train the CNN model whereas 30% was used to test the model. Models with the highest R^2 and lowest RMSE/RMSE% were retained for predicting aboveground biomass in both seasons. Predictive biomass distribution maps were constructed for both seasons using Sentinel-2 spectral bands and derived VIs as input data. A sensitivity analysis was also run to determine which input variables were the most significant in developing the CNN model. Figure 3.5 illustrates the processes and methodology used in this study.

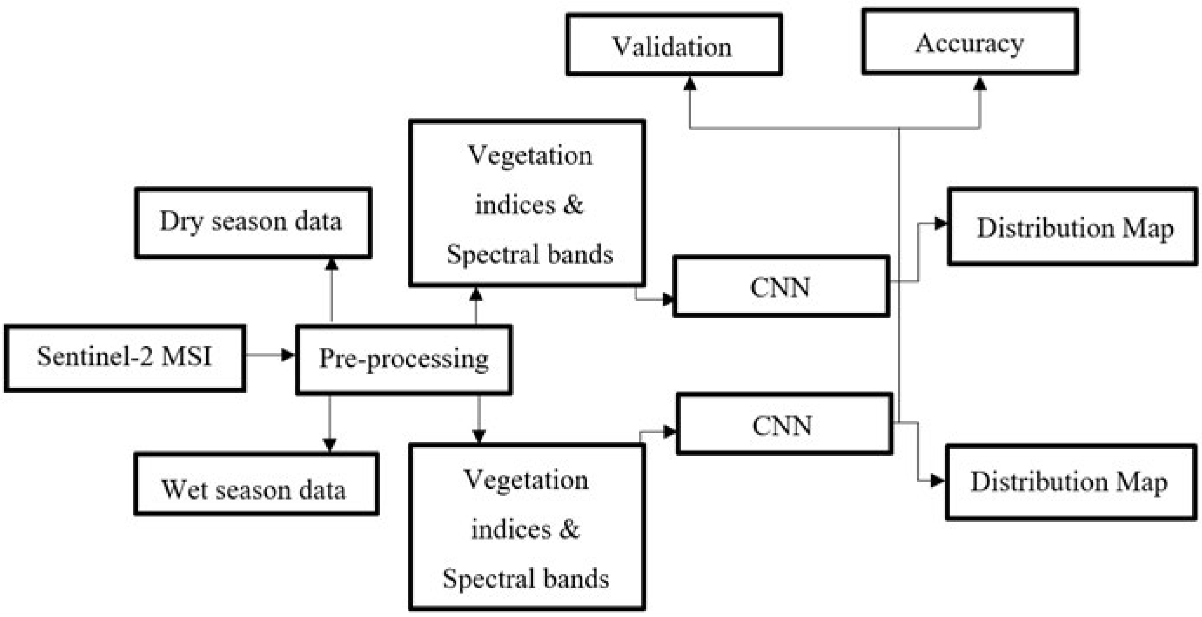


Figure 3.5: Flowchart of the methodology undertaken in this study

3.3 Results

3.3.1 Descriptive Statistics

Table 3.3: Descriptive statistics of the observed biomass (g/m^2) over the wet and dry season

Period	*n*	Mean	Std. Dev	Min.	Max.	Range
Dry	120	47.82	23.38	8.2	123.8	115.6
Wet	120	195.67	72.04	73.8	477.3	403.5

The mean aboveground biomass recorded during the dry season was 47,82 g/m^2 with a standard deviation of 23,38 g/m^2. The range observed during the same period was 115,6 g/m^2 with 123,8 g/m^2 and 8,2 g/m^2 being the highest and lowest biomass values recorded, respectively. Aboveground biomass recorded during the wet season differed substantially, with an average biomass of 195,67 g/m^2 and a standard deviation of 72,04 g/m^2. The wet season had a range of 403,5 g/m^2 with 477,3 g/m^2 and 73,8 g/m^2 being the highest and lowest biomass values recorded, respectively (Table 3.3).

3.3.2 CNN Training History

The CNN models for both seasons were run with a maximum of 140 epochs, however this was stopped after 30 epochs as this was when model performance was optimal (Figure 3.6). An epoch can be defined as one complete cycle of the forward and back-propagation phase. The activation function used was ReLu. The loss function for the models were RMSE.

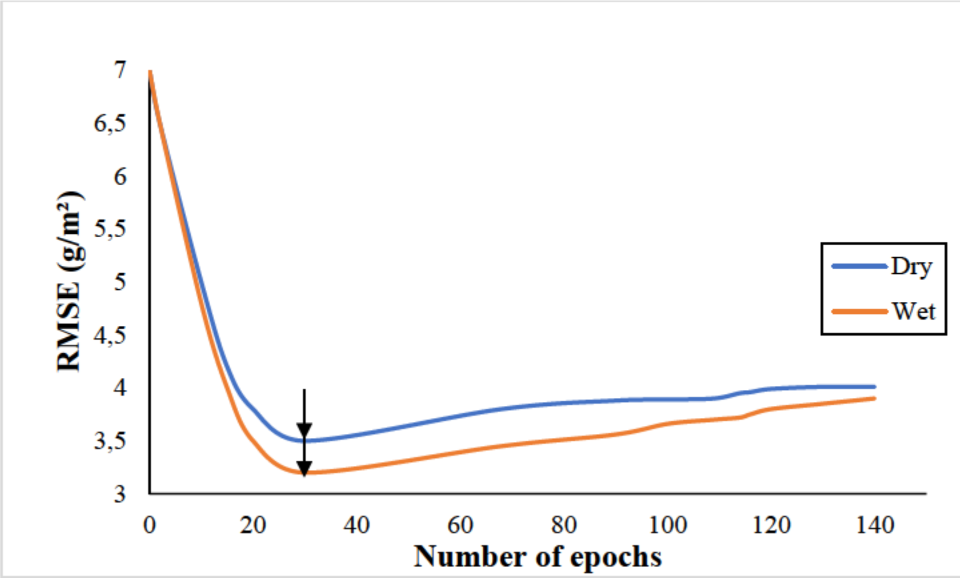
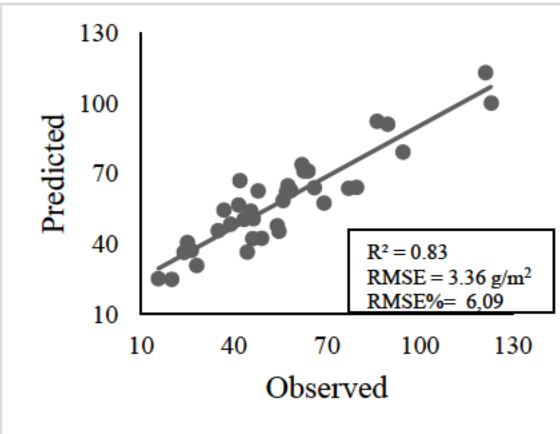


Figure 3.6: Number of epochs for each model. The arrows indicate that for the CNN - Dry and CNN-Wet models the number of epochs that gave the lowest error was both 30.

3.3.3 Dry season vs Wet season

The CNN algorithm used to predict aboveground biomass produced an R^2 of 0,83 with an RMSE of 3,36 g/m^2 and a RMSE% of 6,09 in the dry season (Figure 3.7a). Comparatively, the CNN produced a R^2 of 0,85, RMSE of 2,41 g/m^2 and RMSE% of 3,71 in the wet season (Figure 3.7b).

a. 

