LAND USE AND LAND COVER CHANGE DETECTION AND ASSESS FUTURE CHANGE USING REMOTE SENSING AND MACHINE LEARNING ALGORITHMS: A STUDY OF UMNGENI RIVER CATCHMENT

By

Orlando Bhungeni

Dissertation submitted to the College of Agriculture, Engineering, and Sciences, at the University of KwaZulu-Natal, in fulfilment of the academic requirement of the degree of Master of Science (MSc) in Environmental Science

(Specialization: Remote Sensing)

Westville Campus

November 2024

PREFACE

This dissertation was submitted under the School of Agriculture, Earth and Environmental Sciences (SEAS), University of KwaZulu-Natal, Westville Campus, for the degree of Master of Science by Research in Geography and Environmental Sciences.

This dissertation has been prepared according to Format 2 as outlined in the Information for the guidance of examiners of Higher degrees, which states:

Format 2: As a set of papers which are published, in press, submitted, or intended for submission.

As the candidate's supervisor, I have approved this dissertation for submission.

……..

Supervisor: Date 13-02-2025

Prof. Michael T. Grebreslasie

DECLARATIONS

Declaration 1 – Plagiarism

I, Orlando Bhungeni declare that

1. The research reported in this dissertation is my original research, except where otherwise indicated.

2. This dissertation has not been submitted for any degree or examination at any other university.

3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This dissertation 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 have been referenced, and

b. where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This dissertation 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 section/s.

Signed … [redacted] ……

Orlando Bhungeni

Declaration 2: Publications and manuscripts

Details of contribution to publications that form part and/or include research presented in this dissertation (include 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)

1. Bhungeni, O., Ramjatan, A., and Gebreslasie, M. (2024). Evaluating Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal. *Remote Sensing*, *16*(12), 2219. https://doi.org/10.3390/rs16122219

2. Bhungeni, O., Ramjatan, A., and Gebreslasie, M. LULC Change Detection and Future LULC Modelling Using Rf and MLPNN-Markov Algorithms in the Umngeni Catchment, Kwazulu-Natal, South Africa (intended for submission)

From the aforementioned manuscripts, my role included conducting all the data analysis work and contributing to the drafting of original drafts of the manuscripts together with my supervisor. While other authors contributed to the reviewing and editing

Signed: Orlando Bhungeni

Date: 19 January 2025

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to my academic supervisor, Prof. Michael Gebreslasie, for his passion, patience, and dedication in crafting my growth as an emerging remote sensing scientist. I am also thankful for his invaluable guidance and unwavering support throughout the journey of my research project, which has widened my perspective and enriched my academic journey.

I would also like to show my gratitude to Dr. Sifiso Xulu for sharing his knowledge and experience in data analysis and academic writing.

I would like to extend my acknowledgement to Mr. Thembelani Dlamini, the GIS Lab technician, for his technical support during the data analysis period. His expertise and assistance have been crucial in facilitating my research and have greatly contributed to my overall work.

I would like to acknowledge all my fellow students, Nasiphi Ngcoliso, Bolelang Tsolo, Irvin Doctor Shandu, Collins Matiza, and Fallon Kemp for sharing their experiences and creating a positive learning environment.

I also thank Mrs Ashadevi Ramjatan for her support and for organising fieldwork operations. I would also like to acknowledge her valuable contributions to individual manuscripts of my work.

I would like to express my special gratitude to my mother Nonkululeko Bhungeni, for her dimensional support and understanding in allowing me to pursue further studies have been invaluable, and I am deeply grateful for her love and sacrifices.

I acknowledge the kindness and support provided by Departmental administrative staff Dumsile Ndlovu during my study period. I thank Mr Xolani Gumede for his assistance and support as well.

I thank the South African National Space Agency (SANSA) for funding my MSc studies.

Abstract

Water catchment areas are the key regional strategic water source with a range of ecological benefits such as providing habitat for biodiversity, climate change mitigation, sport, and recreation. Detailed temporal information on the spatial attributes of the catchment areas is paramount for water resource managers to support water resource conservation strategies. Data from remote sensing provide a good playground for the advanced statistical tools to study land cover patterns in the water catchments. Thus, our study leveraged remote sensing data and explored options of Machine Learning Algorithms (MLAs) for the classification of thematic Land Use and Land Cover (LULC) classes within the uMngeni River Catchment area. This study used Landsat 8 Operational Land Imager (L8-OLI) data along with Naïve Bayes (NB), Random Forests (RF), Support Vector Machine SVM), and Artificial Neural Networks (ANN). The RF model proved its robustness with high accurate overall accuracy and kappa of 97.02% and 0.96, respectively. Hence, the RF was subsequently utilised to assess LULC change detection for 20 years (from 2003-2023). The interplay among the land use classes was observed, with common transitions to built-up class and a noticeable constant decline of the forest land use class. Change detection maps enabled the modelling of 2032 land cover maps using the hybrid model from Multi-Layer Perceptron Neural Network (MLPNN) and Markov Chain (MLNN-Markov). The combination of RF and L8-OLI is suitable for mapping the LULC in the case of a South African complex landscape. The uMngeni River catchment possesses fair land cover distribution with an alarming built-up land cover class at the cost of indigenous forests, which is expected to increase by 3.7% in 2032. The findings of this work will also demonstrate the practical application of MLAs in mapping LULC and support the water resource managers in formulating management strategies aimed at safeguarding the integrity of the water resources.

CONTENTS

PREFACE i

Declarations ii

Declaration 1 – Plagiarism ii

Declaration 2: publications and manuscripts ii

ACKNOWLEDGEMENTS iv

Abstract v

CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES ix

1 CHAPTER 1: INTRODUCTION 1

1.1 Background 1

1.2 Problem Statement 1

1.3 Literature Review 2

1.4 Motivation and Objectives 5

1.5 Research Questions 6

1.6 Dissertation structure 6

1.7 Description of the chapter content 7

1.8 Limitations of the study 8

2 CHAPTER 2: EVALUATING MACHINE-LEARNING ALGORITHMS FOR MAPPING LULC OF THE UMNGENI CATCHMENT AREA, KWAZULU-NATAL .9

2.1 Introduction 10

2.2 Materials and Methods 13

2.2.1 Study Area 14

2.2.2 Definition of LULC Classes Umngeni River Catchment 16

2.2.3 Remote Sensing Data Acquisition and Preprocessing 17

2.2.3.1 Calculated Spectral Indices 18

2.2.4 Reference Data Collection 18

2.2.5 Image Classification 19
2.5.1. Selected classifiers and parameter tuning 20
2.2.6 Variable importance 24
2.2.7 Accuracy assessment 25
2.3 Results 26
2.3.1 Mapping and spatial extent LULC classes 27
2.3.2 Comparison of Machine Learning Algorithms Mapping Accuracy 30
2.3.3 Model Feature Importance 32
2.4 Discussion 34
2.5 Conclusion 38
3 CHAPTER 3: HALF-DECADE INTERVALS LULC CHANGE DETECTION AND FUTURE LULC MODELLING USING RF AND MLPNN-MARKOV ALGORITHMS IN THE UMNGENI CATCHMENT, KWAZULU-NATAL, SOUTH AFRICA 40
3.1 Introduction 40
3.2 Methods and material 44
3.2.1 Study area 44
3.2.2 Satellite image collection and Pre-processing 46
3.2.3 Land use classification and accuracy assessment 47
3.2.4 Land use change detection 49
3.2.5 Predictor variable selection and transition sub-model development 49
3.2.6 Transition potential data 49
3.2.7 Transition probability matrix 50
3.2.8 Land cover prediction and validation 50
3.2.9 Model validation 51
3.3 Results 51
3.3.1 Prediction of the LULC scenarios for the year 2032 57
3.4 Discussion 60
3.4.1 Projection 64

3.5 Conclusion 65

3.6 Limitations and recommendations 66

4 CHAPTER 4: SYNTHESIS REPORT OF THE DISSERTATION 68

4.1 Synthesis of the chapters 68

4.1.1 Evaluating Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal 68

4.1.2 Chapter 3: Half-Decade Intervals LULC Change Detection and Future LULC Modelling Using RF and MLPNN-Markov Algorithms in the uMngeni Catchment, Kwazulu-Natal, South Africa 69

References 71

LIST OF FIGURES

Figure 2.1. The conceptual framework of the study. 14

Figure 2.2.The uMngeni River catchment layout with elevation gradients, river network, and dams. 16

Figure 2.3.The variable collinearity heatmap. Blue to red shows an increasing correlation relationship and vice versa. 27

Figure 2.4.Standardised average reflectance values of each land use class 27

Figure 2.5.UMngeni River catchment LULC maps produced by NB, RF, SVM, and ANN models with detailed small portions in (a) to (h). 28

Figure 2.6. The spatial distribution of land use classes produced by each MLA. 29

Figure 2.7. Confusion matrix and accuracy metrics for NB, RF, SVM, and ANN. 32

Figure 2.8. Model feature importance based on the mean absolute SHAP value. 33

Figure 3.1.The uMngeni catchment layout with elevation gradients, river network, and dams. 45

Figure 3.2. Historical LULC thematic maps from 2003 to 2023. 52

Figure 3.3. Half-decadal area coverage (km^2) from 2003 to 2023 and the predicted LULC class for 2032. 53

Figure 3.4. Land use class changes in percentage for five-year time steps from 2003 to 2023. 54

Figure 3.5.UMngeni River Catchment percentage change detection for 2003,2023 and 2032. 54

Figure 3.6.Statistical relationship between land use class changes and time. 55

Figure 3.7. Land use class pixel count for the 2019 reference map and 2019 simulated map.59

Figure 3.8. Simulated land use map for 2032. 60

LIST OF TABLES

Table 2.1. Shows the description of nine thematic LULC classes used in the current study. .16

Table 2.2. L8-OLI spectral bands details and their respective spatial resolutions 17

Table 2.3.Shows the supplementary indices used in the land use classification. 18

Table 2.4.Training and test samples used in the study. 19

Table 2.5.Hyperparameter tuning of the MLA models.....23

Table 2.6.Thematic land use classes area covers a report by RF.....29

Table 3.1.Training (Tr) and test (Ts) count per LULC class for six study terms (2003-2023). 46

Table 3.2. Summary of satellite imagery used from 2003-2023.....47

Table 3.3.Time step land use type coverage in km^2 and percentage from 2003 to 2023. 55

Table 3.4. Accuracy assessment metrics for six-term interval thematic maps from the Random Forests model. 56

Table 3.5. Sensitivity of MLPNN model at different environmental variable combinations with training accuracy. 57

Table 3.6. Transition probability grid for 2019 prediction Trial 1. 58

Table 3.7. Accuracy assessment metrics for the 2019 simulated map.....58

1 CHAPTER 1: INTRODUCTION

1.1 Background

The water catchment areas are important ecological components that provide various ecosystem services that benefit society and biodiversity (Aneseyee et al., 2020; Mekuria et al., 2023). These include the availability of freshwater for consumption, supporting societal needs such as food production, sport, and recreation, and providing a natural habitat (Lynch et al., 2023; Pullanikkatil et al., 2016). In addition, water catchment areas are the major role players in mitigating climate change (Abbas et al., 2023; Gyamfi-Ampadu et al., 2021; Mugo et al., 2020). Such essential services can be sustained when the biophysical environment of the water catchment areas is sustainably managed. The provision of these services relies on the good health of the ecosystem, physical components such as the river network, and other land resources necessary to supply these services. Catchment ecosystems such as that of the uMngeni are vulnerable to both natural disasters stemming from climate change and human intrusion (Ruppert et al., 2015).

1.2 Problem Statement

Natural land cover and freshwater resources in the water catchment areas are threatened by Land use/Land cover changes (LULC-C) in response to societal demands (Nyathi et al., 2020; Staponites et al., 2022). The trajectory of these changes has implications for water quality (Schütte and Schulze, 2017), natural habitat and climate change mitigation, ultimately leading to dysfunction of the ecosystem and human health risks (Mandal et al., 2023; Mugo et al., 2020; Plessis et al., 2014). In a period of 30 years (1960 to 2019), an estimated 32 % of the global land has been converted or modified. Among other land uses, forests lost a considerable total land cover of about 0.8 million km^2 at the expense of global agricultural expansion (Winkler et al., 2021). In the period of 11 years (1994 to 2005), at least 1.2% of South African total land was transformed, where a great proportion of the increase was attributed to urban expansion (Fairbanks et al., 2000). These changes are linked to societal demands and natural factors such as the availability of food to provide for rapid population growth, expansion of human settlements, and industrial developments (Jewitt et al., 2015; Kibena et al., 2014; Mararakanye et al., 2022; Namugize et al., 2018). In addition, natural disasters stemming from

climate change also play a part in land cover modifications, such as floods, droughts, and soil erosion.

Boakye et al. (2008) pointed out that densely populated areas are subject to undergo environmental changes, particularly in developing countries. The population density status of the given city of the region determines the rate of LULC modification. The province of KwaZulu Natal (UKZN) is the second most densely populated province in South Africa, with about 21% of the national population is residing in KZN (Stats, 2011). Making it prone to the transformation of its landscape; for example, between 2005 and 2011, approximately 7.6% of the natural land in the province was lost due to the expansion of agriculture, residential areas, and mining dunes (Jewitt et al., 2015) and the competition on the land space. These changes in the uMngeni River Catchment are expected to be exacerbated by 2050 (Mauck and Warburton, 2014; Moodley et al., 2023).

Many studies on the relationship between LULC and water quality status have been conducted; for example, Namugize et al. (2018) found that water bodies near residential areas and waste treatment plants have higher concentrations of substances. Such concentrations make the water unfit for both domestic use and aquatic life (Jewitt et al., 2015). In addition, Aghsaei et al. (2020) pointed out that in a region with limited water resources, LULC-C can also result in water scarcity. Therefore, it is paramount to find reliable and accurate means to present a better understanding of the LULC configuration. For this reason, research based on remote sensing data, in partnership with recently developed Machine Learning Algorithms (MLAs), emerged as a better solution to combat the water quality related issues. Aiding in decision-support information, playing an integral part in strategy development, surveillance of land use activities, and management of water resources.

1.3 Literature Review

The progress made on the advanced remote sensing technologies in mapping and monitoring the water catchment areas has been determined through rigorous search on both Google Schooler and Scopus. The search method was facilitated by the Boolean operators together with keywords associated with each manuscript's specifications. Other supplementary filters, such as year of publication, region, and disciplines were also specified to narrow the search based on the research theme. The sets of criteria included ‘Catchment’ AND Machine

Learning' AND 'lulc' AND 'remote sensing' AND 'classification' OR 'mapping'. The second scheme was based on the application of remote sensing on change detection and foresting future land use changes; the combination search words included 'forecasting' OR simulation' AND 'lulc' AND 'catchment' AND change detection. The methods were used iteratively to region adjustment from continental to local with a static period set to return only articles published between 2014 and 2023 and limited to only remote sensing, environmental science, natural science, and ecology disciplines. This approach is based on artificial intelligence and is designed to scoop the sets of keywords specified only in the title, abstract, and keywords of the references in the online databases and subsequently return all the references that meet the criteria.

The returned search results were further scrutinised based on the relevance through reading the abstracts. The most relevant literature was reviewed to get insight into the recent remote sensing advancements in mapping the LULC, forecasting the future land use status, and identifying the gaps, as presented in each manuscript. The evidence in the reviewed literature reaffirms that there have been continual improvements in the LULC spatial monitoring of the water catchment areas in the last decade.

Recently, remote sensing remained the reliable source that provides historic time-stamped spatiotemporal data, which is crucial for detecting land cover changes (Campbell and Wynne, 2011). In addition, the introduction of moderate to high-resolution multispectral data enhanced the digital image classification techniques in the remote sensing field. Those techniques today have been widely used in remote sensing as reliable methods for the interpretation of digital images.

The most popular digital image classification methods include pixel-based and object-based image analysis (OBIA). Pixel-based image classification methods only consider the individual pixel's spectral properties to derive each class's signature. It uses the probability density function derived from the training pixels to predict label pixels. Once the probability is calculated, the pixel is assigned to a class with a higher probability (Lillesand et al., 2015). In addition, it assumes the pixel spectral properties of each class in the image is normal (Mather and Koch, 2011). While pixel-based approach, OBIA methods identify the boundaries of the homogeneous attributes based on the unique spectral characteristics of the pixels and create interlocking boundaries to form regions (Campbell and Wynne, 2011). The regions are

segmented using a hierarchical approach which forms the subdivisions in recognition of the geospatial scale of the objects within the regions (Mather and Koch, 2011).

The recent advancement of statistical learning tools, machine learning-based digital image processing, emerged in the remote sensing application. It shows good potential to analyse the trends and the LULC at different scales (Biau and Scornet, 2016). Peterson et al. (2019) defined machine learning as the set of mathematical methods that learn the properties of data from the provided samples and perform classification based on the learned similarities. In contrast to the parametric classifiers, MLAs do not consider the statistical parameters (e.g., the threshold derived from minimum and maximum values of the training pixels) (Mather and Koch, 2011). Instead, machine learning requires the training of the models with known data sets that allow the models to learn the variations and the pattern of the pixels of the same class. This allows precise labeling of the given image, detection of the object, and recognition of the trends in a supervised or semi-supervised manner (Belgiu and Drăguţ, 2016).

Hence, the literature reviewed highlighted the improvements in classification accuracies and recommended machine learning models in conjunction with the various remote sensing data in explaining the trajectory of LULC within the river catchments. The literature revealed that the applications of machine learning methods have been providing valuable information to address water quality-related issues across the world. The capabilities of the MLAs in mapping the LULC have been tested using the hyperspectral and very high spatial resolution (VHR) multispectral data in different terrestrial settings. For example, Akar and Tunc Gormus, (2022); Lekka et al., (2024); Delogu et al., (2023) achieved outstanding classification overall accuracies with a threshold of 88.1% to 97%. While comparative studies leveraged moderate to low-resolution sensors data yielded good overall accuracies ranging from 75% to 95.83.

The most prominent machine learning classifiers and imagery data with rich archives have also been used in jointly not solely for monitoring the catchment land cover changes but also for forming the bases for predicting the future LULC scenarios. Landsat data and a range of MLAs which include—Random Forests (RFs), OBIA, and Support Vector Machines (SVMs) have made the most appearances in the literature on the analysis of the LULC within the catchment areas and are deemed to be efficient in tracing the past long term land use status. While studies

on the LULC the land cover modeling recommended the hybrid models formed statistical methods and machine learning such as cellular Automata (CA), Markov Chain, and multi-layer perceptron (MLP). The main aim of the adoption of hybrid algorithms is to compensate for the shortfalls of the land cover modeling algorithms as each model has its advantages and drawbacks (Hafeez et al., 2019; Loukika et al., 2021).

Evidence presented in the literature reviewed proved the excellence of MLAs in the classification of large and complex data. However, it has also been picked up that their performance varies with the layout of the landscapes (Bayas et al., 2022). As there is no universal method that can be applied to all geographic locations. Thus, we can not generalise and recommend the method in the case of South Africa; as of this date, there is no study that has ever compared these MLAs in the South African water key strategic areas. Hence the current study seeks to identify the most suitable approach in mapping the LULC within the uMngeni River Catchment among the prominent methods highlighted in the literature. Furthermore, there are limited studies of a similar kind that have stretched the application of the identified machine-learning model to unpack historical land use activities with the intention of predicting future land cover scenarios. That is why this study further conducted 20-year change detection analysis from 2003 to 2023 and simulated the land cover maps for the next 10 years (2023 to 2032).

1.4 Motivation and Objectives

Despite the efforts made to control LULC transformations and their related implications, the trajectory of the LULC continues to '*form islands of protected areas in a matrix of anthropogenically transformed areas'* (Jewitt et al., 2015). Therefore, Considering the ecological, cultural, and socio-economic significance of the uMngeni River Catchment, it is necessary to closely monitor the landscape of its nature at a short interval period. Hence, the current study seeks to fulfill the following aims and objectives:

Aim

The study is expected to improve the understanding of LULC configuration over the uMngeni River catchment, a key element in building a LULC change and projection model within the catchment.

objectives

i. To assess the performance of different machine learning classifiers to map the land use classes within the uMngeni River Catchment;
ii. to evaluate Landsat series spatial-temporal data in mapping the historical LULC change for the period of 20 years (2003 to 2023);
iii. to predict the plausible future LULC scenarios in the uMngeni River Catchment.

Such information is not only important for understanding the current and past land cover status, but it also provides the preliminary data that enables the prediction of future LULC scenarios. Unveiling of such quantitative information is paramount to water resource managers and other environmental health organisations in the development of strategic planning, policies, and other intervention measures to safeguard water resources based on the updated information. Hence, the opportunity for the current study exists.

1.5 Research Questions

The proposed study seeks to address the following questions:

i. What are the suitable machine learning algorithms for mapping uMngeni river Catchment land use classes?
ii. What is the land use class distribution status within the uMngeni catchment?
iii. What are land cover classes change rates from 2003 to 2023?
iv. What are the likely plausible land cover scenarios by 2032?

1.6 Dissertation structure

This work is a publication-format dissertation composed of four chapters. The first chapter gives a general overview of the study (Chapter One), followed by three chapters; comparing machine learning methods for mapping catchment land use categories (Chapter Two), land use/Land cover change detection and prediction of future land cover scenarios (Chapter Three), and finally summarize the dissertation (Chapter Four). As stated in the Declaration 2 section, Chapter Two has undergone peer review and has been published in an international journal. Chapter Three is being revised and intended to be submitted to the journal concurrently. The composition and content of these two chapters are preserved in the dissertation

compilation as per journal format. As a result, the layout of manuscripts is composed of an introduction, materials and methods, results, discussion, conclusion and reference list. Certain sections have overlaps and reputations because of the manuscript's structure preservation. Since chapters are the prerequisites of the other and focus on one study setting. For example, methods (study site description, data collection, and references). However, the manuscript's originality is preserved in line with the plagiarism policies.

1.7 Description of the chapter content

Chapter 1

Contains a general overview of the study and a literature review on the application of both remote sensing and machine learning tools in studying the land use developments in regional watershed areas. The literature focussed on the accuracies and ecological settings of studies conducted on advanced land use/land cover classification methods, and those factored in the spatiotemporal resolution of the remote sensing platforms. Ultimately, highlighting the significance, gap, and main theme of the study.

Chapter 2

In Chapter Two, the performance of non-parametric machine learning methods (Random Forests, Support Vector Machines, Artificial Neural Networks, and Naïve Bayes) was compared in classifying the high dimensional land use land cover of the catchment landscape leveraging Landsat imagery. The performances of the classifiers and Landsat data were evaluated using the commonly used recommended methods.

Chapter 3

As the prerequisite of Chapter Two, the best-performed machine learning model was used to generate the primary data for prediction of the future land use land cover status. So, the toolkit comprised of Random Forests, Multi-layer Perceptron Neural Network, Landsat series data, and the range of environmental variable data was used to produce historical land use maps for 2003 to 2023 and simulate the future land use/land cover scenarios for 2032.

Chapter 4

This chapter is a synthesis of the overall dissertation.

1.8 Limitations of the study

The limitations of the study have been listed below;

The RF model struggled to separate land cover classes with similar optical properties, for example, mining quarries and tar roads or parking lots. Due to its insensitivity to outliers tends to be biased to land use class with the higher training sample allocation.

The secondary invasion of harvested areas by the herbaceous and grass species interferes with the interplay between barren land and cropland, leading to grassland overestimation.

The residential areas with defined tree canopies and green grass open spaces (lawn) partially masked the building structure, causing mixed pixels, making it not guaranteed to award the pixels to the built-up class. The transition zones between the two land parcels tend to possess mixed pixels, as a result of confusing the classifier. Ultimately, impacted the accuracy of the prediction maps.

2 CHAPTER 2: EVALUATING MACHINE-LEARNING ALGORITHMS FOR MAPPING LULC OF THE UMNGENI CATCHMENT AREA, KWAZULU-NATAL

remote sensing

MDPI

Article

Evaluating Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal

Orlando Bhungeni *, Ashadevi Ramjatan and Michael Gebreslasie

School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban 4000, South Africa; 822820676@stu.ukzn.ac.za (A.R.); gebreslasie@ukzn.ac.za (M.G.) *
Correspondence: 223146923@stu.ukzn.ac.za

Abstract: Analysis of land use/land cover (LULC) in catchment areas is the first action toward safeguarding freshwater resources. LULC information in the watershed has gained popularity in the natural science field as it helps water resource managers and environmental health specialists develop natural resource conservation strategies based on available quantitative information. Thus, remote sensing is the cornerstone in addressing environmental-related issues at the catchment level. In this study, the performance of four machine learning algorithms (MLAs), namely Random Forests (RFs), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Naïve Bayes (NB), were investigated to classify the catchment into nine relevant classes of the undulating watershed landscape using Landsat 8 Operational Land Imager (L8-OLI) imagery. The assessment of the MLAs was based on a visual inspection of the analyst and commonly used assessment metrics, such as user's accuracy (UA), producers' accuracy (PA), overall accuracy (OA), and the kappa coefficient. The MLAs produced good results, where RF (OA = 97.02%, Kappa = 0.96), SVM (OA = 89.74%, Kappa = 0.88), ANN (OA = 87%, Kappa = 0.86), and NB (OA = 68.64%, Kappa = 0.58). The results show the outstanding performance of the RF model over SVM and ANN with a significant margin. While NB yielded satisfactory results, its sensitivity to limited

training samples could primarily influence these results. In contrast, the robust performance of RF could be due to an ability to classify high-dimensional data with limited training data

Keywords: uMngeni river catchment; machine learning; LULC; Landsat 8; remote sensing; South Africa

2.1 Introduction

The catchment areas are important ecological components that provide various ecosystem services that benefit society and biodiversity (Aneseyee et al., 2020; Mekuria et al., 2023). These include the availability of freshwater for consumption, supporting societal needs such as food production, sport, and recreation, and providing a natural habitat (Lynch et al., 2023; Pullanikkatil et al., 2016). Such essential services can be sustained when the biophysical environment of the catchment areas is sustainably managed. However, natural land cover and freshwater resources in the water catchment areas are threatened by Land use/Land cover (LULC) changes in response to societal demands (Nyathi et al., 2020). It is essential to map and monitor the water catchment landscape dynamics to provide valuable information on the distribution of the land use activities in the catchment, which serve as the tool for initiatives whose mandates aim to strike a balance between developments and the natural environment in the catchment areas. In addition, this information will help water resource managers and other environmental health practitioners establish environmental management plans and policies based on available information (Gyamfi-Ampadu et al., 2020; Obaid et al., 2023). Moreover, the availability of such information is important as it can be used to troubleshoot water quality issues and prompt the identification of land resource degradation (Namugize et al., 2018).

Mapping the water catchment land use activities can be achieved by either traditional field observation or remote sensing methods. Remote sensing emerged as the convenient and cost-effective way to provide the spatiotemporal data required to analyze heterogeneous landscapes at a large scale (Ouchra et al., 2023). In addition, the remote sensing data can be systematically stored, maintained, and openly shared to end users. A variety of remote sensing sensors (optical and active sensors) have been utilized to map the LULC in the catchment areas. Akar and Tunc Gormus, (2022) assessed the influence of the contextual data features and various dimensionality filters, including Discrete Wavelet Transform (DWT), in mapping LULC using Headwall Hyperspec data in Bergama, Turkey. The random forest (RF) and support vector machine (SVM) were employed, and RF and DWT data obtained higher overall accuracy and kappa

scores of 88.13% and 0.88, respectively. In comparison, SVM recorded lower accuracy than RF. Lekka et al., (2024) have used Environmental Mapping and Analysis Program (EnMAP) imagery data to map the LULC using pixel-based machine learning algorithms (MLAs), such as RF and SVM image classification. Their results showed that the SVM image classifier returned an overall accuracy of 92.6%, while the RF model obtained an overall accuracy of 88.1%. Delogu et al., (2023) used PRISMA hyperspectral data with machine learning methods—such as RF, artificial neural network (ANN), and convolutional neural network (CNN), to classify the land use activities in Naples, Italy. The CNN achieved the best results at 0.973 and 0.968 overall accuracy and Kappa, respectively. As the second-best model, ANN recorded 0.963 and 0.956, and RF showed the lowest classification accuracy at 0.887 and 0.867, respectively.

Other studies have demonstrated the potential of very high spatial resolution (VHR) multispectral data in providing the same service as hyperspectral data in different terrestrial settings and achieving higher classification accuracies. For instance, James et al., (2020), utilized Pleiades-1 imagery to map the LULC in Côte d'Emeraude, France. Maximum likelihood (ML) and SVM algorithms were used to implement the classification. ML algorithm yielded a higher overall accuracy of 84.64% compared to SVM, which only produced a map with an overall accuracy of 76.13%. In Fayoum City, Egypt, Mahmoud et al., (2023) evaluated a variety of classifiers in mapping the city land cover classes using PlanetScope imagery. They found the ANN model more accurate with a 0.97 kappa compared to Naïve Bayes (NB), SVM, RF, and decision trees (DT), which recorded 0.93, 0.90, 0.86, and 0.87, respectively.

Some studies explored the capabilities of moderate to low-resolution sensors' capabilities in classifying spatial land use classes. For example, in a comparative study, Gyamfi-Ampadu et al., (2020) employed Landsat-8 (L8-OLI) data and SVM and RF algorithms to map Nkandla Forest cover classes in South Africa. L8-OLI was found efficient in mapping the forest cover, and SVM was found more accurate with overall accuracy and kappa records of 95.83 % and 0.94, compared to 95.24% and 0.93 achieved by RF, respectively. On the west coast of India, a study by Bayas et al., (2022) assessed the efficiency of the MLAs in mapping LULC using Sentinel-2 (S2) imagery. NB, Classification And Regression Trees (CART), gradient tree boost (GTB), and RF MLAs were assessed, and the models achieved an overall accuracy of 81.34%, 90.18%, 91.06%, and 92.14%, respectively. A study by Ouchra et al., (2023) also used L8-OLI data to evaluate the capabilities of MLAs in land use analysis in Casablanca, Morocco. The

RF, SVM, CART, and GTB algorithms were among the tested methods, and RF showed robust performance with an overall accuracy of 95.42%, compared to other classifiers with an overall accuracy of 83%, 91.50%, and 93.46%, respectively. A study by Parashar et al., (2024) compared SVM and RF algorithms using S2 datasets without and with spectral indices to analyze the LULC hilly terrain of Gopeshwar, India. The RF models provided more accurate results in both datasets, with overall accuracy and kappa of 86% and 0.83 for a dataset with only spectral bands and 88% and 0.85 for a dataset with incorporated spectral indices, respectively. While SVM reported the least accuracy, with overall accuracy and kappa of 82% and 0.79 for spectral bands and 87% and 0.83 for incorporated indices, respectively. In the Munneru River Basin, India, Loukika et al., (2021) explored the performance of the MLAs viz. SVM, RF, and CART are used to map the LULC based on L8-OLI and S2 data. They found S2 and RF to dominate their counterparts. The average overall accuracy for RF, SVM, and CART in L8-OLI was 94.85%, 90.88%, and 82.88%. In comparison, S2 average overall accuracy for RF, SVM, and CART was 95.84%, 93.65%, and 86.48%, respectively. Woldemariam et al., (2022) evaluated the MLAs in classifying the land use dynamics using L8-OLI data in the Lake Haramaya catchment, Ethiopia. Object-based image analysis (OBIA), SVM, RF, and ANN were tested, and the SVM dominated other classifiers with an overall accuracy of 94%, while RF and ANN achieved 92% and 89%, respectively. While OBIA reported the least overall accuracy at 75%. Dash et al., (2023), compared the parametric and non-parametric classifiers in mapping LULC based on L8-OLI data in the Big Sunflower River watershed, United States of America (USA). RF, SVM, and the ML were factored in, and their results showed SVM showed a robust performance with an overall accuracy (kappa) of 93.5% (0.88), 88.8% (0.82), and RF with the least performance at 84.6%, (0.72).

The success of land cover mapping depends not only on the remote sensing data but also on the classification methods utilized. As discussed above, various machine-learning algorithms have been used in the remote sensing field to explain the distribution of spatial features in catchment areas, with varying degrees of accuracy reported. Their performance varies with the biophysical features of the landscapes (Bayas et al., 2022). Hence, there is no generic approach for LULC classification for all terrestrial landscapes (Woldemariam et al., 2022). Therefore, it is vital to compare the MLA models to determine their suitability and sensitivity to different geographic settings (Woldemariam et al., 2022). Studies that compared the MLAs in mapping LULC in the case of South African landscapes if is there any, are still limited. Thus, in this

study, we have identified four prominent MLAs, such as NB, RF, SVM, and ANN and were selected to conduct a comparative study to map the thematic land use classes of the uMngeni river catchment in KwaZulu Natal using L8-OLI data. The current study aims to (i) map the LULC that defines the uMngeni River catchment area and (ii) compare the performance of four non-parametric pixel-based MLAs in a supervised classification based on L8-OLI data. The study is expected to improve the understanding of LULC configuration over the uMngeni River catchment, a key element in building a LULC change and projection module within the catchment, to generate information for water resource management strategies.

2.2 Materials and Methods

The analysis of the LULC was carried out in a series of steps, as illustrated in **Figure 2.1.** The acquired data was preprocessed before implementing the machine learning algorithms, and the LULC was computed. The post-classification analysis was carried out, including accuracy assessment and computing variable importance.

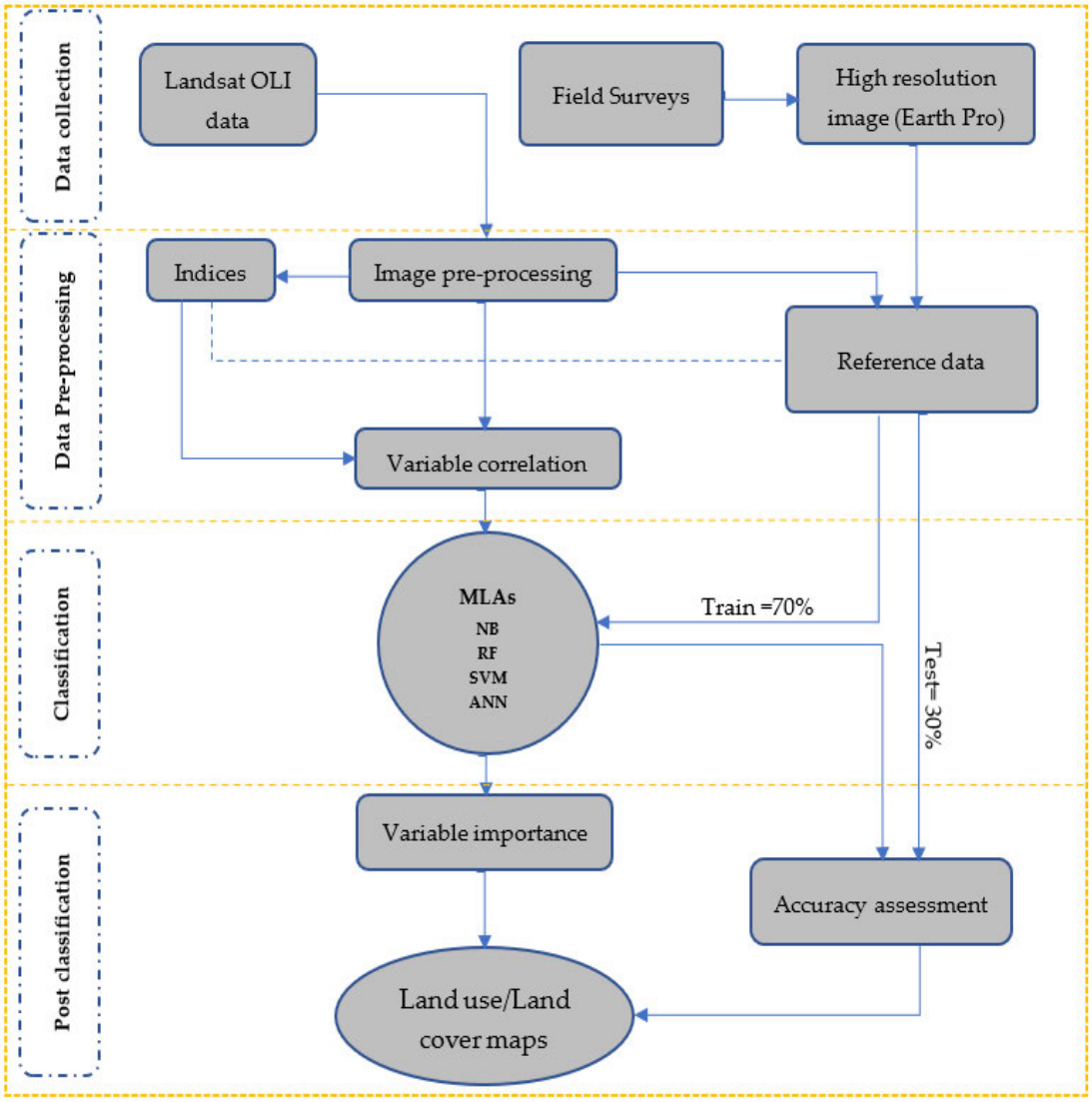

Figure 2.1. The conceptual framework of the study.

2.2.1 Study Area

The study area is located in the uMngeni River catchment, KwaZulu-Natal, South Africa (29°29'05.0"S 29°46'52.8"E and 29°48'41.8"S 31°02'27.0"E. The area is situated in the central-western part of KwaZulu-Natal province (**Figure 2.2**). It covers 4444 km^2, which is subdivided into three main regions (lower, middle, and upper) (Hughes et al., 2018). Its biophysical environment is characterized by a complex river network with an undulating landscape in the upper reaches and gentle slopes with hills towards the lower parts of the catchment, shown in green in **Figure 2.2**.

The river network forms the uMngeni river, which stretches from Drakensberg hills in the inner lands of KZN and flows down through Pietermaritzburg and final deposits in the Indian Ocean

via the Durban north estuary. The uMngeni River has four main dams (Midmar, Albert Falls, Nagle, and Inanda), which serve as a water source for 45% of the KZN province's total population (Namugize et al., 2018), including industrial activities, recreation, and agricultural purposes in the cities along the river path, such as Pietermaritzburg and Durban (Hughes et al., 2018). Land cover change implications not only affect these services but also have an impact on habitat loss, climate change, and biodiversity. To ensure the continuity of such important services, the uMngeni River Catchment was chosen as the study area of interest.

The catchment has an elevation of 1913 m above sea level (Kusangaya et al., 2018). Rainfall occurs during the summer (October to March), ranging from 700 to 550 ml per annum. It has moderate temperatures with an annual average of 12 to 20 °C. Its climate and geographic position play a critical role in the province's economic development. Due to these biophysical characteristics, natural features, and economic activities, the catchment is characterized by a range of land use types.

From our field observation, the land cover types in the catchment include the following but are not limited to built-up: green patches (cropland, plantation, and forest), barren land, and water bodies. The Natural vegetation and cropland dominate the uMngeni catchment (Namugize et al., 2018). The plantation is one of the important land use classes in the catchment, which dominates the upper parts of the catchment. The middle part is characterized by heterogeneous patches of different land use types (plantation and built-up areas) (Namugize et al., 2018). While the lower part is mainly dominated by built-up patches with fragments of green vegetated land cover. It is worth unpacking the information in such biodiversity-rich areas to provide a piece of updated scientific information.

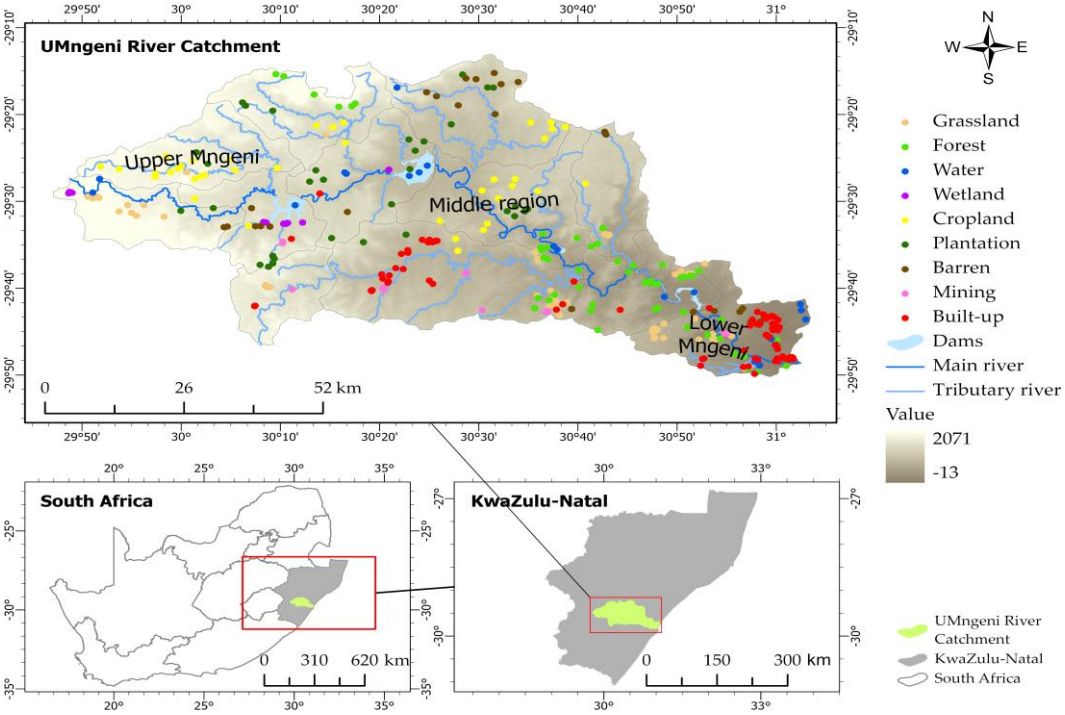

Figure 2.2.The uMngeni River catchment layout with elevation gradients, river network, and dams.

2.2.2 Definition of LULC Classes Umngeni River Catchment

To assess the LULC in the catchment, a total of nine thematic land use classes were identified, which include built-up, cropland, water, plantations, mining, forest, barren, grassland, and wetlands, as described in **Table 2.1**.

Table 2.1. Shows the description of nine thematic LULC classes used in the current study.