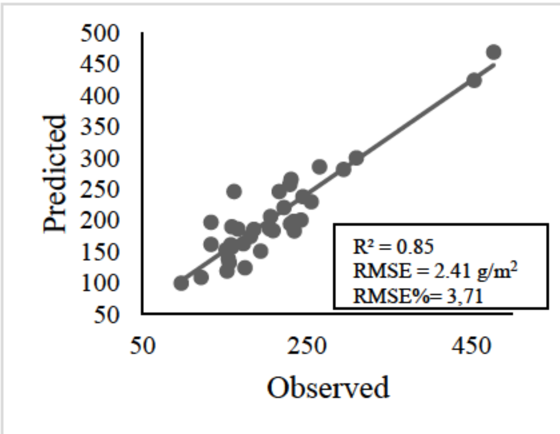
b. 

Figure 3.7: Scatterplots showing observed and predicted biomass over the a. dry season and b. wet season using CNN

Figure 3.8 illustrates the spatial distribution of aboveground grass biomass during the dry and wet seasons. The difference in aboveground grass biomass can be observed between the two seasons with higher biomass indicated in the wet season compared to the dry season. Although higher biomass was predicted in some areas of the study site during the dry season, these areas of high biomass are concentrated in certain parts of the study area. Overall, the wet season depicts higher biomass over a greater spatial scale, with biomass being more evenly distributed across the study area.

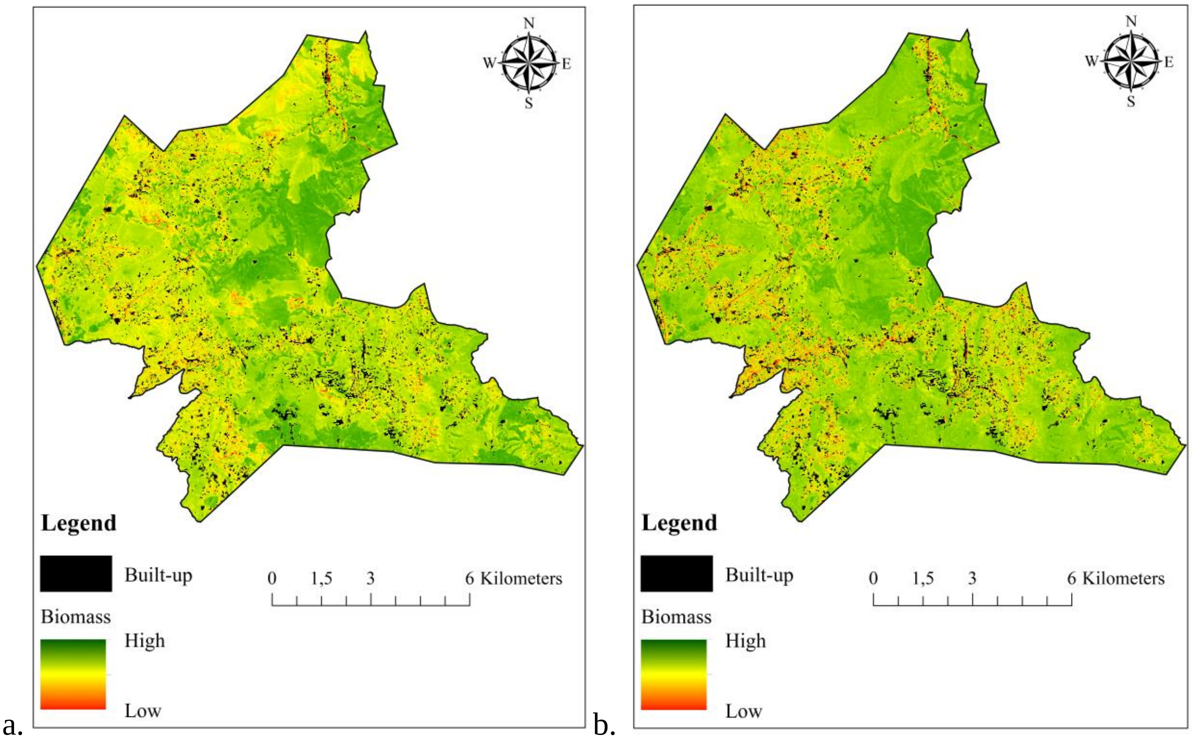


Figure 3.8: Predicted biomass (g/m^2) over the a. dry and b. wet season using CNN

3.3.4 Sensitivity Analysis

Deciphering which input variables are the most significant for model development is critical in ensuring respectable model performance. Figure 3.9 depicts which input variables, from Sentinel 2 spectral bands and derived VIs, were most important for model development in both seasons. It must be noted that only variables with an average impact of >0.1 were included in the model. The top three variables for the dry season included the GNDVI, GCI and the blue band whereas GCI, GNDVI and the blue band were the three most important for the wet season.

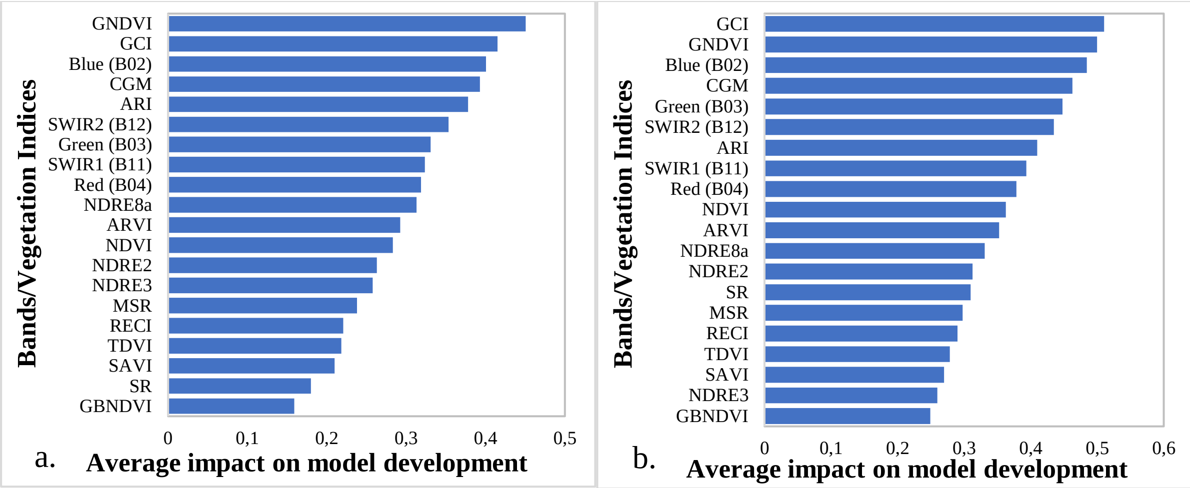


Figure 3.9: Ranking the importance of variables for developing the CNN model for biomass detection in the a. dry season and b. wet season

3.3.5 Changes in Rainfall and Temperature across seasons

Total monthly rainfall in the study area over the dry season decreased from 33,4mm in April 2021 to 2,8mm in July 2021, which was peak winter (Figure 3.10a). Overall, a decreasing trend was evident in total monthly rainfall during the dry season. In contrast, the wet season recorded a highest total monthly rainfall in December 2021 followed closely by January 2022, which was peak summer months (Figure 3.10b). Wet season total monthly rainfall data followed a more normal distribution pattern with a peak in December 2021.