Class ID	LULC classes	Description
1	**Grassland**	Includes grass and shrubs open areas, as well as golf courses, and sport fields grounds
2	**Forest**	Natural forests or transition foarest areas rangeing from low to highly dense canopy cover, dominated with trees of appraximately 5 m high.
3	**Water**	This class considers natural or artificial surface fresh-waterbodies within the study area, which includes main rivers and tributaries, dams, lakes, and ponds
4	**Wetland**	This class is mainly made up of hydrphytes and herbaceous species which are solely based on the moisture of the soil to survive. The wetlands can be permanent or temporary depending on the local climate or its size and water holding capacity.
5	**Cropland**	Composed of both subsistance and ecommercial farming, which might be annual crops, or seasonal. During the post harvest, cultivated lands are charecterised as bare land in the case of seasonal crops.
6	**Plantation**	Range of tree species cultivated for commercial purposes, this class includes greener patches of the mature trees, sub adult, young, jiviniles, and tree stumps.
7	**Barren**	Partial vegetated areas, bare land due to either erosion, natural degradation or human factors, post harvest crop fields, dry river banks, and stripped rock areas.
8	**Mining**	Small patches of mining scars, which comprise of open cast pits, sand mining, and the open raw material processing sites.
9	**Built-up**	Built-up structures, which includes all forms residential areas, economic corridors, industrial, commercial, educational, religious and health infrastructures.

2.2.3 Remote Sensing Data Acquisition and Preprocessing

In this study, a 30 m resolution image from L8-OLI image was used for LULC analysis (**Table 2.2**). The multispectral image was obtained on 15 May 2023 from the open Google Earth Engine (GEE) data catalog, in surface reflectance, level 2 category (https://developers.google.com/earth-engine/datasets/catalog/landsat). The surface reflectance category provides images that are ready for analysis with a set of preprocessing requirements that have already been performed to improve the quality of the image (Kaur et al., 2023). According to (Mandal et al., 2023), surface reflectance images do not require fully blown image preprocessing. Hence the *ee.filter* functions were used to set the criteria to select data that meet the needs of the study. It must also be noted that scenes of different dates but the same season have been collected using *ee.Filter.date().* As the images of the same season will not have an impact on the analysis, since the atmospheric variation is minimal. A total of four scenes (each with a total area of 185 km by 180 km) from two rows and paths were collected to cover our study area. Then, the *filterMetadata* function was set to return only scenes with a cloud cover of less than 10% (Maviza and Ahmed, 2020).

Table 2.2. L8-OLI spectral bands details and their respective spatial resolutions

Bands	Wavelength (λ)	Resolution (m)
Blue	0.45–0.51μm	30 m
Green	0.53–0.59 μm	30 m
Red	0.64–0.67 μm	30 m
Near Infrared (NIR)	0.85–0.87 μm	30 m
Shortwave Infrared 1 (SWIR 1)	1.56–1.65 μm	30 m
Shortwave Infrared 2 (SWIR 2)	2.10–2.29 μm	30 m

However, the filtering processing alone was infeasible to prepare the image for the analysis. Therefore, image pre-processing was performed as the prerequisite for any satellite image in the GEE interface. The median composite was formed from a series of scenes from the winter season, 01 May to 30 August 2022.

To stretch our specifications, the boundary shape file was used to create a subset of our study area from the composite image using the *Clip()* function. Then, the subset was subject to further

processing, as the clouds are unavoidable in the satellite image (Foga et al., 2017); the function of the mask from C code, commonly known as CFmasking, was applied to remove cloud-infested pixels in the image, as used by (Mashala et al., 2023). The approach involves the quality assurance (QA-PIXEL) band, which assigns zero values to the cloud-contaminated pixels and cloud shadows, producing a cloud-free composite (Floreano and de Moraes, 2021).

2.2.3.1 Calculated Spectral Indices

Spectral indices are derived pixel-wise ratios of two or more bands of the multispectral image, which serve as both logistics and added explanatory variables for image classification (Orieschnig et al., 2021). Many studies have achieved meaningful LULC classification results by supplementing multispectral bands with information from spectral indices (Loukika et al., 2021; Orieschnig et al., 2021; Palanisamy et al., 2023). The indices layers facilitate the discrimination of specific features and assign the pixels to predicted classes. Our study area possesses diverse spatial attributes, making distinguishing the class members from others difficult. Hence, the current study employed various mathematical equations **(Table 2.3)** to compute the Normalized difference indices (NDI) for vegetation (NDVI) (Rouse et al., 1974) to improve the separation of vegetation classes (Viana et al., 2019), water indices (NDWI) to distinguish waterbodies from hill shadows (Gao, 1996), and bare soil indices (BSI) (Rikimaru et al., 2002) to identify bare surfaces (Nguyen et al., 2021). These indices were also used as additional explanatory inputs to L8-OLI visible, NIR, and SWIR bands.

Table 2.3.Shows the supplementary indices used in the land use classification.

Index	Equation	Rerefences
NDVI	$NDVI = \frac{NIR - Red}{Red + NIR}$	[30]
NDWI	$NDWI = \frac{Green - NIR}{Green + NIR}$	[32]
BSI	$BSI = \frac{((Red + SWIR) - (NIR + Blue))}{((Red + SWIR) + (NIR + Blue))}$	[33]

2.2.4 Reference Data Collection

Field survey approach and Google Earth Pro 7.3.6 high resolution images were used in collecting sampling data for building and measuring the mapping accuracy of the classifiers. The field survey was carried out on the 16th of August 2023, where coordinates of predefined thematic land use classes (**Table 2.1**) were geo-located using a portable global positioning system (GPS)

through a random sampling method. The same method was also used in Google Earth Pro-based sample collection. This approach widely places the sampling points across the study area (**Figure 2.2**) and ensures all the land cover classes are adequately represented in the reference data (Carlotto, 2009; Foody, 2010). The GPS coordinates were imported into the Google Earth Pro 7.3.6 platform and overlaid on high-resolution images to locate the land cover classes (Olofsson et al., 2014). Around each coordinate, the irregular-sized polygons were then digitized to represent the corresponding land cover class (**Table 2.4**). The digitized reference polygons were imported into GEE and overlayed to L8-OLI image. Then the spectral properties of cover classes were extracted from the L8-OLI image per pixel, as shown in **Table 2.4** and randomly divided into two datasets, a training dataset (70%) and a validation data set (30%) (Obaid et al., 2023). These two datasets were used to train and assess the accuracy of the MLA models—NB, RF, SVM, and ANN.

Table 2.4.Training and test samples used in the study.

LULC Class	No. of polygons	Training pixel count	Test pixel count
Grassland	63	6944	1685
Forest	51	5823	1483
Water	38	1865	435
Wetland	11	357	86
Cropland	38	8965	2163
Plantation	32	7820	1957
Barren	30	4492	1075
Built-up	86	2504	598
Mining	13	667	159

2.2.5 Image Classification

Regions with diverse land use classes can make it difficult for machine learning models to discover the relationship between explanatory features (Gandhi, 2021). Therefore, before commencing the classification stage, it is important to scrutinize the variables based on their spectral similarities in the given class. To address that, the Pearson correlation coefficient was used to perform variable correlation analysis. It is a commonly used pair-wise technique that helps to identify the collinearity between two variables and eliminate noise (Gyamfi-Ampadu et al., 2021). It improves the accuracy of the MLAs by reducing redundancy and improving the computation processing rate. The stepwise correlation analysis was performed in L8-OLI explana-

tory variables; (1) spectral reflectance information for each variable was extracted from sampling sites, (2) then imported into Python to compute variable correlation using *corr()* algorithm (3) Seaborn's package was employed to plot the correlation array as a heat map matrix.

2.5.1. Selected classifiers and parameter tuning

Naïve Bayes

The NB classification approach evolved from the Bayesian and Naive theorem. Bayesian technique assumes the possibility of a recurrence of the event based on learning from the previous event (Costache et al., 2022). While Naive is a probabilistic classifier with a strong independent assumption of the feature from the others (Palanisamy et al., 2023). In other words, dimensions that define a certain variable are independent of each other. Naïve Bayesian theory can be expressed mathematically as the probability of a pixel being in class *k given* X as training sample, which is equivalent to the density of the sample size given the class, multiplied by the probability of the pixel being in class *k*, divided by the density of sample size:

$$P(y = k|X) = \frac{p(X|y=k)(P(y=k)}{p(X)} \quad (1)$$

Where *y* is the unassigned variable, *k* is the class, and X is the training sample. Therefore P(y = *k*|X) is the probability of the *y* and p is the density. NB models can be trained in supervised classification methods either in the form of parametric or non-parametric models by incorporating Gaussian functions.

The NB model is simple to build, with the focus on adjusting the lambda (λ) parameter. Which plays a central role in ensuring that LULC classes not recognized during training are not assigned a zero probability (Gandhi, 2021).The model computes the statistics (mean and the variance of the variables for each class) required for classification from the provided training data (Bayas et al., 2022). In LULC analysis, it has been reported as a promising classifier with satisfactory results (Palanisamy et al., 2023). Its performance is attributed to the independence of the feature dimensions from the others. For the purpose of our study, the NB model from "*smileNaiveBayes"* library in GEE was used.

Random Forests

The RF classifier is an ensemble classifier that uses regression decision trees to classify the dataset (Breiman, 2001). It gained popularity in the remote sensing community due to its reduced computational processing time, unique classification approach, and robust performance (Orieschnig et al., 2021). The RF proposed by Breiman involves bagging the CART split criterion. Bagging involves building bootstrap samples from the original training data. This process is also known as bootstrapping. Then the predictor trees are then built from each bootstrap sample. The class membership of the unknown pixel is determined by the votes from all decision trees (Obaid et al., 2023). The decision to assign the pixels to the class is based on the majority results from multiple RF decision trees (Mather and Koch, 2011; Zhang et al., 2023). The advantage of the RF models is the ability to handle high-dimensional datasets (Bayas et al., 2022) and it is insensitive to outliers and overfitting (Cuypers et al., 2023). The RF models left out some samples for each newly generated tree, commonly known as the out-of-bag samples (OOB), which report on the important features in the classification and misclassification errors of each tree (Kim et al., 2014). For the model to meet analysts' needs, requirements must be predefined in the parameter tunning stages.

Two important parameters to be set when building the model are the number of decision trees to be created (*Ntree*) and the number of variables to be selected per split (*Mtry*). The higher value of *Ntree* does not improve the accuracy of the classifier; instead, it brings confusion. Fewer *Ntree* increases the confidence in features used to build the decision trees and reduces the correlation between the electorates (Cuypers et al., 2023; Orieschnig et al., 2021). In contrast, *Ntree* can be as big as possible, while *Mtry* should be equal to the square root of the number of features (Camargo et al., 2019). There is no standardized way to adjust the parameters for RF models. Therefore, it depends on the discretion of the analyst and the nature of the study. In this study, the RF classifier from the *"smileRandomForest*" library was utilized. To develop the model, Ntree was adjusted to 25, *Mtry* was set to three, which is equal to the square root of explanatory features, and each node was adjusted to have 2 leaves, as detailed in Table 2.5. Each decision tree received a share of 0.5 % from the original sample data for OOB, nodes per tree were set not to exceed 1000. The optimization of the parameters was obtained using 10-fold cross-validation.

Support Vector Machine

The SVM algorithm is a supervised vector-based classifier that has been used for both regression and classification (Yulianto et al., 2023). Different types of SVM models have been widely used for different remote sensing applications which include LULC, water quality, and vegetation mapping. SVM models use discriminative methods by learning the distribution patterns of training data. It uses linear or nonlinear lines (known as hyperplanes) to optimize the boundaries between the vector data. The development of SVM models was designed as a binary classifier that only deals with linear data (Huang et al., 2002). The method used the principle of finding the best marginal line (hyperplane) between the two features in the high-dimensional space, then the optimal hyperplane serves as the decision boundary that separates the two features. This principle generally works well when applied to linear data.

However, it is challenging to separate nonlinear data using the linear SVM approach. Because the inherent design of its model tends to not completely separate the training dataset, leading to classification errors (Lantzanakis et al., 2021). After numeracy attempts to develop SVMs suitable for classifying nonlinear data, the joint effort from SVMs and kernel functions was the viable way to separate the nonlinear data. The primary role of the kernel function in SVM models is to mobilize the training data to a new dimensional space, allowing the marginal lines to find the optimal hyperplane to separate the vectors (Srivastava et al., 2012). The most popular kernel functions for the classification of complex remote sensed data include the radial basis function (RBF) and polynomial kernel (Martins et al., 2016). These two kernels have been reported in the literature for their ability to produce accurate classification results; for example (Roushangar et al., 2023), used RBF-SVM to monitor dynamic land use changes in Urmia Lake Iran, as did (Saraf et al., 2022) in urban expansion mapping using a polynomial kernel function. Martins et al., (2016) compared the performance of the kernels, and they found RBF to be slightly more accurate than the polynomial kernel in delineating the sub-catchment.

As there is no rule of thumb in kernel selection, the analyst's discretion is used based on the nature of the study. Therefore, for the current study, the polynomial kernel was selected based on the trial-error approach. The applicable SVM type for polynomial kernel in sensing data analysis is support vector classifier (SVC), which allows for the cost (C) of non-separatable

vectors (C-SVC). Thus, it is also known as soft margin because it requires only the best hyperplane and accepts some separation errors as defined in the C threshold; if value C is high, it can lead to overfitting. In this way, C-SVC guarantees the results of which their integrity is highly dependent on the training data and parameter tuning of the model (**Table 2.5**). The C-SVC model has been used in different studies for LULC classification and achieved meaningful results, such as (Loukika et al., 2021).

The C-SVC from the "*libsvm"* library in GEE was used. The kernel and C were set to polynomial and 10 respectively, as the main parameters to control marginal lines and error threshold. Additionally, hyperparameters were also tuned, including a degree, gamma, and coef0, to achieve the optimum hyperplane, as shown in **Table 2.5**.

Table 2.5.Hyperparameter tuning of the MLA models.

Classifier	Classifiers	Parameters	Parameter adjustments
NB	*ee.Classifier.smileNaiveBayes()*	λ	0.000001
RF	*ee.Classifier.smileRandomForest()*	Ntree	25.00
		Mtry	null
		MinLeafPopulation	2.00
		MaxNodes	1000.00
		BagFraction	0.50%
SVM	*ee.Classifier.libsvm()*	kernelType	POLY
		svmType	C_SVC
		coef0	0.3
		degree	1.00
		cost	10
ANN	*MLPClassifier (ANN)*	activation	RELU
		hidden_layer	1.00
		nuerons	10

Artificial Neural Network

ANN is a non-parametric algorithm with a complex network of interconnected neurons. The network is composed of an input phase, one or more intermediate phases, and an output phase. In remote sensing applications, the input layer has neurons that serve as receptor sites for data from each variable. They pass the data to the intermediate phase via channels. Each channel

has a numeric value assigned, known as weight, which is the factor of the inputs to be transmitted to the neurons of the intermediate phase (Dahiya et al., 2023). Neurons in the intermediate phase house the numeric values called bias, which are added to the inputs. Then, the total neuron values are subjected to the activation function, which determines the neurons that are going to contribute to the inputs of the next phase (Wu et al., 2014); all the activated neurons transmit the inputs to the next phase; the process is called forward propagation. The inputs propagated through the network to the output phase, where the inputs are assigned to a corresponding class (Chen, Lin, et al., 2023). In addition, the misclassified data is backpropagated through the network model iteratively for revision by adjusting the weights. In this study, the feed-forward propagation ANN model from the *Sci-Kit learn* package applicable in Python (Google Colab cloud computing platform) was used, as defined:

$$\text{Output} = aS(\sum_{i-1}^{n} w_i \, wz_i + b) \qquad (2)$$

Where w_i and zi are weights that are carried by the channels, b is the bias as the remainder of the training samples, i is the number of neurons in the intermediate phase.

According to (Deshpande and Emmanuel, 2016), ANN models, in general, tend to overestimate the classes. The network size and the number of neurons to be used are very important parameters to consider to achieve the best results; the large complex network can lead to overfitting of the input data (Hafeez et al., 2019). The number of intermediates was set to 1 phase with 10 iterations. The nonlinearity was solved by the RELU activation function, as detailed parameter tuning is presented in **Table 2.5**.

2.2.6 Variable Importance

The inherent designs of MLA models are complex in nature, and it is challenging to interpret their interaction with the input data (Lundberg et al., 2017.) However, understanding the model behaviour is important (Wei et al., 2015) to determine the utility features and gain an insight into the mechanisms of the models (Hosseiny et al., 2022; Yilmaz and Kavzoglu, 2022). Thus, the built models (NB, RF, SVM, and ANN) were further investigated to explain the interaction between models and features in classifying given LULC classes (Table 2.1) using the Shapley Additive exPlanation (SHAP) value. Its approach is based on game theory, which explains the

impact of individual features on the prediction of machine learning models (Lundberg et al., 2017.). To compute the SHAP value, the Sci-Kit learn and shap packages in Python were used to build both MLA and SHAP value explainer models, respectively.

2.2.7 Accuracy Assessment

The most used accuracy assessment standard for land use mapping is the Kappa coefficient (Olofsson et al., 2014). The Kappa assesses the status of classification results against the randomly assigned values. The nature of the Kappa values ranges from −1 to 1. These kappa values rank the classification accuracy of the classified image, Where the value is less than zero, no agreement, greater than ≤ 0.40 fair, ≤ 0.75 satisfactory, and ≤ 1 indicates good performance (Hua, 2017). The Kappa can be defined as follows:

$$K = \frac{P(A)-P(E)}{1-P(E)} \tag{3}$$

Where $P(A)$ is the overall accuracy and P(E) chance agreement, which is given by the sum of the predicted values and actual values for each class an agreement by chance (Congalton, 1991).

Note: j= land use class of interest (e.g. water bodies), will be used as an example to determine the accuracies of the class of interest.

The overall accuracy represents the sum of correctly labelled pixels for all the classes in the input image as per provided test sample sites and is mathematically expressed as a fraction of correctly labelled pixels and the total number of the test sample sites (Kang et al., 2021).

$$OA = \frac{\sum \text{correctly leballed pixels}}{\sum \text{pixels in test sample sites}} \tag{4}$$

PA is the number of pixels that the test sample of the given class (j) agreed to labelled pixels as class j in the input image. It is given by the sum of correctly labelled pixels of a class j, divided by the sum of correctly labelled pixels and the pixels claimed to not belong to class j, as adapted from (Olofsson et al., 2014):

$$PA_j = \frac{\sum \text{correctly leballed pixels}j}{\sum \text{correct leballed pixels}j + \text{misleballed as class}j} \quad (5)$$

The user's accuracy is determined by the number of correctly labelled pixels of class divided by the sum of the pixels class claimed to belong to class *j* in the test samples, as proposed by (Story and Congalton, 1986).

$$UA_j = \frac{\sum \text{correctly leballed pixels}j}{\sum \text{correct leballed pixels}j + \text{leballed as class}j} \quad (6)$$

As an effort to assess the performance of NB, RF, SVM, and ANN models, the aforementioned accuracy metrics will be calculated from the confusion matrix of each model, as presented in **Figure 2.7**.

2.3 Results

Figure 2.3 illustrates the result of the L8-OLI variable collinearity test, as discussed in section 2.3. One variable of the highly correlated pairs was removed. For example, ultra-blue and blue bands show a linear spectral relationship of 0.98, as shown in red to light red cells in **Figure 2.3**. In contrast, variables that showed the least correlation or distinct spectral variation were retained to facilitate the classification process. As a result, in a total of eleven variables tested, six multi-spectral bands such as visible (blue, green, and red), NIR, and SWIR (1 and 2) bands were selected as explanatory variables for the LULC classification along with three spectral indices (NDVI, NWDI, and BSI). To demonstrate the unique spectral information of each variable per class, the spectral variability was computed, as illustrated in **Figure 2.4.**

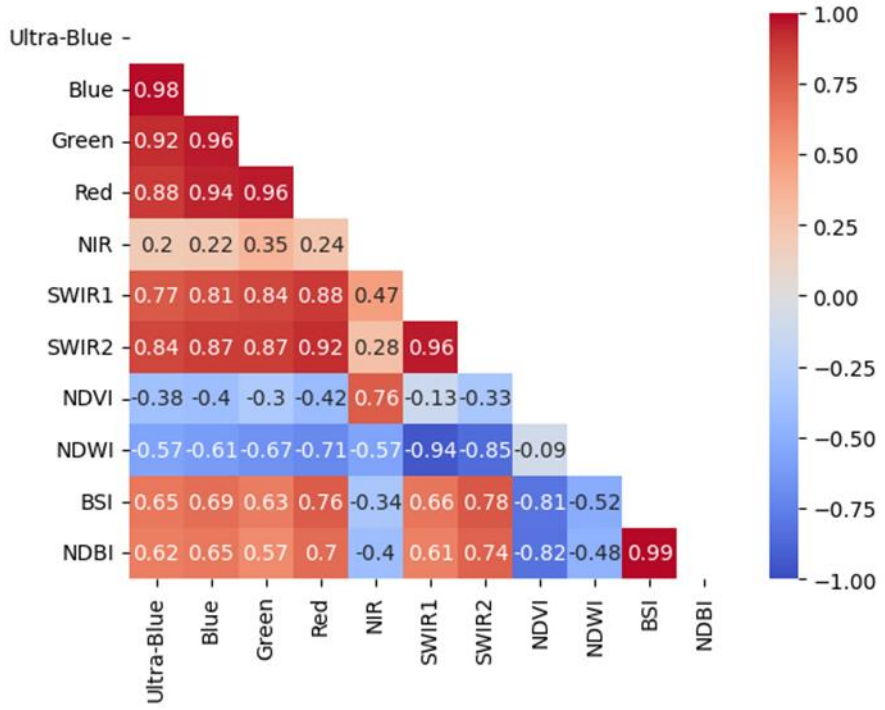

Figure 2.3.The variable collinearity heatmap. Blue to red shows an increasing correlation relationship and vice versa.

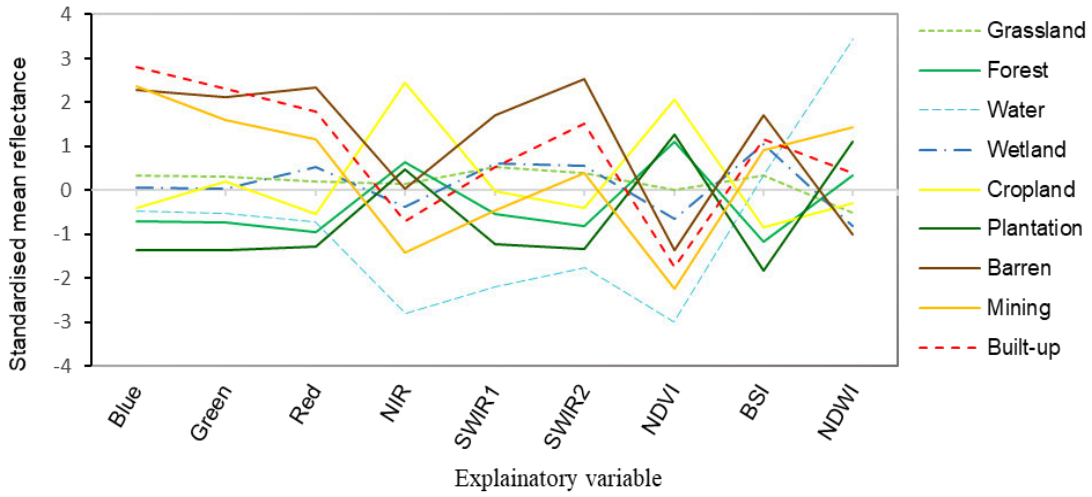

Figure 2.4.Standardised average reflectance values of each land use class.

2.3.1 Mapping and spatial extent LULC classes

This study successfully mapped nine LULC classes covering about 4444 km^2 of the uMngeni River catchment using four MLA strategies (NB, RF, SVM, and ANN). The classification results are shown in **Figure 2.5**, and the graphical representation is in **Figure 2.6**. The MLAs

show estimates of varying extent but agree to class geographic position in the study area. The catchment is dominated by grassland vegetation, covering 1819.31 km^2 being RF and 1206.16 km^2 ANN. Natural Forest is the second dominant class, with an area cover of 1200.13 km^2 determined by NB and 762.2 km^2 ANN. This is followed by planted forest, with an estimate of 738.9 km^2 and 296.5 km^2 by NB.

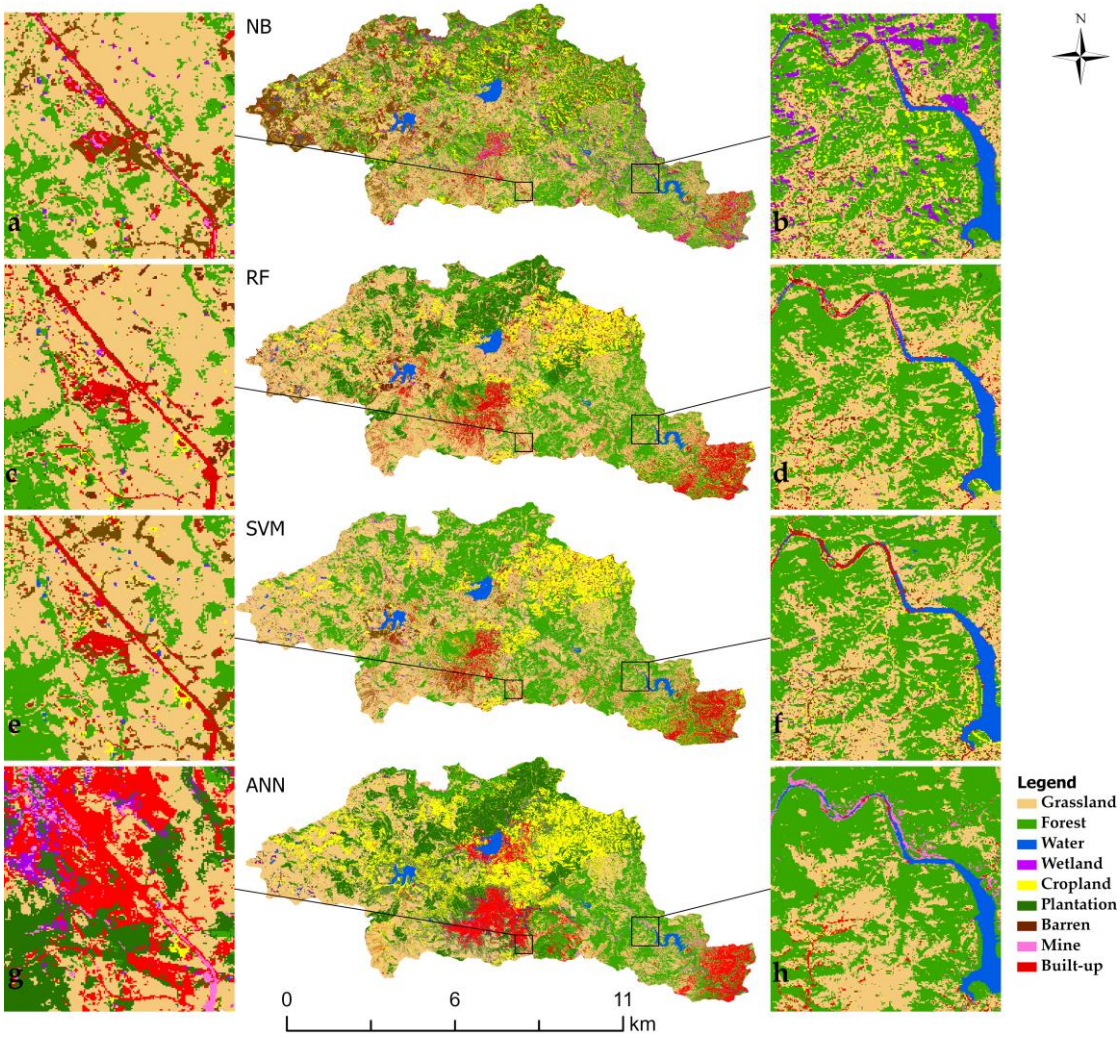

Figure 2.5.UMngeni River catchment LULC maps produced by NB, RF, SVM, and ANN models with detailed small portions in (a) to (h).

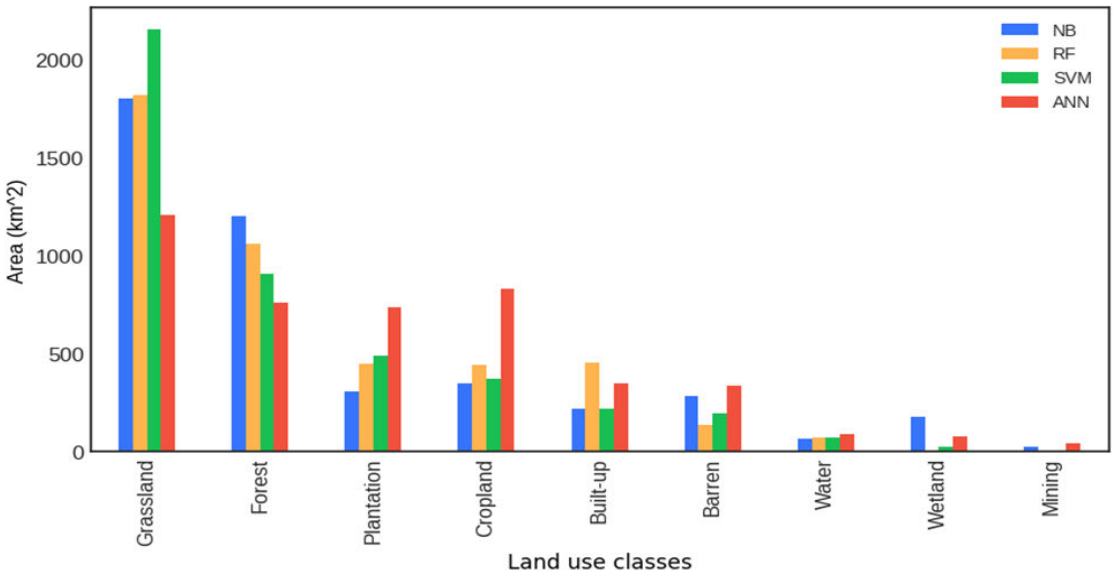

Figure 2.6. The spatial distribution of land use classes produced by each MLA.

Plantation, as the third dominant class, shows slight variations of area cover in NB, SVM, and RF (**Figure 2.6**). Except for distinct area coverage of 738.88 km^2 recorded by the ANN model. ANN showed the significant area extent of 831.27 km^2 in the cropland class. In contrast, RF and SVM reported a statistical insignificant for cropland class cover at 372.73 km^2 and 440.29 km^2, respectively. Built-up and bare soil area cover vary across the four MLAs, ranking them in the fifth and sixth dominant classes. The Water, wetland, and mining classes showed the least area coverage. In this category, the mining class represents the least coverage of the total study area, with area cover at a maximum of 45.04 km^2 determined by ANN (**Figure 2.6**).

Table 2.6. Land use class area coverage from model RF.

LULC CLASS	Area in km^2	Area in %
Grassland	1819.31	40.94
Forest	1058.55	23.82
Built-up	455.69	10.25
Cropland	446.49	10.05
Plantation	440.29	9.91
Barren	134.68	3.03
Water	74.81	1.68
Wetland	5.29	0.12
Mining	3.35	0.08
Total	4444.00	100.00

[1] These are the results (km^2 and percentage) for class area cover that was used in the discussion section since RF produced higher accurate maps than SVM, NB, and ANN.

The geographic positions of thematic land use classes are consistent across four MLAs (**Figure 2.5**). In the matrix of grassland, as it dominates the catchment by 40.99%, the cropland and barren land are observed in the proximal distance from the water bodies and in the periphery of the upper parts of the catchment. Meanwhile, natural forest land cover class cover a large portion of the middle reaches extending to the lower regions, with about 0.08 % of mining classes on the forest banks. The plantation is scattered around the upper region with higher density in the far north of the upper region. The distribution of 10.27% (**Table 2.6**) of the built-up class is observed mainly in the lower region of the study area with dense areas towards the uMngeni river estuary (**Figure 2.6**). The wetland class covered tiny patches (0.12%) in the dam shorelines and visible at the tip of the upper region.

2.3.2 Comparison of Machine Learning Algorithms Mapping Accuracy

The accuracy matrices of the models were generated using 30% of the sample size (**Table 2.4**). The matrices of the MLAs (**Figure 2.7**) provide the statistical report on the agreement between reality on the ground and the maps produced. The OA, Kappa, UA, and PA were used to assess the accuracy of the models. According to the confusion matrix reports, the OA and Kappa were 64.38% and 0.58 for NB, 97.02% and 0.96 for RF, 89.74% and 0.88 for SVM, and 87% and 0.86 for the ANN, respectively. The overall performance of the models was generated from the accuracy of the individual classes.

The classification accuracy for the land use classes varies significantly across the four MLAs as presented in **Figure 2.7**. RF and SVM have well-classified the built-up class at the threshold of >85.60 % for both UA and PA. The NB model struggled to separate built-up and mining classes. Hence, there is observable swapping between the two classes. Similarly, the ANN model showed a high commission of mining pixels to built-up. Thus, NB only achieved 48.54% and 49.39% for UA and PA for built-up, respectively. NB tends to confuse cropland with plantations. In contrast, RF, SVM, and ANN successfully extracted the cropland pixels with high percentages of metrics (**Figure 2.7**). All the MLAs delineated the surface water, plantation, forest, and barren with fairly high confidence. Except for the NB model that commissioned some natural forest pixels to the plantations class.

It is worth reporting that only the RF model provided meaningful results for mapping uMngeni River catchment classes with a small area extent, which includes barren, mining, and wetland classes. The SVM model experienced an error in omission in mining and wetland pixels. In contrast, the ANN model mapped mining and wetlands classes with observable overfitting, which highly correlates with the decline of built-up, grassland, and forest classes (**Figure 2.5**). NB significantly overfitted the barren and wetland classes at the expense of cropland and plantation land cover classes. Consequently, these accuracy metrics form the basis of the overall classification performance of each model.

NB

Predicted classes \ Reference data	1	2	3	4	5	6	7	8	9	Sum row	PA%
1	1797	140	0	14	50	0	328	97	157	2583	**69.57**
2	50	1132	0	409	582	61	0	0	0	2234	**50.67**
3	0	0	694	2	0	1	0	0	0	697	**99.57**
4	8	9	1	103	0	0	2	8	0	131	**78.63**
5	93	342	0	6	2429	400	11	1	2	3284	**73.96**
6	18	629	0	3	595	1585	27	0	2	2859	**55.44**
7	244	11	0	0	0	0	1171	3	239	1668	**70.20**
8	5	0	3	73	0	0	9	71	76	237	**29.96**
9	56	3	2	85	0	0	62	252	449	909	**49.39**
Sum colum	2271	2266	700	695	3656	2047	1610	432	925		
UA	**79.13**	**49.96**	**99.14**	**14.82**	**66.44**	**77.43**	**72.73**	**16.44**	**48.54**		

OA% 64.10
Kappa 0.58

RF

Predicted classes \ Reference data	1	2	3	4	5	6	7	8	9	Sum row	PA%
1	2554	9	0	0	7	0	13	0	0	2583	**98.88**
2	22	2093	0	0	98	21	0	0	0	2234	**93.69**
3	0	0	696	0	1	0	0	0	0	697	**99.86**
4	8	1	0	121	1	0	0	0	0	131	**92.37**
5	60	10	0	0	3210	0	4	0	0	3284	**97.75**
6	11	37	1	0	1	2801	8	0	0	2859	**97.97**
7	64	0	0	0	1	0	1597	0	6	1662	**96.09**
8	5	0	3	4	0	0	1	133	91	237	**56.12**
9	4	0	0	0	0	0	7	0	898	909	**98.79**
Sum column	2728	2150	700	125	3319	2822	1630	133	995		
UA %	**93.62**	**97.35**	**99.43**	**96.80**	**96.72**	**99.26**	**97.98**	**100.00**	**90.25**		

OA% 97.02
Kappa 0.96

SVM

Predicted classes \ Reference data	1	2	3	4	5	6	7	8	9	Sum row	PA%
1	2330	54	0	3	24	3	156	11	2	2583	**90.21**
2	75	1932	0	0	186	38	2	0	1	2234	**86.48**
3	0	0	697	0	0	0	0	0	0	697	**100.00**
4	19	1	3	96	1	0	0	11	0	131	**73.28**
5	134	52	0	6	3080	0	11	0	1	3284	**93.79**
6	16	86	0	1	8	2737	11	0	0	2859	**95.73**
7	297	0	0	0	4	0	1289	0	78	1668	**77.28**
8	17	2	3	18	0	0	24	47	126	237	**19.83**
9	24	1	0	5	0	0	60	27	792	909	**87.13**
Sum colum	2912	2128	703	129	3303	2778	1553	96	1000		
UA'	**80.01**	**90.79**	**99.15**	**74.42**	**93.25**	**98.52**	**83.00**	**48.96**	**79.20**		

OA% 89.18
Kappa 0.87

ANN

Predicted classes \ Reference data	1	2	3	4	5	6	7	8	9	Sum row	PA%
1	2100	111	0	8	285	2	2	68	7	2583	**81.30**
2	119	190	0	2	37	166	0	10	0	524	**36.26**
3	0	0	685	0	0	0	0	11	1	697	**98.28**
4	24	6	0	1011	0	0	0	0	0	1041	**97.12**
5	184	13	0	0	3051	0	36	0	0	3284	**92.90**
6	0	24	0	0	6	2820	9	0	0	2859	**98.64**
7	167	0	0	0	38	0	1258	0	205	1668	**75.42**
8	17	0	0	0	0	0	0	74	146	237	**31.22**
9	45	0	0	0	2	0	1	73	788	909	**86.69**
Sum column	2656	344	685	1021	3419	2988	1306	236	1147		
UA %	**79.07**	**55.23**	**100.00**	**99.02**	**89.24**	**94.38**	**96.32**	**31.36**	**68.70**		

OA% 87.50
Kappa 0.86

Figure 2.7. Confusion matrix and accuracy metrics for NB, RF, SVM, and ANN.

2.3.3 Model Feature Importance

Land cover classes were mapped through the contribution of L8-OLI multispectral bands and derived indices (Figures 2 and 3). The impact of each input feature on labelling classes was analysed based on the average SHAP absolute value. Figure 2.8 shows the feature importance bar in the degressive arrangement. Each bar represents the sum of the average SHAP values of classes per feature. In the y-axis, the longer the bar, the higher the importance of the feature, and the X-axis represents the average SHAP values for each feature.

In Figure 2.8, the model's interpretation graphs depict the two common patterns across four models. The top four contributors include SWIR (B6 and B7), Blue (B2), and Green (B3) bands, and the second half shows the gradual decreasing mean SHAP value trend from BSI, NDWI, Red (B4), and NDVI. Given that our study area is dominated by vegetation land use classes, which explains that the highly ranked features are sensitive to vegetation. Thus, SWIR bands were ranked as the most important feature in overall. They are centered at 1.56-2.29 µm, allowing the detection of water content among vegetation class members (Lukeš et al., 2013). Similar results have been reported in studies based on assessing Landsat features for mapping vegetation foliage, such as (Schultz et al., 2015; Sothe et al., 2017). This has also been observed in other studies that used different sensors (Sentinel-2) and found the dominance of SWIR over the other features (Abdi, 2020; Chen, Yang, et al., 2023; McCarty et al., 2020).

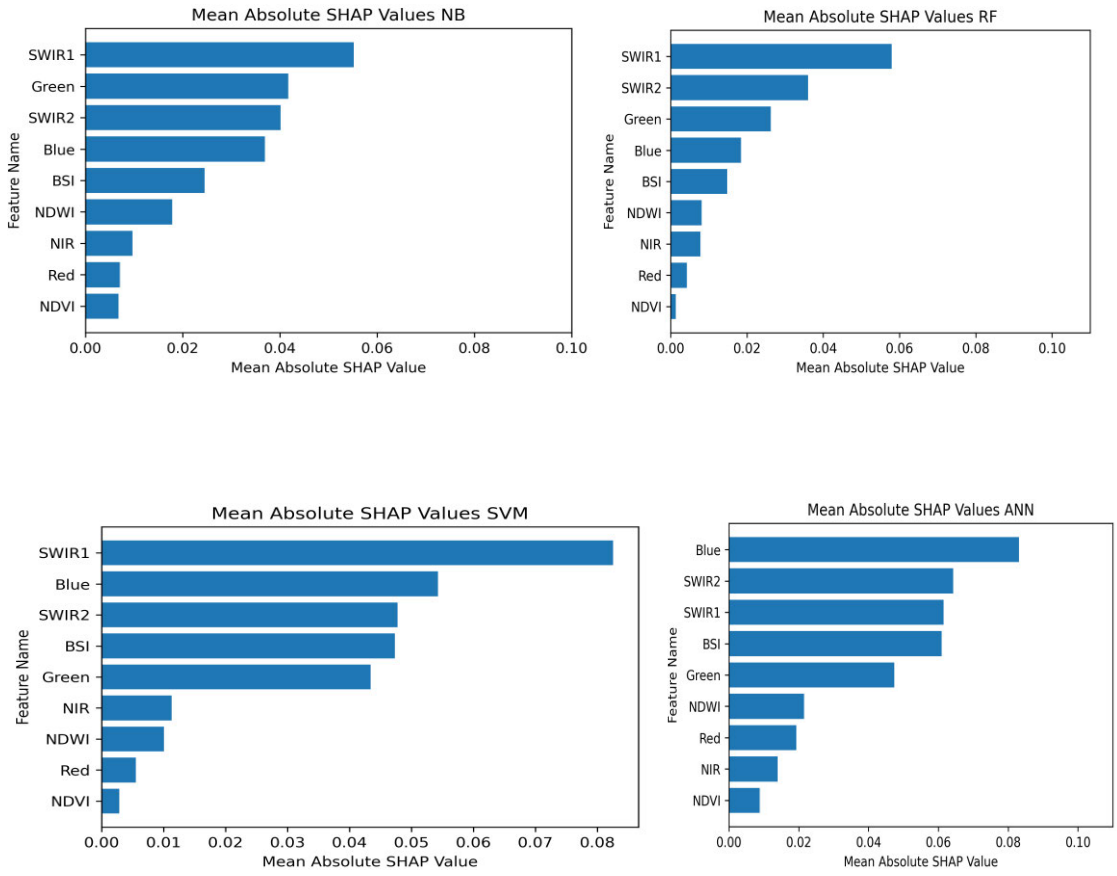

Figure 2.8. Model feature importance based on the mean absolute SHAP value.