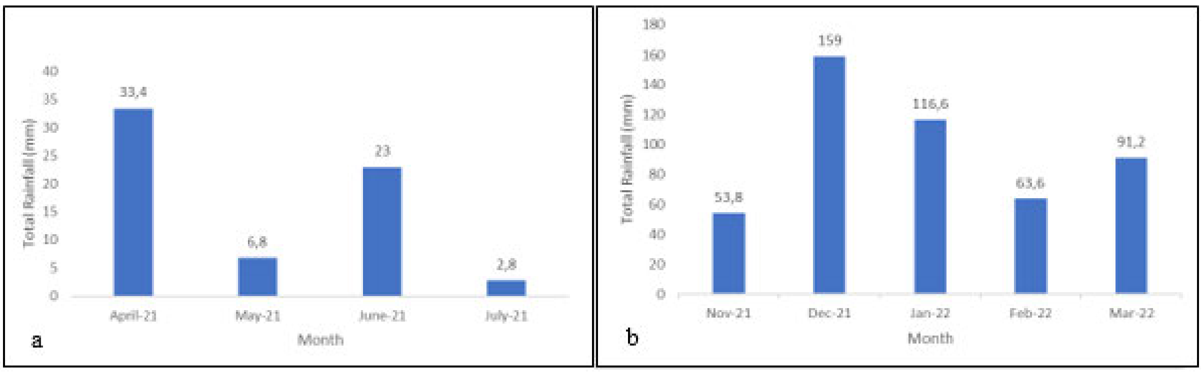


Figure 3.10: Total monthly rainfall (mm) in Vulindela during the a. dry season and b. wet season (Data provided by South African Weather Services)

A gradual decrease was recorded for average maximum daily temperature across the dry season, with a decrease from 24,03C° in April 2021 to 18,13 C° in July 2021 (Figure 3.11a). In contrast, a sharp increase in average maximum daily temperature was recorded at the start of the wet season from November 2021 (24,6 C°) to December 2021 (27,9 C°). Thereafter, the average maximum daily temperature remained fairly uniform for the remainder of the wet season (Figure 3.11b).

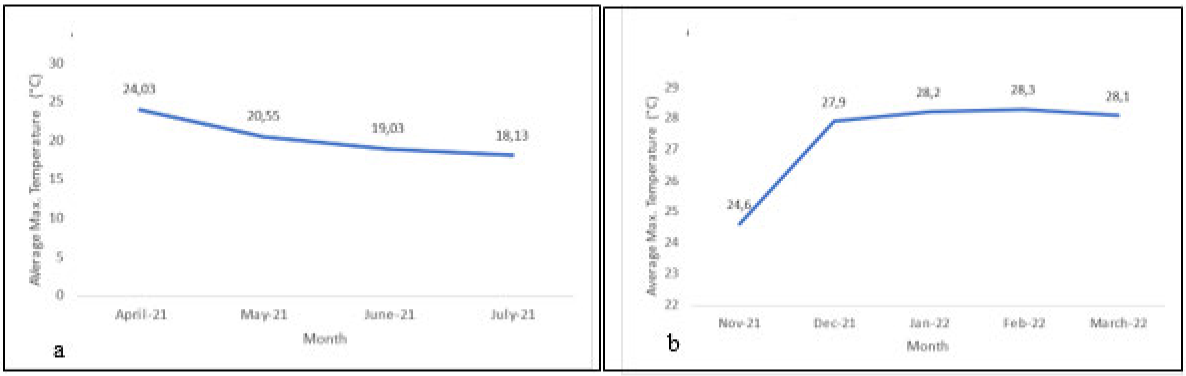


Figure 3.11: Average maximum daily temperature (C°) in Vulindlela over the a. dry season and b. wet season (Data provided by South African Weather Services)

3.4 Discussion

This study estimated and compared aboveground grass biomass between the dry season (April-July) and the wet season (November-March) in the greater Umngeni catchment. Overall, recorded grass biomass increased from ±48 g/m^2 in the dry season to ±196 g/m^2 in the wet season. The predicted biomass maps also depict a significant increase in aboveground biomass across the study area during the wet season. In contrast, biomass is primarily concentrated in small patches across the study area during the dry season.

Grasslands are driven by external factors such as precipitation, temperature and fire (Masenyama et al., 2022). These factors maintain the ecological functionality of the grassland. However, these factors also fluctuate spatio-temporally (Shoko et al., 2018). It has been widely agreed that grassland productivity is directly and significantly related to changes in both rainfall and temperature (Shoko et al., 2018, Dingaan and Tsubo, 2019, Magandana et al., 2020) . Both rainfall and temperature variables fluctuate based on seasonal variations and hence play a significant role in influencing grassland productivity, particularly aboveground biomass (Magandana et al., 2020). Van den Hoof et al. (2018) found a statistically significant relationship between rainfall variability and grassland productivity. Furthermore, Magandana et al. (2020) found statistically significant relationships between changes in rainfall and temperature with changes in aboveground grass biomass.

The findings in this study concur with Van den Hoof et al. (2018) and Magandana et al. (2020), albeit this study originates from a remote sensing background. Average total rainfall from April 2021 to July 2021 had a downward trend with an average total rainfall of approximately 16,5 mm for the dry period (Figure 3.10a). Average rainfall across the wet season, from November 2021 to March 2021, had an overall increasing trend with average total rainfall for the wet period estimated to be 96,84 mm (Figure 3.10b). This indicates an almost six-fold increase in rainfall received in the wet season when compared the dry season. Furthermore, temperature data from the dry and wet seasons follow the same trend, with the average daily maximum temperature decreasing gradually during the dry months and increasing steeply during the wet months (Figure 3.11a and Figure 3.11b). The average daily maximum temperature across the dry and wet period was approximately 20,44 °C and 27,42 °C, respectively. Therefore, the increase in aboveground grass biomass can be linked to rainfall and temperature increase, as also suggested by Van den Hoof et al. (2018) and Magandana et al. (2020).

Although rainfall and temperature are the significant drivers of grassland productivity, the influence of other biophysical factors such as soil and rainfall type cannot be omitted (Van den Hoof et al., 2018). The type of rainfall received is as important as the quantity received over time (Roffe et al., 2019). Gradual rainfall events allow for better water absorption into the soil column than erratic rainfall events, in which most of the rainfall is lost as surface run-off (Van den Hoof et al., 2018, Roffe et al., 2019). The edaphic factors of the grassland also play a significant role in productivity, particularly soil pH, texture and organic matter content (Van den Hoof et al., 2018). An increase in precipitation increases plant production, which in turn increases soil organic matter (Van den Hoof et al., 2018). Soil type is also influential in plant productivity, with fine clay-like soils being more suitable for optimal production than coarse sandy soils (Van den Hoof et al., 2018). This is due to clay-like soils being more adept at nutrient exchange, holding organic content, better bulk density and higher soil organic carbon than sandy soils (Van den Hoof et al., 2018).

The study area consists mainly of two soil types, Acrisols and Ferralsols, as deduced from Fey (2010). Acrisols are brownish-reddish soils with fine granular structure and sandy-loamy texture (Podwojewski et al., 2011). Acrisols are generally unproductive soils that lack sufficient plant nutrients, have a high pH and usually form a substrate for grasslands or savannah (Podwojewski et al., 2011). Acrisols are also highly porous soils and are especially susceptible to soil erosion (Podwojewski et al., 2011). Ferralsols are characterised by reddish-yellow soils with high clay content (Mukangango et al., 2020). Ferralsols are structurally sound soils with good infiltration and drainage. However, they are chemically poor soils with most plant nutrients being stored in the biomass and can only be recycled back into the soil column by moribund (Mukangango et al., 2020). Acrisols and Ferralsols are similar and can often be found together, with both soils being susceptible to dry periods and drought (Mukangango et al., 2020). Since both soils are well drained and poor at water retention, they cannot provide enough moisture for grasses and vegetation, particularly on slopes (Mukangango et al., 2020). The lack of precipitation during the dry season can account for changes in edaphic factors, which inherently affect biomass availability.