2.4 Discussion

The mapping of the nine land use classes in GEE was accomplished by complementing the four non-parametric MLAs (NB, RF, SVM, and ANN). To examine the LULC's current state and the geographic distribution of each class in the uMngeni River basin. The results of this analysis will be useful in determining which planning instruments are best for protecting fresh-water resources. It also sheds light on how MLAs are applied and how well they may explain the LULC in the region of undulating terrain.

Thematic land use classes were classified based on the use of the multi-spectral bands from L8-OLI imagery. It is imperative to underscore the variations in area coverage for land cover classes amongst the MLAs (Figure 2.6); these results align with comparative research (Ashiagbor et al., 2023; Loukika et al., 2021; Woldemariam et al., 2022) that examined machine learning classifiers in mapping LULC classes. This reflects the variability in the inherent design of the machine learning models (Ashiagbor et al., 2023; Foody, 2015). and the dimensionality of the data to be processed (Camargo et al., 2019). Thus, results from the RF model as the best-performed classifier were used in the discussion of this work.

The grassland class was identified as the most dominant class. It is well distributed, with an area extent of 40.99% of the total catchment area. (Namugize et al., 2018) reported similar findings. The grassland class holds the topsoil together, increasing the water infiltration rate (Machado et al., 2019). Thus has a significant impact on groundwater recharge and water quality.

The forest provides similar ecological functions as the grassland; with an average coverage of 23.85% of the catchment area, the forest is the second largest class after grassland. It is found in abundance in the central sections and in somewhat dispersed patches in the upper and lower regions. Most notably, it mitigates climate change and acts as a carbon sink, hence preventing the deterioration of water quality (Abbas et al., 2023; Gyamfi-Ampadu et al., 2020). As a result, efforts to preserve the current natural plant cover should be undertaken, and as suggested by (Gyamfi-Ampadu et al., 2020, 2022; Plessis et al., 2014), actions aimed at increasing the area covered by these classes may be encouraged.

The cropland is the fifth largest class, accounting for 10.06% of the study area. Although it is evenly dispersed throughout the catchment, homogeneous patches are observed in the north-east and stretch towards the tip of the western regions (**Figure 2.5**). Adjacent to cropland in the further north, about 9.92% of the plantation area cover was observed. The areas dominated by croplands and plantations are known for having low water quality (Mararakanye et al., 2022; Namugize et al., 2018) due to the inputs used to enhance plant life. The residuals remain in the soil and are eventually channelled to the waterbodies, impacting the freshwater ecosystem (Mararakanye et al., 2022; Plessis et al., 2014). In light of these environmental risks, water resource managers must ensure control measures are in place to promote sustainable agriculture and forestry.

Among the classes that account for the bigger share of the catchment area, the built-up class is the last at 10.27%. It is mostly concentrated in the southern and lower areas of the catchment. Driven by societal needs (Gyamfi-Ampadu et al., 2020; Namugize et al., 2018; Obaid et al., 2023), it comprises a range of infrastructure types, as explained in **Table 2.1.** The built-up class creates impermeable surfaces with low water infiltration, raising water run-off (Schütte and Schulze, 2017), hiking of the non-point pollutants to the rivers (Hua, 2017; Kibena et al., 2014; Mararakanye et al., 2022; Namugize et al., 2018), and ultimately increasing the stream rate flow (Tena et al., 2019) with implications for water quality and hydrological structures. For such reasons, practical measures to control urban sprawling must be developed, as numerous studies have suggested (Abbas et al., 2023; Jewitt et al., 2015).

Similar to built-up areas, barren areas and mining pose a hazard to freshwater resources; they occupy comparatively minor portions of the watershed, 3.03% and 0.08%, respectively (**Table 2.6**). Water turbidity is caused by barren, which permits dirt particles to splash into waterbodies (Twesigye, 2011). The mining spots are somewhat randomly dispersed along the catchment's forest banks. Their byproducts endanger human health as well as the aquatic biodiversity (Hua, 2017). In order to minimize the influence on waterbodies, it is advised that mining developments and barren areas within the catchment be constantly monitored and subject to management measures.

Water and wetland classes cover 1.69% and 0.12% of the study area, respectively. The wetlands are found in the north-west periphery, similar findings were also reported by(Moodley et al., 2023). Both water and wetlands are the major role players in ensuring the health of the ecosystem in general (Nayak and Bhushan, 2022). Therefore, to guarantee the continuity of their service, initiatives aimed at water harvesting and natural restoration systems should be established, as recommended by (Schütte and Schulze, 2017).

The quantitative analysis of the LULC in the catchment unpacked the underlying data regarding the LULC distribution status. This data will serve as a reference for freshwater resource managers and environmental health specialists in developing management plans and policies for freshwater resources based on informed decisions.

Since various models yield different results in different classes, it is imperative to evaluate the MLAs' performance at the class level(Palanisamy et al., 2023). In light of this, only the ANN model confidently delineated the river paths (**Figure 2.5h**). However, could not detect small dams and ponds. In contrast, RF, SVM, and NB struggled to distinguish between river trails and built-up (**Figure 2.5b, d and f**). This has been reported in prior studies that used pixel-based methods in LULC mapping (Loukika et al., 2021; Nayak and Bhushan, 2022; Obaid et al., 2023) and linked to spectral overlap between built-up and riverbeds or river rocks. A strong ability to map small waterbodies demonstrated by the RF and SVM. Similar results have been reported by (Kang et al., 2021). Highlighting the SVM efficiency in mapping small waterbodies. But performed poorly in defining mining pits. On the other hand, NB and ANN showed bidirectional misclassification between of the built-up and mining classes. These findings are in line with (Camargo et al., 2019; Chen, Lin, et al., 2023). While RF succeeded in separating these classes with a high user accuracy score. Similar results have been reported (Chen, Lin, et al., 2023; Dahiya et al., 2023; Orieschnig et al., 2021), claiming that RF possesses a strong generalization technique for LULC analysis, making it less sensitive to the small variations in input vector data.

It is also worth reporting that all MLAs struggled to delineate the wetland areas except the RF model. NB and ANN tend to overfit the wetland class, while SVM commissioned some of the wetland pixels to grassland. These findings were observed in other studies, (Dash et al., 2023)

suggesting that this can be associated with composite signatures from the mixed pixels caused by the L8-OLI 30 m resolution (Campbell and Wynne, 2011) and small samples allocated to the wetland class due to its small size. The ability to compile the MLAs' overall performance is made possible by the precise evaluation at the class level.

Overall, all the models performed well and produced more realistic LULC maps that resemble the reality on the ground. The findings of this study indicate that RF outperformed all the MLAs with slightly higher OA and Kappa values from those recorded by the SVM model as the second-best classifier, yielding higher accurate results than ANN and NB. Previous studies (Ashiagbor et al., 2023; Loukika et al., 2021; Ouchra et al., 2023; Parashar et al., 2024) have debated the comparison of RF and SVM and concluded that RF is resilient in classifying complex LULC classes. Its success is credited to its insensitivity to sample size and the advantage of the random multi-decision trees in generalization—which are immune to outliers. In contrast, other studies found SVM to be superior to RF; for example, (Gyamfi-Ampadu et al., 2020) found SVM to outperform the RF class in mapping natural forest cover classes, indicating that the model's success can be attributed to its default settings. Woldemariam et al. (2022) evaluated MLAs for dynamic land use, found SVM to be superior to RF and credited the model's performance to the usage of the kernel functions to optimize the hyperplanes to find the optimum margin.

An emerging MLA in the remote sensing field, NB, was found unfit for analysis of the complex land use landscapes. Other researchers (Camargo et al., 2019; Orieschnig et al., 2021; Palanisamy et al., 2023) also reported similar findings. This highlights the limitations in the inherent design of the model in dealing with high-dimensional data. The performance can also be associated with the sensitivity of the model to sample size. The NB classifier can be effective in large areas with more linear land cover patterns.

The results of our study proved the excellence of the RF model, which produced a more accurate LULC analysis of the uMngeni River catchment. These results will fill the gap in the literature by identifying the suitable MLA for quantitative analysis of dimensional LULC in undulating watershed areas. In addition, this can be a useful tool for water resource managers and environmental health stakeholders to facilitate the development of freshwater resource

management strategies by unpacking the visual and qualitative LULC information for the watershed areas.

The L8-OLI multi-spectral bands with integrated indices provided good results. SWIR (1 and 2) emerged as the most influential band in the delineation of LULC classes, as discussed in section 3.3. For such reasons, L8-OLI data was found to be reliable for analysis of the complex LULC in the undulating landscape of the catchment areas. However, it was found to be challenging to classify the pixels in the edges of the big parcels of class and built-up class members in the core of barren and/or grassland or partially masked by trees. These findings are consistent with (Dash et al., 2023), who found that pixels in the transition zone between two parcels have been misclassified due to mixed-pixel composite signatures, as explained by (Campbell and Wynne, 2011; Lillesand et al., 2015). It is challenging to discriminate smaller classes from others in L8-OLI 30 m resolution images due to mixed pixels (Loukika et al., 2021). Other studies factored in the limitations of pixel-based algorithms, (Maviza and Ahmed, 2020; Nayak and Bhushan, 2022) associated mixed pixel effects with the restrictions of pixel-based models to go beyond the pixel level.

2.5 Conclusion

The LULC thematic maps production based on moderate-resolution satellite data is commonly used in various land cover mapping studies. Using such data gives environmental managers the timely quantitative spatial information they need to implement best practices for natural resource conservation. In this study, four pixel-based non-parametric MLAs—NB, RF, SVM, and ANN were attested in the classification of the thematic LULC class in the uMngeni River Catchment landscape, South Africa. Moderate resolution (30 m) from the L8-OLI sensor was harnessed for mapping nine identified LULC classes. The referenced data acquired from field surveys and Google Earth Pro high-resolution imagery was used for both training models and accuracy assessment. The findings of this study demonstrated the disparities in the MLAs' performances at the class level, resulting in variations in LULC class area coverage across four MLAs. Overall, the RF model outperformed all other models, producing a map that closely resembles the reality on the ground with 97.02% overall accuracy and 0.96 kappa value and in line with number of studies on mapping LULC in the catchment areas. We discovered that the ANN model performed well in defining the river trails compared to other models. While RF proved to be efficient in mapping classes with tiny area sizes. When viewed from the viewpoint of other models, SVM tends to experience the commission of mining class members to build

up, hence producing low accuracies in the mining class. As evidenced by their low accuracies, NB and ANN either overfitted or underestimated the LULC classes covered a small area. These errors stem from two directions: first, the variations in the designs of the models hinder them from analysing the data at the subpixel level; second, the 30 m resolution data used makes it difficult for MLA models to classify the impure Landsat pixels due to low resolution. The findings of this work will demonstrate the practical application of MLAs in mapping LULC and serve as an anchor for water resource managers and other relevant environmental health professionals to update or establish the policies and management plans aimed at conserving water resources based on informed decisions. In addition, RF for classification with balanced hyperparameter tuning and Landsat data used in this study proved to be efficient in mapping LULC in the Umngeni River Catchment. As a result, the method used in this study can be adapted to other research in similar geographic settings.

Author Contributions: Conceptualization, O.B., A.R. and M.G.; Methodology, O.B., A.R. and M.G.; Formal analysis, O.B. and M.G.; Supervision, M.G.; Writing - original draft, O.B.; Writing - review and editing, O.B., A.R. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the South African National Space Agency (SANSA) (www.sansa.org.za).

Data Availability Statement: The surface reflectance Landsat 8 data can be accessed in Earth Engine data catalog at USGS Landsat 8 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog | Google for Developers.

Conflicts of Interest: The authors declare no conflicts of interest.

3 CHAPTER 3: LULC CHANGE DETECTION AND FUTURE LULC MODELLING USING RF AND MLPNN-MARKOV ALGORITHMS IN THE UMNGENI CATCHMENT, KWAZULU-NATAL, SOUTH AFRICA

Abstract: Water catchment areas are the key strategic water sources with a variety of ecological benefits. However, the trajectory of Land Cover and Land Use Changes (LULC-C) poses a significant threat to water catchment areas, negatively affecting water quality. Thus, the adoption of remote sensing data and Machine Learning Algorithms (MLAs) is a novel approach that provides spatiotemporal data on the environmental changes resulting from LULC dynamics. Hence, this work harnessed Landsat imageries and the Random Forests (RF) classification as well as a hybrid model from the Multi-Layer Perceptron and Markov chain (MLPNN-Markov) to detect changes in LULC and forecast future changes. At every five years interval, the RF model generated more accurate maps for 2003–2023. The LULC prediction for 2019 also produced acceptable values for the kappa accuracy matrices, which were 65.50%, 58.4%, 90.90%, and 0.52 for overall accuracy, kappa location, kappa histogram, and kappa overall, respectively. The findings highlighted the decline of forest areas, with a strong negative correlation with built-up and mining areas. The secondary invasion of the abandoned cropland occupied by grassland members was observed. Thus, grassland displayed increasing trends between 2019 and 2023. Wetlands and water, however, exhibited a steady trend with minor variations. On the other hand, each of these trends persisted in the future, with the exception of grassland areas that displayed scaling-down behaviour in 2032. The outcomes of this work will offer a piece of updated information on the LULC-C and hints at the possible future direction for the trends by 2032. This is crucial to local bodies tasked to protect the integrity of the water catchment areas with the aim of improving the water quality.

Keywords: Random Forest; LULC; Landsat; change detection; uMngeni River Catchment; South Africa; MLPNN-Markov

3.1 Introduction

The biophysical attributes of river catchments are important for the health of terrestrial and freshwater ecosystems (Jewitt et al., 2015) and for mitigating the effects of climate change

(Abbas et al., 2023; Gyamfi-Ampadu et al., 2021; Mugo et al., 2020). In addition, the catchments provide essential social demands, such as water supply, food, settlement, and recreational opportunities (Kaval, 2019; Mugo et al., 2020; Pullanikkatil et al., 2016). The provision of these services relies on the good health of the ecosystem, physical components such as the river network, and other land resources necessary for the supply of these services. Catchment ecosystems such as that of the uMngeni River Catchment are vulnerable to both natural disasters brought on by climate change and human intrusion (Ruppert et al., 2015).

In view of the health status of river catchment ecosystems, the interference of anthropogenic activities and unprecedented natural disasters are considered the driving forces behind the land use and land cover changes (LULC-C) in the catchment areas worldwide (Floreano and de Moraes, 2021; Stosch et al., 2017). Specifically, anthropogenic activities, including settlements, cultivation, timber, and mineral extraction, are among the factors that influence the modifications of catchment landscapes (Pullanikkatil et al., 2016). These catchment land cover modifications are linked to various implications such as water quality, groundwater recharge, wildlife habitat, climate change mitigation, soil conservation, and other catchment ecosystem benefits (Floreano and de Moraes, 2021; Indraja et al., 2024; Mararakanye et al., 2022; Mugo et al., 2020; Pullanikkatil et al., 2016; Yao et al., 2016). According to Mugo et al. (2020), the first step in addressing these environmental issues should focus on understanding the land cover trends and configuration. Therefore, research on LULC monitoring and remote sensing-based forecasting is, therefore, considered a viable approach (Schoeman et al., 2013).

Remote sensing technologies with rich data infrastructures have been the main data source for land cover monitoring studies (Schoeman et al., 2013). Remote sensing provides extensive archives with time-stamped spatiotemporal data, which is crucial for detecting land cover changes in the catchments (Campbell and Wynne, 2011). The LULC-C analysis approaches vary as there is no standard rule for selecting a method (Woldemariam et al., 2022). However, method selection can be influenced by the nature of the remote sensing data, the physical properties of the study area, and the objectives of the study. Consequently, a range of methods in the literature have been employed for land cover change detection. Machine learning, deep learning, Multi-Layer Perceptron Neural Networks (MLPNN), Markov chain and Cellular Automata (CA) are among the prominent methods frequently used in the published work on land cover change detection and prediction.

Mashala et al. (2023) used Landsat series data to analyze the historical LULC in the Letaba watershed in South Africa from 1990 to 2021. The study achieved an overall classification

accuracy of 93% to 99% across the study periods and found a sharp increase in built-up and shrubs by 47% and 41%, respectively. These increases were at the expense of grassland and indigenous forests. In a similar study, Obaid et al. (2023) assessed spatiotemporal LULC-C between 1986 and 2021 in Vaal Dam Catchment, South Africa, based on Thematic Mapper (L5-TM), Enhanced Thematic Mapper (L7-ETM+), and Operational Land Imager (L8-OLI) with a range overall accuracy of 87% to 95% and kappa coefficient ranging from 0.79 to 0.92. They observed a steady expansion of built-up areas from 0.87% to 1.43% during the study period at the cost of woody vegetation. Another study carried out by Mugo et al. (2020), assessed the land cover changes in the Lake Victoria Basin from 1985 to 2014 based on Landsat series data. They reported an 800% increase in built-up and a 90% increase in small farm areas. A study by Aghsaei et al. (2020) used L5-TM and L7-ETM+ images to assess the dynamics of the LULC-C in the Anzali wetland catchment, Iraq, between 1990 and 2013. The study revealed the complete degradation of the wetlands, which initially accounted for 2.5% of the catchment area. These losses coincided with the rise of the cropland and the built-up areas. Namugize et al., (2018) conducted a time series analysis of LULC in the Upper uMngeni River catchment, South Africa, between 1994 and 2011. The study highlighted the expansion of both built-up and agricultural areas by 6% and 4.5%, respectively, which came at the expense of a 17 % loss of vegetation cover during the study period.

Research on land use change detection in the catchment areas is of great importance, not only for monitoring the catchment land cover spatial extents but also as a tool for predicting future LULC scenarios. Studies on the LULC prediction have adopted the hybrid algorithms, to model future land use changes. The primary aim of adopting hybrid algorithms is to compensate for the shortfalls of the land cover modelling algorithms, as each model has its advantages and drawbacks (Hafeez et al., 2019; Loukika et al., 2021). For instance, (Moodley et al., 2023) used L5-TM and L8-OLI to create historical LULC maps for 1990 and 2018 and simulated the LULC for 2030 in both the uThukela and uMngeni catchment areas in South Africa using a hybrid model of CA and Markov Chain model (CA- Markov). The outcomes of the predicted land use classes showed slight adjustments in built-up, plantation, and cropland areas, which increased at the expense of grassland areas.

(Mandal et al., 2023) modelled the decadal future land cover changes from 2030 to 2050 in the Pare River basin, India. They used a hybrid approach of MLP and CA-Markov models. The input classifications based on the Landsat imagery data (L5-TM and OLI-L8) achieved overall

accuracies of 0.91%, 0.85%, and 0.91% with kappa coefficient values of 0.88, 0.82, and 0.89. Their projected maps showed a positive linear relationship between cropland and built-up, while forest land cover experienced a decline by 2050. (Maviza and Ahmed, 2020) used a combination of (MLPNN-Markov) to project future LULC scenarios from 2023 to 2038 in the Upper Mzingwane sub-catchment, Zimbabwe. The accuracy of input classifications from L5-TM and L8-OLI were 0.9846%, 0.7763%, and 0.9791–0.6633 for overall accuracy and Kappa coefficient, respectively. Their simulation indicated significant land cover changes in 2038 with a loss of 441 km^2 of woodland and an increase of 185 km^2 for grassland.

Hybrid models for change detection and land cover forecasting have not only been used in the case of the catchment but also experimented in different scenarios. For example, (Adam et al., 2023) modelled future urban scenarios in Maseru city, Lesotho, for 2050. The L5-TM and OLI-based thematic maps with an overall accuracy report of 88% and 95% with ANN-CA were used to produce the prediction maps. The study shows a continuous increase in built-up areas, which is expected to cover 5434.5 hectares by 2050, significantly reducing croplands, grasslands, and woody vegetation. (Abbas et al., 2023), conducted a study to project future urban status by 2030 in Islamabad City, Pakistan. Sentinel-2 images and MLPNN-Markov were used to produce prediction maps, indicating that 58.84% of Islamabad will be covered by a built-up, while other LULC classes such as forests are expected to decrease with varying rates. In different a landscape, (Floreano and de Moraes, 2021) employed the CA- Markov hybrid model and the Landsat imagery to prepare data for forest land cover forecasting in Rondônia, Brazil. The results showed that approximately 30% of the forest areas will be encroached by non-forest land use classes by 2030.

Studies examining past and future land cover in different landscapes have been seen crucial for providing valuable information for planning and policy-making to promote sustainable development. According to the evidence gathered from the literature, there are a limited number of studies that have extended historical land use analysis to predict the future land cover state of the catchment area in South Africa. Therefore, knowledge of the past and future land cover scenarios should be generated for watershed areas as they are the key strategic freshwater sources (Indraja et al., 2024). Such knowledge will serve as a critical tool for developing water catchment management strategies. This study aims to examine the spatial and temporal land use status of the uMngeni River Catchment between 2003 and 2023 at half-decade intervals and project plausible future scenarios for 2032. Our study site, the uMngeni River Catchment,

is subject to different forms of landscape modifications rising from both anthropogenic and natural factors (Mauck and Warburton, 2014; Moodley et al., 2023; Schoeman et al., 2013). Factors such as urban expansion, mining, agriculture, and unprecedented natural factors such as floods are the main threats to the uMngeni River catchment hydrological integrity. The qualitative and statistical information presented in this study will contribute to the knowledge and understanding of change rates and patterns, providing insight into likely future land cover distribution, which will serve as a foundation for the conservation of water resources. Additionally, the findings have the potential to influence national initiatives aimed at improving the water supply and protecting the catchment areas.

3.2 Methods and material

3.2.1 Study area

The current study focuses on the uMngeni River catchment as the selected region of interest. The catchment is located between (29.7156, -29.1349 E) and (31.1137, -29.6383 S) in the KwaZulu Natal (KZN), South Africa. Its estimated total surface area is 4 400 km^2, separated into three main regions (upper, middle, and lower) (Hughes et al., 2018). Its landscape exhibits the intricate network of rivers with wavy terrain in the upper riches, whereas moderate slopes and hills define the catchment's lower regions. (Figure 3.1).

The uMngeni River, the principal river formed by the network of rivers, originates from the Drakensberg and flows through Pietermaritzburg before emptying into the Indian Ocean through the Durban north estuary. This lengthy, meandering river, with major dams (Midmar, Albert Falls, Nagle, and Inanda) along its trail, supplies water for 45 % of the people living in the KZN province (Namugize et al., 2018). The local climate and watershed topography hold a high stack on the composition of the uMngeni River. The catchment is 1913 m above sea level (Kusangaya et al., 2017), with rainfall occurring during the summer seasons (October to March), ranging from 700-1550 ml and mild temperatures between 12 to 20 °C annually. These features are key factors contributing to the province's economic growth.

One of South Africa's major economic centers and main trading port is located in the catchment (Hughes et al., 2018). Accordingly, it contributes up to 65% of KZN's overall economic output. The province's Gross Domestic Products also heavily relies on industries such mining, recreation, forestry, and agriculture (Strydom et al., 2020). The catchment is distinguished by a variety of land use types as a result of the previously mentioned industrial and economic activities.

Therefore, the consequences of land cover change not only affect economic opportunities but also affect water quality, ecosystems, and climate change. The uMngeni River Catchment was selected as the study area of interest to ensure the continuation of these crucial services.

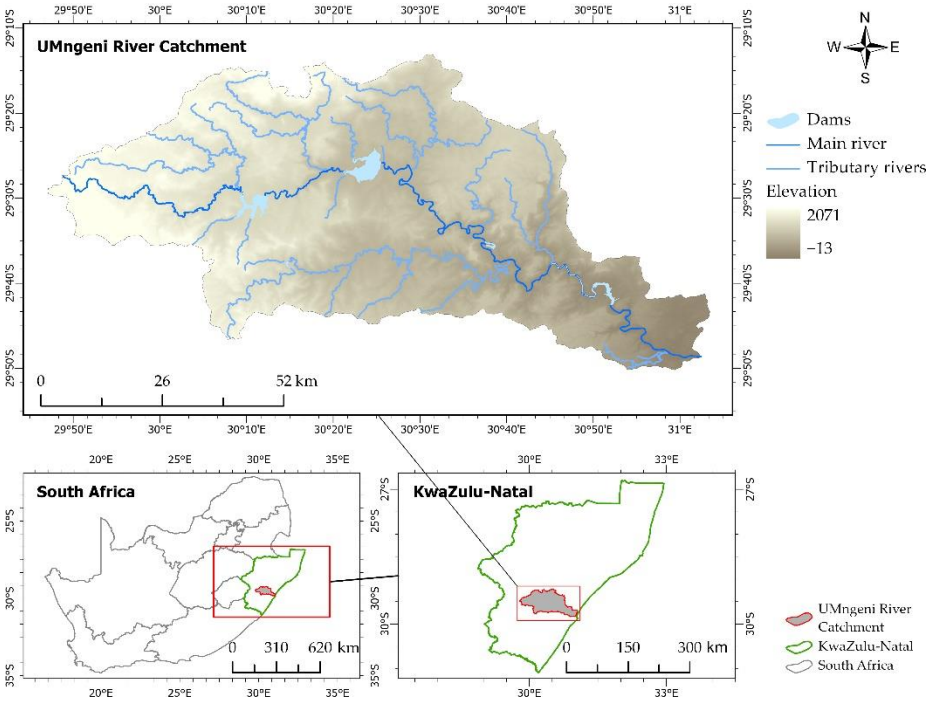

Figure 3.1.The uMngeni catchment layout with elevation gradients, river network, and dams.

According to our field observations, the catchment's land cover categories include but are not limited to, built-up areas, barren ground, and water bodies. Green patches, which include forests, plantations, and crops. The uMngeni catchment is dominated by farmland and natural vegetation (Namugize et al., 2018). One of the major land use classes that predominates in the north of the basin is the plantation. Heterogeneous patches of various land use types (plantations and built-up areas) define the middle region. While the lower part is mainly dominated by built-up patches with fragments of green vegetated land cover. It is worth unpacking the information in such diverse land cover areas to provide a piece of updated scientific information.

3.2.2 Reference data

The geographic reference data for the land use classes was collected in August 16, 2023, randomly using a portable global positioning system (GPS). High-resolution Google Earth Pro archives and historical South African National Land Cover (SANLC) (Department of Forestry, Fisheries and the Environment, 2023) were also used to gain further details on the historical data needed. The following criteria guided the design of the sampling process: sites dispersed

across the study site with respect to the area covered by each land use type. Google Earth Engine (GEE) street maps and Google Earth Pro were used to accomplish this. The polygons were digitised based on the land cover type indicated by the point with reference to a time-stamped Google Earth Pro image. The digitised polygons were further imported to the GEE platform and over-traced by locally digitised polygons with unique digital labels for each land use class. Then, the spectral properties of each reference site were extracted from the image using the *sampleRegions()* function. Two datasets were created from the reference data containing the spectral information: 70% were used for training the models, and 30% were used to interpret the model performances (Table 3.1). It is also worth noting that the sampling methods were standardized across the six study periods of the uMngeni River catchment.

Table 3.1.Training (Tr) and test (Ts) count per LULC class for six study periods (2003-2023).

	2003		**2007**		**2011**		**2015**		**2019**		**2023**	
LULC Classes	**Tr**	**Ts**	**Tr**	**Ts**	**Tr**	**Ts**	**Tr**	**Ts**	**Tr**	**Ts**	**Tr**	**Ts**
Grassland	17777	7734	20909	8944	16299	7008	24046	10372	23074	9810	20674	8728
Forest	17827	7610	19945	11088	24793	10995	29842	12795	18962	8097	16004	7003
Water	8804	3641	9083	3880	6077	2647	7210	3094	6652	2784	5744	2448
Wetland	2254	1012	2786	1244	2289	946	2293	987	1399	597	2858	1197
Cropland	22927	9617	19971	8591	17067	7393	19867	8523	13468	5848	19748	8530
Plantation	24901	10789	33561	14303	29749	12584	24129	10571	16264	6946	16750	7043
Barren	13263	5594	6977	2996	9885	4294	12610	5419	12352	5238	5232	2207
Mining	709	305	1052	452	659	271	663	294	678	318	526	215
Built-Up	11018	4702	11015	4744	11856	5221	14384	5991	1148	4769	14308	6085
Total	**119480**		**125299**		**118674**		**135044**		**93997**		**101844**	

3.2.3 Satellite image collection and Pre-processing

In the uMngeni River Catchment, LULC-C was examined for six research terms—2003, 2007, 2011, 2015, 2019, and 2023—using Landsat surface reflectance images from the level 2 category. The composite images were collected from the Earth engine cloud data hub (https://developers.google.com/earth-engine/datasets/) accessed on 04 February 2024. A total of 24 scenes were collected during the crop growing season between May and August where the crop fields are distinctive from other Land use classes (Table 3.2).

Table 3.2. Summary of satellite imagery used from 2003-2023.

Year	Acquisition date range	Mission	Sensor	Path/Row	Spectral bands	Resolution (m)
2003	01 May-30 August	Landsat-5	TM	168/080,168/081	6	30 m
2007	01 May-30 August					
2011	01 May-30 August	Landsat-5	ETM	168/080,168/081	6	
2015	01 May-30 August	Landsat-8	OLI	168/080,168/081	6	
2019	01 May-30 August					
2022	01 May-30 August					

Using a range of GEE functions and filters, the collected datasets underwent several pre-processing steps. The pre-processing of the images was done step-by-step; scenes with less than 10% cloud cover were filtered, and all pixels covered by the cloud and associated cloud shadows were eliminated using the *CFmasking* function. To restore the halo spaces left by the deleted contaminated pixels, the *focal-mean* function was applied. Next, a median composite image was created by applying the *median()* function to the cloud-free dataset tiles using the region of interest (ROI) shapefile as a guide. The subset of the region of interest was then clipped from the median composite image. It is also crucial to mention that the Landsat 5 image's horizontal stripes were corrected using the focal-mean method. As explanatory variables, the Normalized difference indexes for vegetation, water, and bare soil were calculated and added to the image.

3.2.4 Land use classification and accuracy assessment

The GEE cloud computing platform offers a range of supervised Machine Learning algorithms (MLAs) appropriate for land use analysis. Naïve Bayes (NB), Random Forests (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN) are some examples of these MLAs. The abilities of these MLAs were attested in mapping the LULC within the uMngeni River catchment (Bhungeni et al., 2024). The RF model performed better than the other algorithms, according to a comparison examination. Numerous MLA comparative research employing RF models in watershed areas have revealed similar outcomes. For instance, Loukika et al. (2021) proved the dominance of the RF model over CART and SVM in their MLA comparative study in the Munneru River Basin, India. Balha et al., (2021) reported similar findings

in their study conducted in Delhi, India. With this evidence, the RF algorithm was chosen as the suitable MLA to map the historical LULC-C with the uMngeni River catchment.

The RF classifier is a supervised learning algorithm based on the decision tree model, proposed by Breiman in 2001. It constructs an ensemble of decision trees using random subsets of the dataset, created through a process known as bagging (Breiman, 2001). The bagging process generates multiple random sets of samples from the original reference data (also known as bootstrapping). Then multiple decision trees are formed from each set of samples. Each decision tree serves as the electorate, which aids in allocating the unlabelled pixels to the appropriate class (Obaid et al., 2023). The majority results of decision trees is used to determine which class to allocate the pixels to (Mather and Koch, 2011; Zhang et al., 2023). The RF models are known for their capabilities to handle complex datasets, have minimal processing time (Bayas et al., 2022), and are insensitive to outliers and overfitting (Cuypers et al., 2023). The choice of explanatory variables and parameter adjustments is deemed to have a major impact on the model's performance.

The two primary parameters of the RF models are the number of explanatory variables per split (*Mtry*) and the number of decision trees to be constructed (*Ntree*). The "*smileRandomForest"* classifier from the GEE library was turned as follows: *Ntree was* equal to 25, *Mtry* was set to be equal to the square root of variables and each node was adjusted to have 2 leaves. Six multispectral bands (visible, blue, green, and red), NIR, and SWIR (1 and 2) and three additional spectral indices (NDVI, NWDI, and BSI) were among the explanatory features used for the LULC classification for 2003, 2007, 2011, 2015, 2019, and 2023. This combination of predictor variables has been adopted in many studies and proven to provide accurate results (Bhaga et al., 2023; Liu et al., 2022; Zha et al., 2003).

For each research period, we evaluated how well the generated maps corresponded to reality. For the years 2003, 2011, 2015, 2019, and 2023, a subset of 30% of the sample size was used to test the accuracy of the classification results. Using the *errorMatrix* function in GEE, the test dataset was employed to calculate the confusion matrix and determine the assessment metrics. The assessment metrics used in this study include user's accuracy (UA), producer accuracy (PA), overall accuracy (OA), and Kappa coefficient (Congalton, 1991)

3.2.5 Land use change detection

Analysis of land cover changes at five-year intervals from 2003-2007, 2007-2011, 2011-2015, 2015-2019, and 2019-2023 was conducted using the Land Change Modeler (LCM), a built-in spatial analysis modeler in TerrSet version 2020 software.

3.2.6 Predictor variable selection and transition sub-model development

A given geographic area is not universally suitable for all land use activities. Rather, environmental appropriateness (texture and climate) and social factors (infrastructure developments and legislative limits) dictate the occurrence of land use activity. These indicators are deemed to be effective factors in simulating future land cover scenarios, as they detect the suitability of the different land cover types (Indraja et al., 2024). For these reasons, the Digital Elevation Model (DEM), aspect, slope, and distance from rivers and roads were sourced in the GEE data catalogue and incorporated in Land Change Modeler (LCM) as predictor variables. These predictor variables were imported into ArcGIS Pro 3.0.0 along with time series classifications to ensure the data have similar properties. Once the necessary preprocessing steps were fulfilled, the data was imported to LCM in the TerrSet software to form the part of sub-models. In the transition sub-model tool within LCM, the sub-models were defined based on data from two successive land cover classification periods (t_1 and t_2) and environmental predictor indicators.

3.2.7 Transition potential data

The machine learning technique, MLPNN, was leveraged to compute transition potential maps for each of the defined sub-models. The MLPNN is the supervised prominent learning method that proved to effectively handle complex land cover modelling analysis (Ahmed and Ahmed, 2012). The MLPNN model comprises of input layer, a hidden layer, and an output layer. The input layer has neurons that serve as receptor sites for the data from explanatory variables (Clark Labs (2016). Through channels, they transfer the data to the intermediate phase. Each channel has a numeric value assigned, known as weight, which is a factor of the inputs to be transmitted to the neurons of the intermediate phase (Dahiya et al., 2023). The activation function in the neurons facilitates the non-linearity of the data and identifies the neurons that will contribute to the inputs of the subsequent phase (Wu et al., 2014); all the activated neurons transmit the inputs to the next phase; the process is called forward propagation. The inputs were propagated through the network to the output layer, where the data was classified based on the likelihood of change (Chen, Lin, et al., 2023). The MLPNN model's optimization for transition

potentials was guided by the MLP model, utilizing the default parameters recommended by Clark Labs (2016).

The Model of 7 hidden layers driven by the sigmoid activation function was used to generate transition maps for each class in the sub-model based on 50% of 300 random samples and the other half reserved for validation, and 10000 iterations were set. The produced transition maps are incorporated as the predictor variables in the change prediction phase.

3.2.8 Transition probability matrix

The Markov Chain Model (Markov) model proposed by Burnham (Burnham, 1973) feeds from the transition potential maps to dictate the likelihood of LULC $class_i$ pixel change to $class_j$ between two periods (producing two matrices such as transition matrix and transition the probability matrix). The transition areas matrix summarizes pixels of one class to change to all classes, and the probability matrix specifies the transition between two classes (Burnham, 1973). The transition probability matrix is defined in the equation 7

$$P_{ij} = \begin{Vmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{Vmatrix}, \left(0 \le p_{ij} \le 1\right) \quad (7)$$

Where P represents the transition probability matrix; P_{ij} is the likelihood of one state i to be converted to another state j; P_n represents the state probability term and P_{ij} values range between 0 and 1.

3.2.9 Land cover prediction and validation

MLPNN, together with the Markov Chain model, have been widely used for predicting future LULC status in the remote sensing field (Indraja et al., 2024; Ozturk, 2015; Soni et al., 2022; Vinayak et al., 2021). These studies proved the efficiency of these algorithms combined in spatial modelling and handling quantitative data. With such evidence, the MLPNN-Markov from the Land Change Modeler (LCM) in the TerrSet 2020 software was used to perform LULC prediction. The LCM requires input data to predict future land use scenarios. GEE and ArcGIS Pro were employed to process historical LULC maps and predictor variables.

The MLPNN-Markov is the integrated model of MLPNN and Markov aiming to improve the modelling results by compensating for the shortfalls of the individual algorithms. In addition,

MLPNN-Markov joint operation allows the integration of multiple variables and enhances the scope of the model by including other crucial variables in one model. Making the MLPNN-Markov reliable and efficient in predicting LULC changes.

The Markov models exhibit a strong statistical ability to predict probabilities of future land cover, while the MLPNN model learns the historical and land use patterns, which are used as the base information to weigh the transition likelihood among the classes. The main advantages of MLPNN-Markov include its ability to handle the dynamic environment of the complex land use classes and socio-economic factors, its requirement for less user input, and the ease of setting the parameters and calibration.

3.2.10 Model validation

The Kappa index Agreement (KIA) is the most commonly used standard to measure the degree of agreement between reference and simulated maps. However, its feasibility in explaining the agreement between the compared maps has been criticised (Olofsson et al., 2014; Pontius and Millones, 2011). As a result, other researchers recommended the use of the components of agreement and disagreement as part of the validation technique of the results. This study employed three unbiased validation kappa indexes, which include correctness (overall accuracy), Kappa histogram, and kappa location, which provide the accuracy of the model in matching the pixel locations. These validation metrics were computed in the MOLUSCE plugin within the QGIS 2.18 environment. The validation module approach is based on the correlation between the input image and the simulated output image. The three variants determine the Kappa overall value as:

$$\text{Kappa} = \frac{(P_a - P_e)}{(P_i - P_e)} \qquad (8)$$

Where kappa is the Kappa index agreement, *Pa* is the actual accuracy, *Pe* anticipated accuracy, and *Pi* is the desired accuracy, which is 100%.

3.3 Results

Historical land cover analysis was undertaken over a 20-year period, with data collected at half-decade laps (2003, 2007, 2011, 2015, 2019, and 2023). Over the last two decades, the interplay among the nine land use classes has been observed within the uMngeni River catchment, as graphically presented in Figure 3.2 and Figure 2.3.

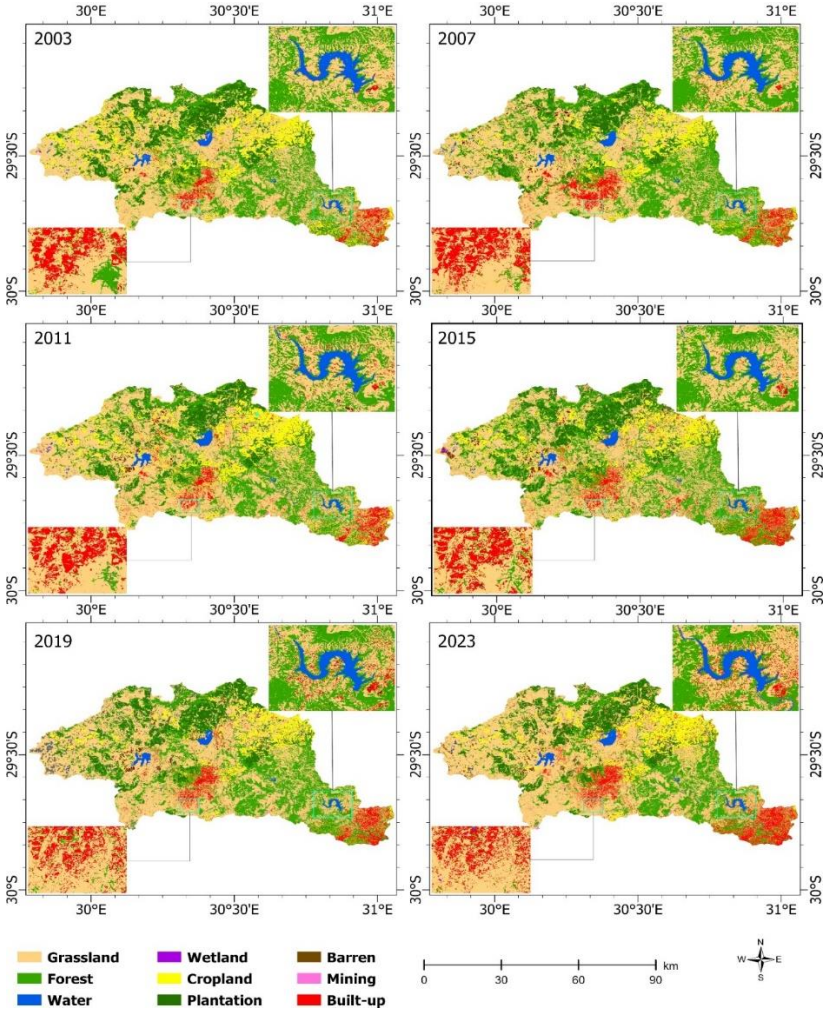

Figure 3.2. Historical LULC thematic maps from 2003 to 2023.