Grasslands are naturally maintained by grazing and fire, two non-climatic factors influencing plant productivity (Koerner and Collins, 2014). The grassland in this study is utilised as a communal rangeland by the local community for their livestock (Cho et al., 2021). However, the lack of a formal rangeland management plan has resulted in adverse conditions within the grassland, is mainly due to fire and overgrazing (Cho et al., 2021). Fire is administered by the local community whenever deemed fit, even though it may be contrary to scientific guidelines. This not only affects the ability of grasses to regenerate, but it also affects the soil characteristics (nutrients, moisture content, organic content) which severely reduces productivity (Reinhart et al., 2016).

Furthermore, livestock is allowed to graze freely resulting in uneven forage distribution and soil erosion in some areas. This was evident and observed within the study area during data collection (Figure 3.2 and Figure 3.3). Continuous grazing by livestock hinders grass productivity as the grass does not have the ability to regrow, particularly in the dry season when stored nutrients are scarce (Koerner and Collins, 2014). Grazing factors such as stocking rates are significant in maintaining grassland productivity, as high stocking rates affect grasslands negatively if not conducted in a controlled manner (O’Connor et al., 2011). Cho et al. (2021) state that the local community face challenges with effective rangeland management, which has resulted in a shortfall of forage, especially during the dry season. The need for an effective and collaborative rangeland management plan, with appropriate stocking rates and rotational grazing, is imperative to improve grassland productivity in Vulindlela (Cho et al., 2021).

To the best of the authors' knowledge, this study can be considered a pilot study as it is one of the first studies, to the best of the authors' knowledge, to predict vegetation biomass using deep learning and Sentinel-2 MSI. Remote sensing has been extensively used in biomass studies, with relatively good levels of success (Mutanga et al., 2016). The advent of machine learning has enabled extensive and complex data analysis in remote sensing, often producing more reliable and accurate results as compared to traditional statistical methods (Ali et al., 2015). Machine learning has advanced through time and contemporary deep learning approaches to data analysis appears to be the outlook for the foreseeable future (Zhu et al., 2017). Neural networks are the foundation of deep learning approaches, and the CNN is one of the most promising deep learning algorithms for vegetation remote sensing applications (Kattenborn et al., 2021). Deep learning differs from typical shallow learning mainly by how the algorithm processes data. In typical machine learning, a human has to ensure that structured data is organised and pre-processed in order for learning to take place, also termed as supervised machine learning (Yuan et al., 2020). However, with deep learning, the algorithm can learn and decipher which data components should be used for feature extraction, resulting in less dependency on supervised learning and pre-processed data (Zhu et al., 2017).

The use of deep CNNs for vegetation biomass studies are sparse, however, they are gaining momentum in academia (Yuan et al., 2020). Most of the studies utilising CNNs have an agricultural background, and have used unmanned aerial vehicles (UAVs) data in small-scale spatial contexts. This study utilised a CNN to estimate grass biomass using open-access and readily available satellite data at a larger spatial scale. Karila et al. (2022) state that two broad types of CNNs that can be used for vegetation monitoring; 2D and 3D CNNs. 2D CNNs are simple CNNs that only utilise a single image (mono-temporal) as an input whereas 3D CNNs have multiple images as input data (Karila et al., 2022, Varela et al., 2022). This study used a simple 2D CNN as only single images from respective dry and wet seasons were used.

Karila et al. (2022) used a UAV with an RGB and hyperspectral sensor (1024 x 648-pixel size and 36 bands between 500-900 nm) to estimate grass biomass, amongst other variables. Their 2D CNN model recorded a NRMSE of 21% whereas their 3D CNN yielded a NRMSE of 10%. Karila et al. (2022) only used NRMSE for model accuracy assessments hence R^2 and RMSE values are not included. Similarly, Varela et al. (2022) predicted aboveground biomass of

Miscanthus grass using UAV imagery with RGB, near infrared and red edge bands (1,4 cm spatial resolution) using 2D and 3D CNNs. Their 2D CNN recorded an R^2 of 0,59 and an RMSE of 180g whereas their 3D CNN produced an R^2 of 0,69 and an RMSE of 149g. Alves Oliveira et al. (2022) utilised UAV RGB data and 3D CNNs of different architectures to estimate aboveground grass biomass. Their best model recorded an R^2 of 0,88 and an RMSE of 482,12 kg/ha, with model performance being significantly influenced by the type of architecture. In comparison, the simple CNN model in this study performed well with an R^2= 0,83/RMSE%= 6,09 and an R^2= 0,85/ RMSE%= 3,71 in the dry and wet seasons, respectively. Taking into consideration that Sentinel-2 imagery was used in this study as opposed to high resolution UAV data, this study shows that CNNs have the potential to be used with freely available satellite data and can be used at regional spatial contexts.

Chen et al. (2021) is arguably a study that can be directly compared to the findings in this study. Chen et al. (2021) used Sentinel-2 imagery paired with a deep sequential neural network (SNN), which is a subset of Recurrent Neural Networks (RNN), to estimate pasture biomass. Their study only used the ten applicable spectral bands, used in this study as well, and NDVI. However, they also included climate data in their models which could not be included in the CNN models in this study due to the lack of complete climate datasets for our study area, with data from the SAWS being relatively disjointed and incomplete to be able to be included in model development. According to Lakhal et al. (2018), the main difference between CNNs and RNNs is that the latter is specialised in processing temporal information or information that follows a set sequence. This was apt for Chen et al. (2021) as they utilised time series Sentinel-2 data from 2017 to 2018 to study pasture biomass, albeit at a paddock-level spatial scale. Their SNN model performed adequately with an R^2 of 0,6 and an RMSE of 356 kg/ha. Furthermore, their study also observed that seasonal patterns in aboveground pasture biomass were distinct, with biomass increasing in the wet season and decreasing in the dry season. They also associate this with changes in climatic conditions, with water availability being highly influential to pasture biomass (Chen et al., 2021).

Jin et al. (2020) utilised mono-temporal Sentinel-2 imagery with a deep neural network to estimate maize biomass. Their study used fifteen VIs and leaf area index (LAI) data as input data to predict maize biomass. Their model performed well with the best R^2 of 0,91, RMSE of 1,49 t/ha and RRMSE of 20,05%. In terms of a sensitivity analysis, Jin et al. (2020) found that the three band water index (TBWI), normalised difference infrared index (NDII) and normalised difference moisture index (NDMI) were the most important VIs for biomass estimation. In this study, the most important VIs for model development for both seasons were GNDVI, GCI and CGM. Théau et al. (2021) study found that GNDVI has a high correlation with grass biomass, particularly in grasslands with low vegetation levels of <0,5kg/m^2, which was the case in our study particularly in the dry season. Hamada et al. (2021), using Sentinel-2 for grass biomass predictions, found that CGM, GCI and GNDVI all highly correlated with biomass. Jin et al. (2020), Chen et al. (2021) and Hamada et al. (2021) found that NDVI was moderate to poorly correlated to biomass and hence was not a relatively important variable in model development. Findings in this study concur as NDVI was moderately significant for dry and wet seasons. Théau et al. (2021) and Hamada et al. (2021) found that green and blue spectral bands were more important for biomass predictions than the red band, which was also found in this study.

Many authors agree that using CNN models for biomass estimation is preliminary, novel and pioneering (Ma et al., 2019, Dong et al., 2020, Alves Oliveira et al., 2022). The same sentiment can be iterated in this study as no known studies have attempted to use CNNs and satellite imagery for biomass predictions. CNNs require large amounts of training data to operate accurately, which may prove to be a limitation as large datasets are not always available (Kattenborn et al., 2021). Using CNNs for small datasets has been done before, as recorded by Narayanan et al. (2021); however, they may require some pre-training and transfer learning to ensure that they are optimised for biomass estimation (Narayanan et al., 2021). Furthermore, the architecture and hyperparameters of CNNs are highly influential in model performance and these must be further studied to improve the generalizability of CNNs (Alves Oliveira et al., 2022). This study was also limited to using only a single image as inputs for model training. Studies show that using multi-temporal imagery significantly improves CNN model accuracy (Karila et al., 2022, Varela et al., 2022). Future studies can attempt to improve on model performance using multi-temporal satellite data.

3.5 Conclusion

This study evaluated the change in aboveground biomass from the dry season to the wet season using Sentinel-2 remotely sensed imagery and simple convolutional neural networks. Sentinel-2 MSI bands and derived VIs were used as input proxy data to train the CNN model for both seasons while ground data was used as a benchmark to assess model accuracy. A significant difference between dry and wet season grass biomass was discovered, with the wet season biomass increasing four times of dry season biomass. These changes can be primarily related to significant changes in rainfall and temperature which also bring about influential changes in other biophysical factors such as soil. Overall, the findings in this study concur with previous studies studying seasonal biomass changes.

This study can also be considered a pilot study as it attempted to utilise a deep learning approach to predict grass biomass. Model performance produced promising results, albeit with a simple CNN and a limited dataset. This research could prove useful to farmers and rangeland managers in planning and decision-making as remote sensing allows for fast and accurate estimation of grassland productivity. However, future research can improve the reliability and practicality of CNN modelling by incorporating multi-temporal data and utilising larger datasets. Using more complex and intricate CNN models in future may also improve predictive performance.

CHAPTER FOUR

CONCLUSIONS

4.1 Introduction

The focus of this study was to distinguish the difference in aboveground grass biomass between the dry and wet seasons using remote sensing and neural network algorithms. In this chapter, the research conducted will be evaluated against the aim and objectives of the study, as mentioned in Chapter One. This will examine how close the study has come to meeting the targets that were set out. The findings in this study will then be synthesised and consolidated in the concluding remarks. Limitations in this study will be acknowledged and recommendations for future research proposed.

4.2 Aim and objectives reviewed

4.2.1 Aim

The aim of this study was to predict inter-seasonal aboveground grass biomass using Sentinel-2 MSI and machine learning algorithms in the Umngeni catchment, KwaZulu-Natal.

4.2.2 Objectives reviewed

In this study, two broad objectives were established in order to achieve the aforementioned aim. This section will review how the study came to achieving these objectives.

- **Compare the performance of traditional Artificial Neural Networks (ANN) and deep Convolutional Neural Networks (CNN) in assessing aboveground biomass using Sentinel-2 data.**

In order to meet the first objective, this study utilised two neural networks, one conventional and one deep, to assess which one performed better in aboveground grass biomass predictions when using Sentinel-2 MSI spectral and VI data. Results from these investigations demonstrated that the contemporary deep CNN (R^2= 0.83, RMSE= 3.36 g/m^2, RMSE%= 6.09) outperformed the commonly-used ANN (R^2= 0.75, RMSE= 5.78 g/m^2, RMSE%= 8.90) in predicting aboveground grass biomass. The sensitivity analysis suggested that the GNDVI, GCI and blue band were the most important variables for model development for the CNN (in decreasing importance). In contrast, the blue band, GCI and GNDVI were the most important variables for the ANN (in decreasing importance) for biomass prediction in the dry season. Previous studies have also observed that deep CNNs tend to outperform ANNs in biomass monitoring applications (Ma et al., 2019, Karila et al., 2022, Varela et al., 2022). Therefore, it was concluded that deep CNNs are more accurate than ANNs and hence have a higher potential to yield more accurate and precise model predictions.

- **Predict inter-seasonal (dry and wet season) aboveground grass biomass using Sentinel-2 and deep learning technique (CNN).**

Following on from the first objective, the second objective was to utilise the better performing algorithm (CNN) to predict and distinguish aboveground grass biomass between the dry and wet seasons by also utilising Sentinel-2 MSI data and derived indices. Findings showed that the average aboveground grass biomass increased from 47.82 g/m^2 in the dry season to 195.67 g/m^2 in the wet season. This correlated with a significant increase in rainfall and temperature from the dry season to the wet season. Mean total rainfall increased from 16.5mm in the dry season to 96.84mm in the wet season. Concurrently, mean maximum daily temperature increased from 20.44 °C in the dry season to 27.42°C in the wet season. Findings by Van den Hoof et al. (2018) and Magandana et al. (2020) also showed an increase in grass biomass in the wet season due to increased rainfall and temperature. The CNN model performed slightly better in the wet season biomass predictions (R^2= 0.85, RMSE= 2.41 g/m^2, RMSE%= 3,71) as compared to the dry season biomass predictions (R^2= 0.83, RMSE= 3.36 g/m^2, RMSE%= 6.09). Similar findings were found by Jin et al. (2020) and Chen et al. (2021) on CNN model performance for grass biomass estimation between seasons.

4.3 A synthesis

This study has demonstrated the potential of integrating deep learning algorithms in analysing satellite imagery for biomass monitoring in KwaZulu-Natal, South Africa. It is evident from the investigations in this study that the contemporary deep CNN (R^2= 0.83, RMSE= 3.36 g/m^2, RMSE%= 6.09) outperformed the traditional ANN (R^2= 0.75, RMSE= 5.78 g/m^2, RMSE%=

8.90) in predicting aboveground grass biomass. The sensitivity analysis suggested that the GNDVI, GCI and blue band had the strongest correlation to biomass in developing the CNN model (in decreasing importance) whereas the blue band, GCI and GNDVI had the strongest correlation for the ANN model (in decreasing importance) for biomass prediction in the dry season.

Furthermore, results in this study show that the average aboveground grass biomass increased from 47.82 g/m^2 in the dry season to 195.67 g/m^2 in the wet season. During this period, mean total rainfall increased from 16.5mm in the dry season to 96.84mm in the wet season whereas mean maximum daily temperature increased from 20.44 °C in the dry season to 27.42°C in the wet season, respectively. This reiterated the notion that grass biomass is highly influenced by climatic changes. In addition, the CNN model performed slightly better in the wet season biomass (R^2= 0.85, RMSE= 2.41 g/m^2, RMSE%= 3,71) as compared to the dry season biomass predictions (R^2= 0.83, RMSE= 3.36 g/m^2, RMSE%= 6.09). The findings in this study indicated that emerging deep machine learning techniques have the potential to be paired with freely available satellite data and will perform better at data processing than previous shallow algorithms. Furthermore, a snapshot analysis of seasonal fluctuations in temperature and precipitation does result in aboveground changes of biomass, which inherently affected grassland productivity during drier periods.

4.4 Limitations and Recommendations

This section will first elaborate on the limitations encountered during of the study and thereafter will provide recommendations for future research on possible ways to address these limitations.