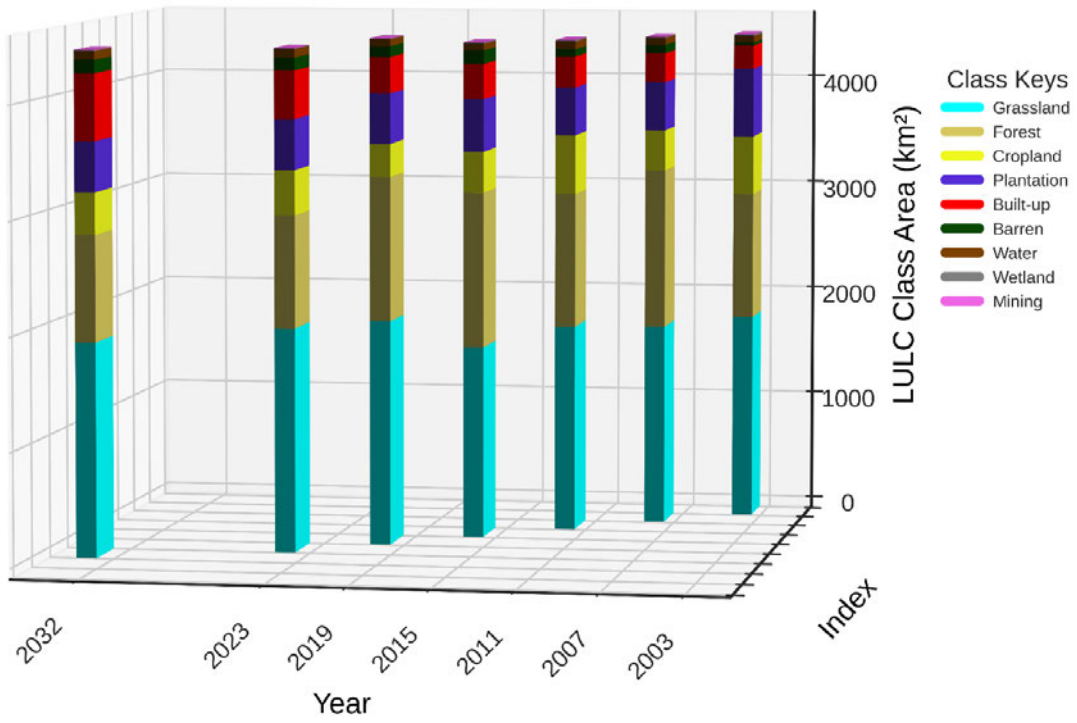

Figure 3.3. Half-decadal area coverage (km^2) from 2003 to 2023 and the predicted LULC class for 2032.

The composition of LULC varied significantly throughout the investigation, according to our findings. In particular, a variety of human activities and natural causes, have transformed native forests, plantations, and farmland. Consequently, native forests experienced a sharp decline (Figure 3.6) with an estimated loss of 3.18% during the research period, having previously covered 25.60% in 2003 (Figure 3.4).

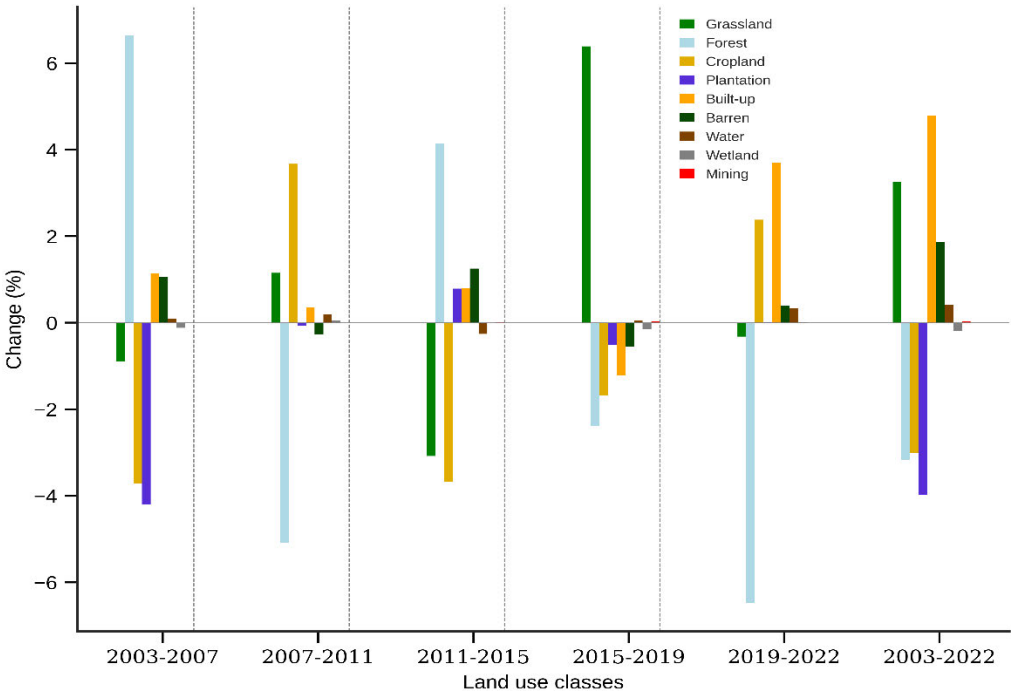

Figure 3.4. Land use class changes in percentage for five-year time steps from 2003 to 2023.

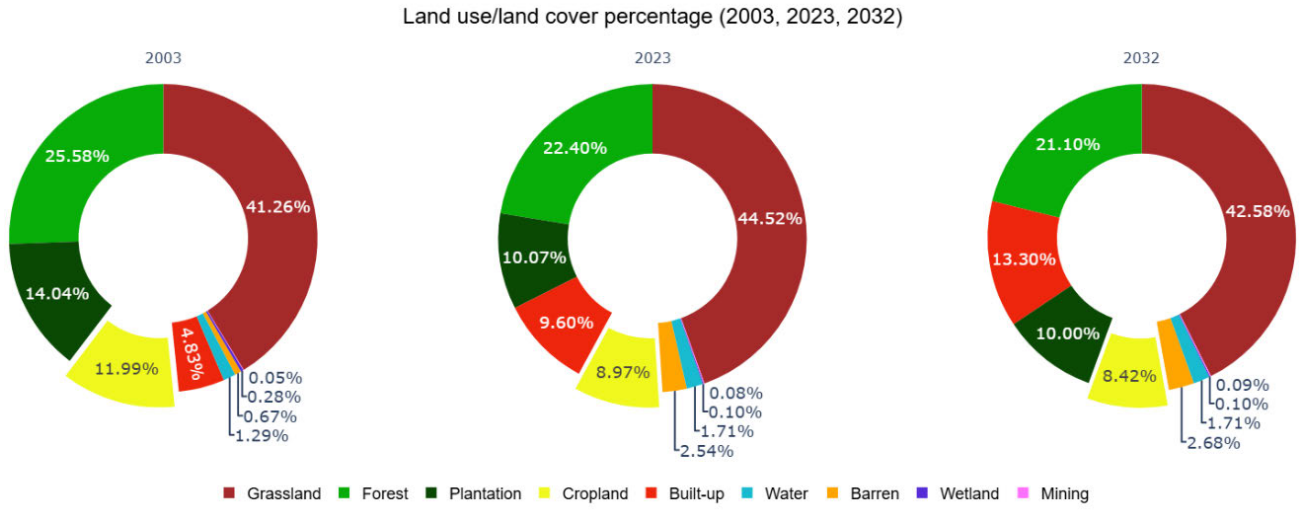

Figure 3.5.uMngeni River Catchment percentage change detection for 2003,2023 and 2032.

Throughout the study period, plantation and agricultural land gradually decreased, from 14% to 10.10% and 12% to 8.97%, respectively. In contrast, built-up land has shown notable increases in percentage cover with a strong alarming trend of 0.92 coefficient of determination (R^2) (Figure 3.6). The built-up category showed gradual growth over time, increasing from covering 4.83% to 9.6% (Figure 3.5), which translates to a total of 5.55% gains in 2023 (Figure 3.4).

Among the primary land use classes in the catchment area, grassland constitutes 44.5% of the catchment. In 2003, it accounted for 41.3% of the total area. Fluctuating patterns were noted between 2003 and 2015 until stability was reached by 2019. A gently alarming shift was noted in bare soil areas, which accounted for 29.71 km^2 (0.67%) in 2003 and gradually expanded to 120.17 km^2 by 2015, indicating an overall increase of 2.57% by 2023 (Table3.4).

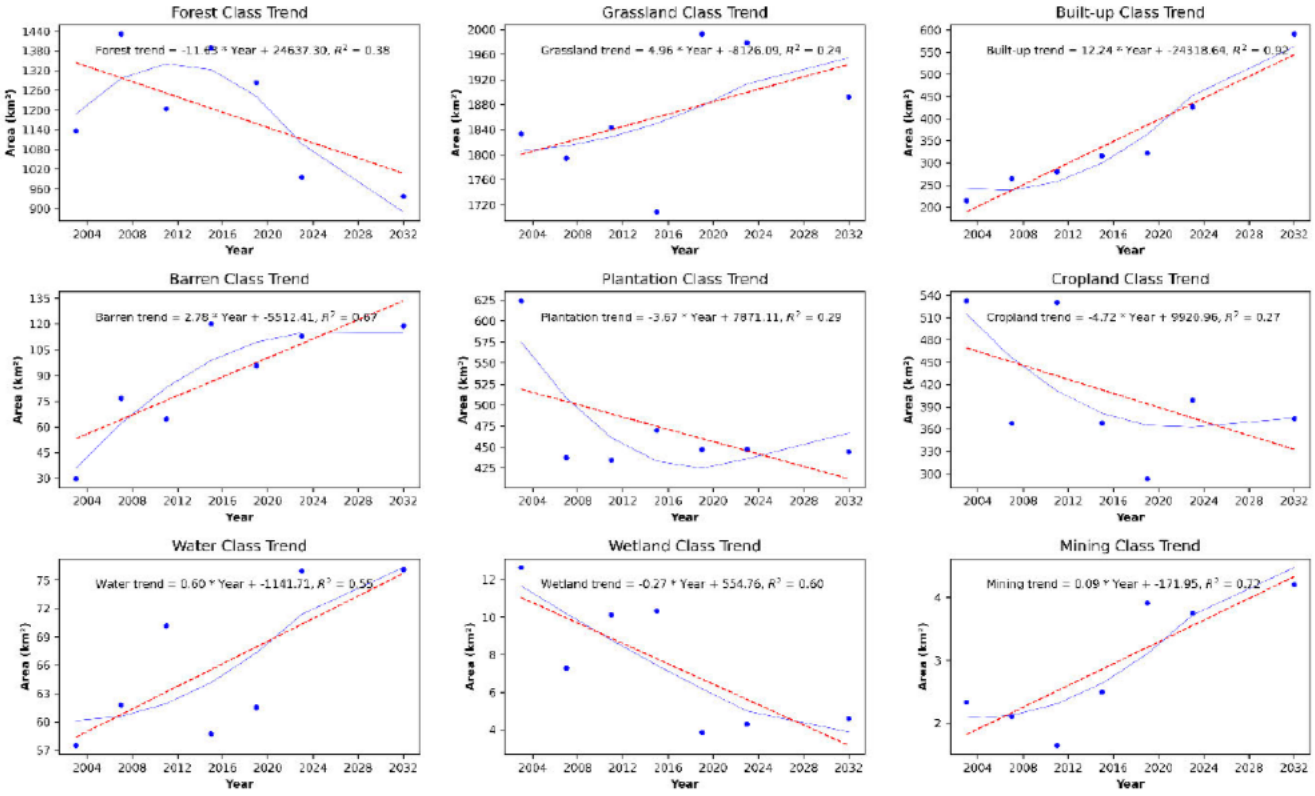

Figure 3.6.Statistical relationship between land use class changes and time.

On the other hand, consistent trends of change with minor variations were recorded for both water and wetland classes, with percentages remaining relatively stable at 0.10% and 1.71% (Table 3.4), respectively. The mines and quarries class, covering a small portion of the catchment, experienced small area gains over time. It is a constantly positive trend as evidenced in R^2 of 0.72 in Figure 3.6.

Table 3.3.Time step land use type coverage in km^2 and percentage from 2003 to 2023.

LULC class	2003		2007		2011		2015		2019		2023		2032	
	Area km^2	%	Area km^2	%	Area km^2	%	Area km^2	%	Area km^2	%	Area km^2	%	Area km^2	%
Grassland	1833.7	41.26	1794.2	40.37	1843.8	41.53	1708.8	38.45	1992.9	44.25	1978.7	44.52	1892.4	42.58

Forest	1136.7	25.58	1431.6	32.21	1204.3	27.12	1389.2	31.26	1283.4	28.49	995.41	22.40	937.82	21.10
Cropland	532.81	11.99	367.95	8.28	530.66	11.95	368.09	8.28	293.21	6.51	398.77	8.97	374.27	8.42
Plantation	624.14	14.04	437.38	9.84	434.5	9.79	470.1	10.58	447.03	9.93	447.32	10.07	444.54	10.00
Built-up	214.46	4.83	265.02	5.96	280.23	6.31	316.11	7.11	322.48	7.16	426.76	9.60	591.09	13.30
Barren	29.71	0.67	76.69	1.73	64.59	1.45	120.17	2.70	95.69	2.12	113.01	2.54	118.9	2.68
Water	57.5	1.29	61.8	1.39	70.18	1.58	58.74	1.32	61.53	1.37	75.98	1.71	76.13	1.71
Wetland	12.64	0.28	7.29	0.16	10.12	0.23	10.33	0.23	3.86	0.09	4.31	0.10	4.62	0.10
Mining	2.33	0.05	2.1	0.05	1.64	0.04	2.49	0.06	3.91	0.09	3.75	0.08	4.2	0.09
Total	4444	100	4444	100	4440	100	4444	100	4504	100	4444	100	4444	100

The accuracy of the classified maps was assessed using standard accuracy metrics (Table 3.4) kappa coefficients recorded were 0.77, 0.78, 0.76, 0.73, 0.75 and 0.76, with corresponding overall accuracies of 80.71%, 82.35%, 80.35%, 77.45%, 80.12%, and 80.33% for the years 2003, 2007, 2011, 2015,2019 and 2023, respectively. The overall accuracy for each map explains closely how the maps produced by the RF model represent the reality on the ground. Achieving an accuracy of 80 % or above shows that classifications fall within the classification acceptable standards. Therefore, the six thematic maps produced were deemed suitable for simulating future land use scenarios.

Table 3.4. Accuracy assessment metrics for six-period interval thematic maps from the Random Forests model.

Year	2003		2007		2011		2015		2019		2023	
LULC class	PA %	UA %	PA %	UA %	PA %	UA %	PA %	UA %	PA %	UA %	PA %	UA %
Grassland	80.44	60.02	84.44	64.44	88.79	60.35	79.07	63.98	88.34	70.29	92.34	60.79
Forest	84.90	75.50	86.63	81.36	89.07	81.11	84.71	63.97	81.92	71.35	85.34	82.80
Water	87.83	99.49	91.47	99.75	98.60	99.31	99.38	99.94	98.76	99.86	99.41	99.74
Wetland	78.02	85.20	67.38	88.81	57.03	73.83	87.13	95.78	70.73	99.340	10.88	85.47
Cropland	91.14	92.36	86.11	83.21	81.94	77.49	78.21	81.32	80.42	90.75	82.66	88.30
Plantation	86.04	94.00	85.26	94.28	79.85	93.99	78.77	92.15	75.65	89.36	91.67	97.67

Barren	58.35	68.42	57.08	73.35	37.61	73.89	65.18	74.05	75.98	81.13	75.01	88.28
Mining	24.85	92.47	54.31	60.23	27.68	97.70	41.56	91.10	45.54	73.82	23.82	87.50
Built-up	66.40	81.95	66.82	86.10	83.01	87.93	56.66	73.22	56.72	81.13	50.80	80.33
Kappa coefficient		0.77		0.78		0.76		0.73		0.75		0.76
Overall accuracy (%)		80.71		82.35		80.35		77.45		80.12		80.33

3.3.1 Prediction of the LULC scenarios for the year 2032

To predict the 2032 land cover map, RF-produced maps from 2003 and 2023 were used as the primary inputs in the LCM module. Before the actual simulation, the trial to test the fitness of the selected MLPNNN-Markov model was conducted by simulating the 2019 land cover map using 2003 and 2011 land cover maps as the inputs. Based on the combination performance at the training stage, the predictor variables, including aspect, DEM, slope, prior anthropogenic disturbance, and Euclidean distance from highways and rivers, were employed. With a training accuracy of 89.75%, trial 1 was chosen as the combination that performed the best (Table 3.5). Integrated with sub-models to produce the transition potential maps later. For every transition sub-model, the parameters and sample distribution were determined by the MLPNN model by default on the fly, as recommended by Clak's lab 2020.

Table 3.5. Sensitivity of MLPNN model at different environmental variable combinations with training accuracy.

Variables (code)
All to Built-up transitions (1)
Slope (2)
Aspect (3)
River Euclidian distance (4)
Roads Euclidian distance (5)
DEM (6)

Trials (T)	Model	Variables included	Accuracy (%)	Skill measure
T1	With all variables	All variables	89.75	0.88
T2	(4) constant	[1,2,3,5,6]	89.15	0.88
T3	[4,3] constant	[1,2,5,6]	89.20	0.88
T4	[4,3,2] constant	[1,5,6]	87.45	0.86
T5	[6,4,3,2] constant	[1,5,]	75.45	0.72
T6	[6,5,4,3,2] constant	[1]	62.70	0.57

Then, transition potential maps were used to generate a transition probability matrix by the Markov probabilistic approach, ascertaining the probability of a class switching to another class, as shown in the probability grid in Table 3.6.

Table 3.6. Transition probability grid for 2019 prediction Trial 1.

Classes	Grassland	Forest	Water	Wetland	Cropland	Plantation	Barren	Mining	Built-up
Grassland	0.6775	0.092	0.0063	0.0031	0.0992	0.0314	0.018	0.0004	0.0717
Forest	0.2265	0.588	0.0054	0.0055	0.0625	0.0768	0.008	0.0003	0.0266
Water	0.007	0.063	0.9179	0.0004	0.0027	0.0019	0.0009	0.0003	0.0058
Wetland	0.5971	0.226	0.0036	0.02	0.07	0.0383	0.0138	0.0004	0.0313
Cropland	0.3009	0.1465	0.0017	0.0015	0.4342	0.0535	0.0207	0.0005	0.0405
Plantation	0.1586	0.4109	0.0012	0.0047	0.0475	0.345	0.017	0.0012	0.0143
Barren	0.45	0.0899	0.008	0.0053	0.1466	0.0214	0.1113	0.0012	0.1657
Mining	0.5233	0.1754	0.0856	0.01	0.0421	0.0152	0.0273	0.0412	0.0795
Built-up	0.3068	0.0479	0.0146	0.0014	0.0775	0.0072	0.0089	0.0009	0.5342

The trial 1 and sub-model criteria design facilitated the MLPNN model to achieve satisfactory accuracy metrics where overall accuracy: 65.50%, $Kappa_{location}$: 58.4%, $Kappa_{Histogram}$: 90.90%., and $Kappa_{overall}$ and 0.52 (Table 3.7). The accurate $Kappa_{Histogram}$ *and* overall accuracy for the 2019 prediction were above 90 % and 65%, respectively. These are the sufficient scores to declare the model's fitness for modelling future land cover scenarios (Hua, 2017). This is also supported by more similar pixel weights per LULC class between the 2019 reference and simulated map, as shown in Figure 3.7, which further reinforces this. These outcomes demonstrated the effectiveness of the MLPNN-Markov hybrid approach in forecasting future LULC. With confidence, the MLPNNN-Markov model was adopted to predict future LULC scenarios for the next 10 years (2032).

Table 3.7. Accuracy assessment metrics for the 2019 simulated map.

Accuracy Matrices	Trial 1
Overall accuracy	65.50%,
$Kappa_{location}$	58.40%,
$Kappa_{Histogram}$	90.90%

$Kappa_{overall}$	0.52

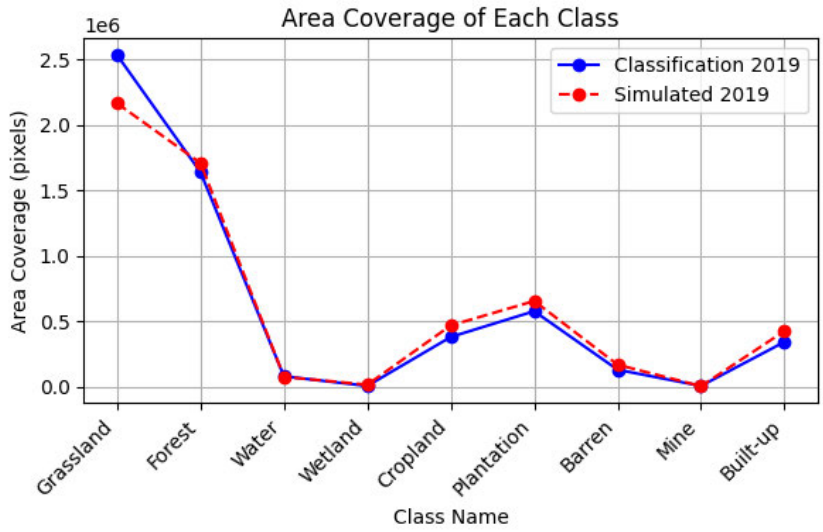

Figure 3.7. Land use class pixel count for the 2019 reference map and 2019 simulated map.