4.4.1 ANN and CNN requirements

One of the biggest limitations of the neural networks utilised in this study are the need for large datasets and computational power. This is especially so for deep learning algorithms such as the CNN as they are an emerging trend in the remote sensing academia. Numerous studies have mentioned that deep CNNs require large datasets to improve accuracy, and this often poses a challenge when data is limited or disjointed. Although CNNs have been previously tested on small datasets, future studies should utilise larger and more detailed datasets to improve on model development. This could entail encompassing different input data such as texture, lidar and topographical data to build more complex and productive CNN models.

4.4.2 CNN architecture and type

The type of CNN heavily depends on the type and amount of input data and variables that one has available. This study used a simple 2D CNN which, in simple terms, means that only single images were used as input data. Hence, future studies could attempt to incorporate multi-temporal data which would necessitate the use of 3D CNNs and thus potentially improve model performance. It is emphasised in the literature that the architecture and hyperparameters of CNNs, and machine learning algorithms in general, are paramount to respectable model performance. Hence, these statutes of model design must be geared appropriately in order to prevent poor model performance either by under- or over-fitting. Therefore, refining and improving deep CNN architecture for vegetation remote sensing applications is also another important research gap that can be fulfilled in future.

4.4.3 Utilise higher spatial resolution sensors

Sentinel-2 MSI was used in this study due to it being an open-access and readily available data source with high spatial and temporal resolution. Utilising a CNN in tandem with Sentinel-2 in this research yielded respectable results. However, future research could investigate the utility of different sensors, both open- and restricted-access, in CNN model performance. This could be in the form of newer satellite platforms like Landsat 8, higher spatial resolution satellites like WorldView-2 or with ultra-high-resolution imagery from UAVs, albeit at much smaller spatial scales.

4.5 Concluding remarks

The aim of this study was to predict inter-seasonal aboveground grass biomass using Sentinel-2 MSI and deep machine learning algorithms in communal grasslands in KwaZulu-Natal, South Africa. By comparing predictive performance of traditional and deep neural networks, this study decisively demonstrated that the highly advanced CNN algorithm predicted aboveground grass biomass between both dry and wet seasons amicably. This conclusion is consolidated based on observations throughout this research and answers the research questions posed in Chapter One:

- Which machine learning technique, between the Artificial Neural Network and the Convolutional Neural Network, performs more aptly at estimating aboveground biomass of grass when paired with Sentinel-2 bands and derived indices?

  After testing the performance of both algorithms in biomass predictions using Sentinel-2 MSI bands and 30 derived indices, it was found that the CNN was more adept at biomass predictions than the ANN. This was evident in the accuracy assessments as the CNN yielded the best R^2 of 0.83, RMSE of 3.36 g/m^2 and RMSE% of 6.09. In contrast, the ANN could only conjure the best R^2 of 0.75, RMSE of 5.78 g/m^2 and RMSE% of 8.90. Therefore, the CNN performed better in biomass predictions as compared to the ANN.

- Can remote sensing and deep learning be used to estimate and assess the difference in grass aboveground biomass between two distinct seasons in South Africa, being the dry season and the wet season? And what can this change in biomass be attributed to?

  This study successfully demonstrated the application of deep learning algorithms to remote sensing research on vegetation biomass, specifically grass biomass. Furthermore, this study conducted this research over two different seasons that occur within a South African climatic context, with slightly increased model performance in the wet season as compared to the dry season. Findings in this study reiterated the substantial effects of precipitation and temperature on biomass fluxes.

Overall, this study was a pilot study and a first attempt at applying deep CNN algorithms to vegetation research and monitoring in a southern Africa. This research could be used to establish rangeland management plans not only in communal grasslands, but also assist in informed decision-making in both small-scale and large-scale grassland ecosystems. This research also advances the theoretical and practical aspects of machine learning in the remote sensing academia.

REFERENCES

ALI, I., CAWKWELL, F., DWYER, E. & GREEN, S. 2016. Modeling managed grassland biomass estimation by using multitemporal remote sensing data—A machine learning approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 10**,** 3254-3264.

ALI, I., GREIFENEDER, F., STAMENKOVIC, J., NEUMANN, M. & NOTARNICOLA, C. 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing,* 7**,** 16398-16421.

ALVES OLIVEIRA, R., MARCATO JUNIOR, J., SOARES COSTA, C., NÄSI, R., KOIVUMÄKI, N., NIEMELÄINEN, O., KAIVOSOJA, J., NYHOLM, L., PISTORI, H. & HONKAVAARA, E. 2022. Silage Grass Sward Nitrogen Concentration and Dry Matter Yield Estimation Using Deep Regression and RGB Images Captured by UAV. *Agronomy,* 12**,** 1352.

BANNARI, A., ASALHI, H. & TEILLET, P. M. Transformed difference vegetation index (TDVI) for vegetation cover mapping. IEEE International geoscience and remote sensing symposium, 2002. IEEE, vol.5, 3053-3055.

BOVAL, M. & DIXON, R. 2012. The importance of grasslands for animal production and other functions: a review on management and methodological progress in the tropics. *Animal,* 6**,** 748-762.

BRODRICK, P. G., DAVIES, A. B. & ASNER, G. P. 2019. Uncovering ecological patterns with convolutional neural networks. *Trends in Ecology & Evolution,* 34**,** 734-745.

CHEN, J. M. 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing,* 22**,** 229-242.

CHEN, Y., GUERSCHMAN, J., SHENDRYK, Y., HENRY, D. & HARRISON, M. T. 2021. Estimating pasture biomass using Sentinel-2 imagery and machine learning. *Remote Sensing,* 13**,** 603.

CHO, M. A., ONISIMO, M. & MABHAUDHI, T. 2021. Using participatory GIS and collaborative management approaches to enhance local actors' participation in rangeland management: the case of Vulindlela, South Africa. *Journal of Environmental Planning and Management***,** 1-20.

CLEVERS, J. G. & GITELSON, A. A. 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation,* 23**,** 344-351.

DAS, M., GHOSH, S. K., CHOWDARY, V. M., MITRA, P. & RIJAL, S. 2022. Statistical and Machine Learning Models for Remote Sensing Data Mining—Recent Advancements. *Remote Sensing,* 14(8), 1906.

DASS, P., HOULTON, B. Z., WANG, Y. & WARLIND, D. 2018. Grasslands may be more reliable carbon sinks than forests in California. *Environmental Research Letters,* 13**,** 074027.

DEB, D., SINGH, J., DEB, S., DATTA, D., GHOSH, A. & CHAURASIA, R. 2017. An alternative approach for estimating above ground biomass using Resourcesat-2 satellite data and artificial neural network in Bundelkhand region of India. *Environmental Monitoring and Assessment,* 189**,** 1-12.

DEDEKIND, L. J., MCCOSH, J., EVERSON, T. M., MORRIS, C. D. & TROTTER, D. 2020. The value of rested sourveld in a communal grazing area in the Eastern Cape, South Africa. *African Journal of Range & Forage Science,* 37**,** 191-195.

DINGAAN, M. N. & TSUBO, M. 2019. Improved assessment of pasture availability in semi-arid grassland of South Africa. *Environmental Monitoring and Assessment,* 191**,** 1-12.

DONG, L., DU, H., HAN, N., LI, X., ZHU, D. E., MAO, F., ZHANG, M., ZHENG, J., LIU, H. & HUANG, Z. 2020. Application of convolutional neural network on lei bamboo above-ground-biomass (AGB) estimation using Worldview-2. *Remote Sensing,* 12**,** 958.

DUBE, T. & 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., PANDIT, S., SHOKO, C., RAMOELO, A., MAZVIMAVI, D. & DALU, T. 2019. Numerical assessments of leaf area index in tropical savanna rangelands, South Africa using Landsat 8 OLI derived metrics and in-situ measurements. *Remote Sensing,* 11**,** 829.

DUSSEUX, P., GUYET, T., PATTIER, P., BARBIER, V. & NICOLAS, H. 2022. Monitoring of grassland productivity using Sentinel-2 remote sensing data. *International Journal of Applied Earth Observation and Geoinformation,* 111**,** 102843.

ELLERY, W., SCHOLES, R. & SCHOLES, M. 1995. The distribution of sweetveld and sourveld in South Africa's grassland biome in relation to environmental factors. *African Journal of Range & Forage Science,* 12**,** 38-45.

FERNÁNDEZ-MANSO, A., FERNÁNDEZ-MANSO, O. & QUINTANO, C. 2016. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation,* 50**,** 170-175.

FEY, M. A short guide to the soils of South Africa, their distribution and correlation with World Reference Base soil groups. *Proceedings*, 2010. 32-35.

FYNN, R., MORRIS, C., WARD, D. & KIRKMAN, K. 2011. Trait–environment relations for dominant grasses in South African mesic grassland support a general leaf economic model. *Journal of Vegetation Science,* 22**,** 528-540.

GAMON, J. & SURFUS, J. 1999. Assessing leaf pigment content and activity with a reflectometer. *The New Phytologist,* 143**,** 105-117.

GITELSON, A. A. & MERZLYAK, M. N. 1997. Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing,* 18**,** 2691-2697.

GUERINI FILHO, M., KUPLICH, T. M. & QUADROS, F. L. D. 2020. Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. *International Journal of Remote Sensing,* 41**,** 2861-2876.

GXASHEKA, M., BEYENE, S. T., MLISA, N. & LESOLI, M. 2017. Farmers' perceptions of vegetation change, rangeland condition and degradation in three communal grasslands of South Africa. *Tropical Ecology,* 58**,** 217-228.

HAMADA, Y., ZUMPF, C. R., CACHO, J. F., LEE, D., LIN, C.-H., BOE, A., HEATON, E., MITCHELL, R. & NEGRI, M. C. 2021. Remote Sensing-Based Estimation of Advanced Perennial Grass Biomass Yields for Bioenergy. *Land,* 10**,** 1221.

HUETE, A., DIDAN, K., MIURA, T., RODRIGUEZ, E. P., GAO, X. & FERREIRA, L. G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment,* 83**,** 195-213.

HUETE, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment,* 25**,** 295-309.

HUNTER, F. D., MITCHARD, E. T., TYRRELL, P. & RUSSELL, S. 2020. Inter-seasonal time series imagery enhances classification accuracy of grazing resource and land degradation maps in a savanna ecosystem. *Remote Sensing,* 12**,** 198.

JENSEN, R. R., HARDIN, P. J. & YU, G. 2009. Artificial neural networks and remote sensing. *Geography Compass,* 3**,** 630-646.

JIN, X., LI, Z., FENG, H., REN, Z. & LI, S. 2020. Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index. *The Crop Journal,* 8**,** 87-97.

KARILA, K., ALVES OLIVEIRA, R., EK, J., KAIVOSOJA, J., KOIVUMÄKI, N., KORHONEN, P., NIEMELÄINEN, O., NYHOLM, L., NÄSI, R. & PÖLÖNEN, I. 2022. Estimating Grass Sward Quality and Quantity Parameters Using Drone Remote Sensing with Deep Neural Networks. *Remote Sensing,* 14**,** 2692.

KATTENBORN, T., LEITLOFF, J., SCHIEFER, F. & HINZ, S. 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 173**,** 24-49.

KAUFMAN, Y. J. & TANRE, D. 1996. Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from AVHRR to EOS-MODIS. *Remote sensing of Environment,* 55**,** 65-79.

KOBAYASHI, N., TANI, H., WANG, X. & SONOBE, R. 2020. Crop classification using spectral indices derived from Sentinel-2A imagery. *Journal of Information and Telecommunication,* 4**,** 67-90.

KOERNER, S. E. & COLLINS, S. L. 2014. Interactive effects of grazing, drought, and fire on grassland plant communities in North America and South Africa. *Ecology,* 95**,** 98-109.

LAKHAL, M. I., ÇEVIKALP, H., ESCALERA, S. & OFLI, F. 2018. Recurrent neural networks for remote sensing image classification. *IET Computer Vision,* 12**,** 1040-1045.

LI, C., ZHOU, L. & XU, W. 2021. Estimating aboveground biomass using Sentinel-2 MSI data and ensemble algorithms for grassland in the Shengjin Lake Wetland, China. *Remote Sensing,* 13**,** 1595.

LIU, X., HAN, F., GHAZALI, K. H., MOHAMED, I. I. & ZHAO, Y. A review of convolutional neural networks in remote sensing image. Proceedings of the 2019 8th International Conference on Software and Computer Applications, 2019. 263-267.

MA, J., LI, Y., CHEN, Y., DU, K., ZHENG, F., ZHANG, L. & SUN, Z. 2019. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *European Journal of Agronomy,* 103**,** 117-129.

MAGANDANA, T. P., HASSEN, A. & TESFAMARIAM, E. H. 2020. Seasonal herbaceous structure and biomass production response to rainfall reduction and resting period in the semi-arid grassland area of South Africa. *Agronomy,* 10**,** 1807.

MAIN-KNORN, M., PFLUG, B., LOUIS, J., DEBAECKER, V., MÜLLER-WILM, U. & GASCON, F. Sen2Cor for sentinel-2. Image and Signal Processing for Remote Sensing XXIII, 2017. International Society for Optics and Photonics, 1042704, 37-48. 2017 Oct 4, San Francisco.

MAS, J. F. & FLORES, J. J. 2008. The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing,* 29**,** 617-663.

MASEMOLA, C., CHO, M. A. & RAMOELO, A. 2020. Sentinel-2 time series based optimal features and time window for mapping invasive Australian native Acacia species in KwaZulu Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, 93**,** 102207.

MASENYAMA, A., MUTANGA, O., DUBE, T., BANGIRA, T., SIBANDA, M. & MABHAUDHI, T. 2022. A systematic review on the use of remote sensing technologies in quantifying grasslands ecosystem services. *GIScience & Remote Sensing,* 59**,** 1000-1025.

MOYO, B., DUBE, S. & MOYO, P. 2013. Rangeland management and drought coping strategies for livestock farmers in the semi-arid savanna communal areas of Zimbabwe. *Journal of Human Ecology,* 44**,** 9-21.

MUCINA, L. & RUTHERFORD, M. C. 2006. *The vegetation of South Africa, Lesotho and Swaziland*, South African National Biodiversity Institute, Pretoria, South Africa.

MUKANGANGO, M., NDUWAMUNGU, J., NARAMABUYE, F. X., NYBERG, G. & DAHLIN, A. S. 2020. Supplementing grass-based cattle feeds with legume leaves and its effects on manure quality and value as a soil improver for an Anthropic Ferralsol in Rwanda. *Experimental Agriculture,* 56**,** 483-494.

MUTANGA, O., DUBE, T. & AHMED, F. 2016. Progress in remote sensing: vegetation monitoring in South Africa. *South African Geographical Journal,* 98**,** 461-471.

NARAYANAN, B., SAADELDIN, M., ALBERT, P., MCGUINNESS, K. & MAC NAMEE, B. 2021. Extracting pasture phenotype and biomass percentages using weakly supervised multi-target deep learning on a small dataset. *arXiv preprint arXiv:2101.03198*.

O'MARA, F. P. 2012. The role of grasslands in food security and climate change. *Annals of Botany,* 110**,** 1263-1270.

O'CONNOR, T. G., MARTINDALE, G., MORRIS, C. D., SHORT, A., WITKOWSKI, T. & SCOTT-SHAW, R. 2011. Influence of grazing management on plant diversity of highland sourveld grassland, KwaZulu-Natal, South Africa. *Rangeland Ecology & Management,* 64**,** 196-207.

PALMER, A., SHORT, A. & YUNUSA, I. A. 2010. Biomass production and water use efficiency of grassland in KwaZulu-Natal, South Africa. *African Journal of Range & Forage Science,* 27**,** 163-169.

PANG, H., ZHANG, A., KANG, X., HE, N. & DONG, G. 2020. Estimation of the Grassland Aboveground Biomass of the Inner Mongolia Plateau Using the Simulated Spectra of Sentinel-2 Images. *Remote Sensing,* 12**,** 4155.

PIRES DE LIMA, R. & MARFURT, K. 2020. Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing,* 12**,** 86.

PODWOJEWSKI, P., JANEAU, J. L., GRELLIER, S., VALENTIN, C., LORENTZ, S. & CHAPLOT, V. 2011. Influence of grass soil cover on water runoff and soil detachment under rainfall simulation in a sub-humid South African degraded rangeland. *Earth Surface Processes and Landforms,* 36**,** 911-922.

RAMOELO, A. & CHO, M. A. 2014. Dry season biomass estimation as an indicator of rangeland quantity using multi-scale remote sensing data.

RAMOELO, A., CHO, M. A., MATHIEU, R., MADONSELA, S., VAN DE KERCHOVE, R., KASZTA, Z. & WOLFF, E. 2015. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation and Geoinformation,* 43**,** 43-54.

RASCH, S., HECKELEI, T. & OOMEN, R. J. 2016. Reorganizing resource use in a communal livestock production socio-ecological system in South Africa. *Land Use Policy,* 52**,** 221-231.

REINHART, K. O., DANGI, S. R. & VERMEIRE, L. T. 2016. The effect of fire intensity, nutrients, soil microbes, and spatial distance on grassland productivity. *Plant and Soil,* 409**,** 203-216.

RICHARDSON, F., HOFFMAN, M. T. & GILLSON, L. 2010. Modelling the complex dynamics of vegetation, livestock and rainfall in asemiarid rangeland in South Africa. *African Journal of Range & Forage Science,* 27**,** 125-142.

ROFFE, S., FITCHETT, J. & CURTIS, C. 2019. Classifying and mapping rainfall seasonality in South Africa: a review. *South African Geographical Journal= Suid-Afrikaanse Geografiese Tydskrif,* 101**,** 158-174.

ROFFE, S. J., FITCHETT, J. M. & CURTIS, C. J. 2020. Determining the utility of a percentile-based wet-season start-and end-date metrics across South Africa. *Theoretical & Applied Climatology,* 140.

ROUJEAN, J.-L. & BREON, F.-M. 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment,* 51**,** 375-384.

ROYIMANI, L., MUTANGA, O., ODINDI, J., SIBANDA, M. & CHAMANE, S. 2022. Determining the onset of autumn grass senescence in subtropical sour-veld grasslands using remote sensing proxies and the breakpoint approach. *Ecological Informatics,* 69**,** 101651.

RUST, J. & RUST, T. 2013. Climate change and livestock production: A review with emphasis on Africa. *South African Journal of Animal Science,* 43**,** 255-267.

SAMIMI, C. & KRAUS, T. 2004. Biomass estimation using Landsat-TM and-ETM+. Towards a regional model for Southern Africa? *GeoJournal,* 59**,** 177-187.

SANTOSO, H., GUNAWAN, T., JATMIKO, R. H., DARMOSARKORO, W. & MINASNY, B. 2011. Mapping and identifying basal stem rot disease in oil palms in North Sumatra with QuickBird imagery. *Precision Agriculture,* 12**,** 233-248.

SCHREIBER, L. V., ATKINSON AMORIM, J. G., GUIMARÃES, L., MOTTA MATOS, D., MACIEL DA COSTA, C. & PARRAGA, A. 2022. Above-ground Biomass Wheat Estimation: Deep Learning with UAV-based RGB Images. *Applied Artificial Intelligence***,** 1-15.

SCOTT-SHAW, R. & MORRIS, C. D. 2015. Grazing depletes forb species diversity in the mesic grasslands of KwaZulu-Natal, South Africa. *African Journal of Range & Forage Science,* 32**,** 21-31.

SHOKO, C., MUTANGA, O., DUBE, T. & SLOTOW, R. 2018. Characterizing the spatio-temporal variations of C3 and C4 dominated grasslands aboveground biomass in the Drakensberg, South Africa. *International Journal of Applied Earth Observation and Geoinformation,* 68**,** 51-60.

SIBANDA, M., MUTANGA, O., ROUGET, M. & KUMAR, L. 2017. Estimating biomass of native grass grown under complex management treatments using worldview-3 spectral derivatives. *Remote Sensing,* 9**,** 55.

SIBANDA, M., ONISIMO, M., DUBE, T. & MABHAUDHI, T. 2021. Quantitative assessment of grassland foliar moisture parameters as an inference on rangeland condition in the mesic rangelands of southern Africa. *International Journal of Remote Sensing,* 42**,** 1474-1491.

SINGH, L., MUTANGA, O., MAFONGOYA, P. & PEERBHAY, K. Y. 2018. Multispectral mapping of key grassland nutrients in KwaZulu-Natal, South Africa. *Journal of Spatial Science,* 63**,** 155-172.

THÉAU, J., LAUZIER-HUDON, É., AUBÉ, L. & DEVILLERS, N. 2021. Estimation of forage biomass and vegetation cover in grasslands using UAV imagery. *PloS one,* 16**,** e0245784.

TUCKER, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment,* 8**,** 127-150.

VAN DEN HOOF, C., VERSTRAETE, M. & SCHOLES, R. J. 2018. Differing responses to rainfall suggest more than one functional type of grassland in South Africa. *Remote Sensing,* 10**,** 2055.

VARELA, S., ZHENG, X.-Y., NJUGUNA, J., SACKS, E., ALLEN, D., RUHTER, J. & LEAKEY, A. D. 2022. Deep convolutional neural networks exploit high spatial and temporal resolution aerial imagery to predict key traits in miscanthus. *AgriRxiv***,** 20220405560.

VUNDLA, T., MUTANGA, O. & SIBANDA, M. 2020. Quantifying grass productivity using remotely sensed data: an assessment of grassland restoration benefits. *African Journal of Range & Forage Science*, 37**,** 247-256.

WANG, J., XIAO, X., BAJGAIN, R., STARKS, P., STEINER, J., DOUGHTY, R. B. & CHANG, Q. 2019. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154**,** 189-201.

XIE, Y., SHA, Z., YU, M., BAI, Y. & ZHANG, L. 2009. A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China. *Ecological Modelling*, 220**,** 1810-1818.

YANG, S., FENG, Q., LIANG, T., LIU, B., ZHANG, W. & XIE, H. 2018. Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. *Remote Sensing of Environment*, 204**,** 448-455.

YUAN, Q., SHEN, H., LI, T., LI, Z., LI, S., JIANG, Y., XU, H., TAN, W., YANG, Q. & WANG, J. 2020. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241**,** 111716.

ZHU, X. X., TUIA, D., MOU, L., XIA, G.-S., ZHANG, L., XU, F. & FRAUNDORFER, F. 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5**,** 8-36.