The predicted LULC map from MLPNN-Markov shows an increase in the built-up arial extent (Figure 3.8) to 591.09 km^2, which accounts for 13.30 % of the total catchment area in 2032. These gains explain the declines observed in other projected land use classes, among those classes, forests are expected to decline from 995.41 in 2023 to 937.82 km^2 by 2032 (1.30% area cover loss) (Figure 3.3). On the other hand, grassland will decrease from 1978.7 to 1892.40 km^2, which is about 1.94% area cover lost. Other land cover classes showed insignificant area declines, such as cropland covering 8.42 km^2, and plantation with a slight decline at 444,54 km^2. Water bodies and wetlands show constant trends at 76.13 km^2 and 4.62 km^2, respectively, while mining and barren show increasing trends with insignificant statistics of 4.20 km^2 and 118.90 km^2 (Table 3.3).

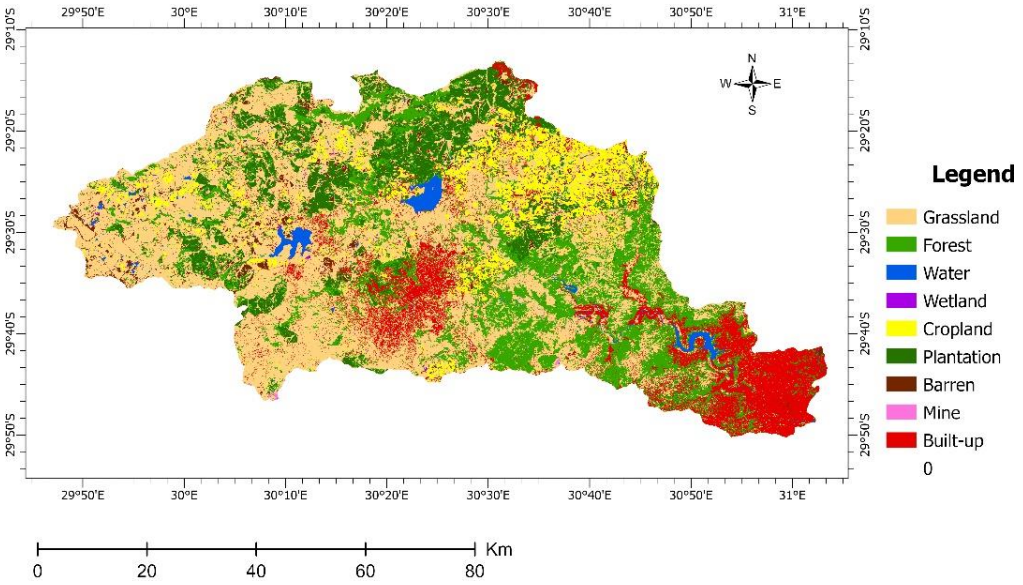

Figure 3.8. Simulated land use map for 2032.

3.4 Discussion

Understanding the past, present, and future land use patterns within the catchment area serves as a foundation for decision-making processes. It allows the enactment and implementation of policies aimed at safeguarding the hydrological resources and biodiversity while ensuring the continuous provision of catchment ecological services. The information unpacked is crucial for sustainable management of river catchment resources. It provides detailed information such as trends and interactions among the land use classes. This information helps in the prompt identification and anticipation of threats to the integrity of the catchment areas. Hence, the findings of this research can be used by environmental resource managers to develop effective plans and management strategies for South African river catchments. This was achieved by leveraging rich open remote sensing data alongside advanced statistical algorithms tailored for LULC analysis.

The Landsat series satellites provided useful spatio-temporal data to achieve meaningful results for mapping historical land use activities. The extensive archives of remote sensing data allow researchers to study both past and current conditions, facilitating the generation of knowledge and understanding to better the future. The success of the Landsat series data in mapping temporal LULC-C in different scenarios has been widely observed globally (Mugo et al., 2020). Its contribution to land cover analysis in river catchment biophysical environments is extensively documented. To complement the Landsat imagery, its routine approach for data captur-

ing has facilitated the acquisition of feasible spatial and temporal data for mapping and simulating land use activities within the uMngeni River Catchment. In the mapping process, the multispectral bands (visible, NIR, and SWIR), along with spectral indices such as NDVI, BSI, and NDWI as explanatory variables, were used in the classification. Their respective sensitivity to different land use types enabled the discrimination of land use classes effectively. With such good logistics, it is possible for MLAs like RF to be capable of producing high-quality maps.

The RF model for classification from the GEE library proved its robust performance in accurately delineating and mapping the land use types within the uMngeni River Catchment. This success stems from the RF model's ability to handle the spectral variations of different land use classes using information contained in the variables (Indraja et al., 2024). In addition, there are several factors that also contributed to the efficiency of the model, which include the inherent design of the classifier and extensive parameter turning, as reported in other studies (Chen, Lin, et al., 2023; Dahiya et al., 2023). Other RF model advantages include proficiency in complex datasets, and its insensitivity to outliers and overfitting, and less sensitivity to different sample sizes (Bhungeni et al., 2024). Due to proficiency results achieved by RF, it was able to unpack the spatiotemporal changes and the interplay among the land use classes at five-year intervals with meaningful results. Such kind of results provided good grounds for the prediction of the future LULC (Soni et al., 2022). Hence, the MLPNN-Markov hybrid yielded reliable results in predicting the future distribution of LULC.

The study found that the uMngeni River catchment has experienced the interplay among the land use classes, leading to observations of spatial changes that occurred at five-year time steps between 2003 -2023. The changes are distributed across the study area, and significant changes are observed in the middle regions spreading along or from the edges of water bodies, southeast, southwest and southern parts of the catchment. The mobility of change shows the swapping and replacement among the land use classes. These changes have been associated with societal demands and natural factors. In this study, it was also found that both human and natural factors contributed to observed changes within the uMngeni River Catchment.

Human-induced activities such as built-up, growing crops, timber, and mining collectively played a role in land cover dynamics observed in the last 20 years. The built-up class showed a steady increase from period to period and claimed 4.78 % of the total catchment by 2023, as Namugize et al. (2018) also reported similar results. It is worth noting that these developments were largely observed in urban outskirts, as (Jewitt et al., 2015) made the same observation. This is encouraged by the need for human settlement, industrial facilities (Adam et al., 2023;

Mashala et al., 2023), and road construction in rural areas (Abbas et al., 2023; Floreano and de Moraes, 2021; Jewitt et al., 2015; Moodley et al., 2023; Ozturk, 2015) and the sparse traditional layout of rural residences (Adam et al., 2023). Resulting in the modifications of the water catchment landscape (Namugize et al., 2018), which has a critical role in ensuring freshwater availability and mitigating climate change. To address this, focus on containing the urban sprawling should be on the top of the agenda, as suggested by (Abbas et al., 2023; Cuypers et al., 2023; Mugo et al., 2020). Thus, the sparse customary layout of residences in rural areas should be contained by promoting dense settlement, and boundaries for land dedicated to the settlement must be demarcated (National Planning Commission. 2012); (Jewitt et al., 2015).

The cropland and barren classes showed haphazard behaviour across the study periods. The unstable tolls of the cropland class are not surprising, as the harvested patches are left bare or undergo pre-planting secondary invasion by herbs and reclaimed later. Similar observations were experienced by the plantation class. As a result, the inversely proportional relationship between these two land cover classes and the barren class is manifested in Figure 3.3. However, cropland experienced a net loss of 3.02% of its positions between 2003 and 2023. This can be associated with many factors, such as abandoned agricultural land later reclaimed by grassland and built-up class members (Abbas et al., 2023), as this transition was observed in the southeast of the catchment. On the other hand, plantation areas showed a sharp decline of 4.20% in the period between 2003 and 2007 and were fairly constant from 2007 to 2023. The reduction of the plantation areas by 2007 can be associated with the ceasing of timber plantation licences by the Department of Forestry and Fisheries as per applicable regulations (Jewitt et al., 2015). In addition, cropland made some advancements during this period taking portions covered by plantation patches in the northern part of the catchment. Making them cover 8.97% and 10.10% by 2023, respectively. The balance between these land cover classes, and the integrity of the water quality classes should be maintained. Many studies proved that water quality crisis correlates with the change in land cover (Abdulkareem et al., 2018; Mararakanye et al., 2022; Mugo et al., 2020; Yao et al., 2016).

Among the human-made developments, mines account for a small area cover of the catchment. Taking its total size into account, it made significant expansion between 2003 and 2023, initially covering 2.33 km^2 and advancing to 3.75 km by 2023 (0.80%). These results agree with (Schoeman et al., 2013), who reported the small Mining area advancement at a national scale. The mines are scattered around the catchment, particularly in the forest banks in the south, and middle regions of the catchment. These are open-cast small sites, starting as a pit and expanding

both outward and downward concurrently as mining continues. In addition, they are very dynamic environments subject to being filled up with water in the case of heavy rains, with observable bare soil areas resulting from outward advancements. This mobilizes the mine by-products, such as effluents, dust particles and non-point pollutants, which pose a health risk to water biodiversity (Plessis et al., 2014). Its implications on terrestrial and freshwater ecosystems reflect those of the cropland class, as previously discussed. However, mines introduce the residuals of heavy metals, making the environment ill-suitable to sustain life. These human activities are the main drivers behind changes or substitutions of naturally occurring classes like grassland.

The forest has shown a constant decline from period to period with a net loss of 3.18% of area cover by 2023 (Figure 3.6). These losses coincide with the constant expansion of built-up, these findings are consistent with a number of previous studies (Jewitt et al., 2015; Mandal et al., 2023; Mashala et al., 2023; Moodley et al., 2023; Namugize et al., 2018; Obaid et al., 2023), plantation and grassland, as their advancement towards the forest areas observed in the southern regions of the catchment, as illustrated in Figure 3.2. This also explains the gains of the grassland over the past 20 years, as it initially covered 41.3% and increased to 44.5% by 2023, representing a net increase of 3.26%. Among many contributing factors, grassland is improved by the recovery of previously disturbed land cover, for example, abandoned croplands or pre-planting season herbs invasion, post-built-up vegetation recovery, and deforestation (Jewitt et al., 2015). The relationship between development and forests raise concerns about water security (Abbas et al., 2023; Maviza and Ahmed, 2020) and expose land to harsh climate conditions (Abbas et al., 2023; Gyamfi-Ampadu et al., 2022; Yao et al., 2016). This reduces groundwater recharge and mobilises the non-point pollutants to freshwater bodies. Therefore, the post-development initiative aimed at rehabilitating the disturbed areas should be considered rather than relying on a natural healing method.

Wetlands were constant across the study term, with negligible changes, such findings have been reported in other studies (Mandal et al., 2023; Moodley et al., 2023). While water bodies show a mild increasing trend from 57.5 km^2 in 2003 to 75.89 km^2, as (Moodley et al., 2023; Namugize et al., 2018) reported similar findings. This can be attributed to the rise of the dam's water levels in response to unprecedented rainfall per annum due to climate change. In addition, the construction of dams and ponds in the west and northeast to retain water for irrigation could have contributed to the area extent of the water bodies.

3.4.1 LULC Projection

The MLPNN-Markov provided insight into the possible future trends of land use changes with acceptable accuracy values. However, it has some limitations, Floreano and de Moraes (2021) highlighted that including only spatial factors in the LULCC prediction model might lead to the exclusion of other major role players, e.g. socio-economic and political influence in configuration of LULC-C. In addition, the limitation of Landsat data stemming from its 30 m has affected the results obtained by MLPNN-Markov; these findings are in line with (Ozturk, 2015; Soni et al., 2022). Moreover, the inability of the RF model to go beyond the pixel level also played a role in this (Adam et al., 2023).

The half-decadal spatiotemporal analysis has enabled the projection of the future land cover. The results of this work will not only give insight into the possible land cover but also give a heads-up to the water resource managers, enabling them to establish policies to prevent the transpiring water quality issues. Overall, the predicted land cover classes do not vary greatly from the 2023 land cover map. Our results showed that Built-up is expected to increase by 3.7% by 2032. In the southeast, the intensified density in the existing build-up areas, and in the northeast, there is also noticeable scattered built-up, as Moodley et al., (2023) made similar observations. This can stem from a commitment by eThekwini Municipality to build houses for 26% of KZN province dwellers at a rate of 5000 houses per annum (IDP 2020). In addition, the South African government aims to upgrade the informal settlement and develop dense townships to reduce urban sprawling by 2030 (National Planning Commission. 2012). Moreover, the scattered developments are encouraged by the construction of road lines in rural areas, as it makes it easy to access the place and resources (Abbas et al., 2023; Jewitt et al., 2015). These projected developments are expected to take place at the cost of natural vegetation and equally affect the water quality, as reported by (Plessis et al., 2014).

In 2032, forest and grassland cover classes are expected to surrender 1.30% and 1.94% to other classes, respectively. In particular, these losses are associated with human intrusion. Due to their large area coverage, developments are likely to temper with the natural vegetation. Hence, some patches of these two land cover classes are being replaced by the built-up and mining activities in the middle and the southeast regions. Natural factors like forest encroachment and land degradation stemming from climate change could be the reason for the decline of forest area extent. Other land cover classes showed insignificant area declines, such as cropland covering 8.42 km^2, and plantation with a slight decline at 444,54 km^2. This decline can be explained by the counter area covered by post-harvest barren soil. However, the bare soil total

area cover can also account for partially vegetated areas and fire belts observable in the tip of the western region.

Water bodies and wetlands show constant trends at 76.13 km^2 and 4.62 km^2, respectively. Notably, water increased about 2 km^2 which could be associated with water level variations stemming from climate change as previously discussed. While mining and built-up showed increasing trends with insignificant rates (Table 3.3). However, the detrimental impact rising from small mining areas is quite concerning. Therefore, an integrated approach should be established to balance the mining development and water quality, as the runoff is more frequent and intensified by climate change.

3.5 Conclusion

Our study provided insight into the historical land use configurations at half-decade intervals between 2003 and 2023 and elevated the viewpoint for possible future land use cover patterns in the uMngeni River Catchment. The combination of RF and L8-OLI proved to be an effective logistic tool by providing the required data for predicting future land cover. Land cover maps with an overall accuracy of greater than 80%, except in 2007, were produced and subsequently used as the base data for the 2032 prediction using a hybrid MLPNN-Markov model. Prior to the land cover prediction, 2003 and 2011 were used to test the fitness of the MLPNN-Markov model by predicting 2019 land cover classes. The model was approved for the actual simulation as it produced a map that was closely more similar to that of actual 2019, manifested in the $Kappa_{histogram}$, and overall accuracy, which all were above 65%.

According to the outcomes of RF and MLPNN-Markov, the built-up area cover showed a strong positive trend of 0.92%. These gains occur at the cost of grassland and forest, and this antagonistic relationship is expected to maintain the same pace by 2032. It is also worth highlighting small advancements observed in the mining land use class, as its impact does not match its spatial cover. Plantation and Cropland exhibited mild changes across the study periods. However, holding big stakes of the catchment's total land coupled with their related environmental impacts.

The findings of this study recommend shifting from the customary layout of rural and township development to overlayed dense buildings with a clear demarcation of urbanized areas. Mining activities should be practiced precisely with the regulations in place aimed at protecting against water contamination. The impact of cultivation and timber industry residuals in the water ecosystems is non-negligible, as spelt out in the literature. Therefore, water resource managers and

the agricultural sector must strike a balance between cultivation practices and water sanity. Wetlands and water support life to aquatic species and serve as living systems for natural water purification. Hence, we recommend that their ecological integrity should be prioritised in the planning strategies aimed at the preservation of the catchment natural state.

Overall, our study proved the effectiveness of the RF classification algorithm together with Landsat imagery for unpacking the historical land use scenarios and as a vehicle for meaningful projection results achieved by the MLPNN-Markov Model. The outcomes of this study will not only demonstrate the application of remote sensing and GIS in acquiring the historical LULC underlying data but will also serve as the early detection tool for future LULC scenarios and their accompanying environmental impact. This enables the water resource managers and other local environmental authorities to be on high alert and be able to put preventative measures to change the predicted scenarios. The methods used in the study can also be adapted and implemented in other landscapes to gain a detailed understanding of LULC dimensions and trends. However, it should be noted that 30 m resolution data have reduced the accuracy of the classifications. In addition, the inability of the RF model to go beyond the pixel level also played a role. Therefore, future research should attempt to reproduce the historical satellite data by mimicking the properties (calibration) of high-resolution images such as PlantScope and RapidEye.

3.6 Limitations and Recommendations

Two land cover classes with more similar optical properties, for example, mining quarries and tar roads or parking lots, the RF model, struggle to separate and tend to be biased to land use class with the higher training sample allocation.

MLPNN-Markov tends to ignore the scattered customary layout of the rural development and bind the dense development along the provided vector lines (rivers and roads).

MLPNN-Markov has some limitations, as also pointed out by (Floreano and de Moraes 2021). The MLPNN-Markov only accommodates spatial factors in the LULC prediction, which might lead to the exclusion of other major role players, e.g. socio-economic and political influence in LULC scenarios.

The secondary invasion of harvested areas by the herbaceous and grass species interferes with the interplay between barren land and cropland, leading to grassland overestimation.

The residential areas with defined tree canopies and green grass open spaces (lawn) partially masked the building structure, causing mixed pixels, making it not guaranteed to award the pixels to the built-up class. The transition zones between the two land parcels tend to possess mixed pixels, as a result of confusing the classifier. Ultimately, impacted the accuracy of the prediction maps.

To enhance the accuracy of the change detection and simulation results, it is recommended the use of high-resolution images in the classification stage. Hence, we propose that future research should attempt to build a model capable of learning the spectral properties of high-resolution images and reproduce the historical satellite data by mimicking those properties of high-resolution imagery data such as PlantScope and RapidEye and comparative work between high-resolution and moderate resolute LULC simulation.

4 CHAPTER 4: SYNTHESIS REPORT OF THE DISSERTATION

4.1 Synthesis of the chapters

The research objectives set out in this study were achieved. The overall results of this work demonstrate how the remote sensing capabilities in studying the heterogeneity of catchment optical properties. Particularly, categorically quantifying the water catchment LULC, historical land use analysis and projecting the potential distribution of LULC types in the future. All these three objectives were accomplished through leveraging remote sensing applications. Such studies prove the effectiveness of remote sensing applications in the realm of natural resource conservation. The methodologies presented in this work can be used as guides and applied by other studies examining LULC in similar ecological settings. Moreover, served as an anchor for the regional water resource managers and third parties in establishing the water resource management strategies based on informed decisions. The synthesis of this work is sub-sectioned based on the manuscripts.

4.1.1 Evaluating Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal

The water catchment areas not only provide hospitality services to humans and wildlife but are also a key regional strategic water source. However, the trajectory of the LULC is posing a serious threat to those services and is accompanied by implications for the water quality. The outcomes of our work reaffirm that remote sensing mapping is a viable way to provide a timely quantitative and graphical representation of the status of LULC in the catchment areas. Which is the fundamental tool facilitating initiatives aimed at protecting the integrity of the catchment and safeguarding the sanity of water resources.

The Landsat 8 and advanced MLAs (NB, RF, SVM, and ANN) successfully classified the dimensional land use classes of the catchment with varying accuracy results. The performance of these classifiers was compared using the commonly used components of agreements. The RF model outperformed other classifiers with overall accuracy and a Kappa score of 97.02% and 0.96, respectively. Similar findings have also been reported in other MLA comparable

studies in classifying complex land use classes (Ashiagbor et al., 2023; Loukika et al., 2021; Ouchra et al., 2023; Parashar et al., 2024). The performance is awarded to the inherent design of the RF model, which makes it less sensitive to sample size, the ability to handle complex data, and the advantage of the random multi-decision trees in generalization—which are immune to outliers.

The model performance is determined by a range of factors. Among those factors, in this study, the variable collinearity and importance based on the SHARP value analysis were implemented to remove noise elements and get insight into the utilisation of the variables passed to the model. A total of six multi-spectral bands and three indices were selected to facilitate the classification process. The SHARP value model interpretation results showed that bands from SWIR were found to be the most important bands. This can be accredited to sensitivity to vegetation foliage, given that our study area is largely covered by vegetation. Similar observations have been reported in other studies such as (Abdi, 2020; Chen, Yang, et al., 2023; Gyamfi-Ampadu et al., 2020; McCarty et al., 2020; Schultz et al., 2015; Sothe et al., 2017). Hence, we recommend the inclusion of SWIR bands in the LULC classification in the areas with high vegetation content. In fact, the toolkit and the methodology used in this study are recommended to be adopted in similar studies with comparable geographic physical features.

4.1.2 Chapter 3: LULC Change Detection and Future LULC Modelling Using RF and MLPNN-Markov Algorithms in the uMngeni Catchment, Kwazulu-Natal, South Africa

The study aimed to analyse LULC-C that took place over the period of 20 years from 2003 to 2023 within the uMngeni River Catchment area. The findings of this study revealed the interplay of LULC classes during the study periods, leading to transitions between the classes (dynamic and static). Results showed the grassland as the most dominant land cover class, followed by the forest, with wetlands and mines making the least of the catchment area coverage. These are ideal quantities, as they favour rain harvesting and water retention. Most importantly, these classes play a pivotal role in local climate mitigation. However, the built-up showed a constant increasing trend from period to period which correlates with the constant decline of the forest land use class. This poses a serious threat to the health of catchment

ecological services. The LULC-C observed during the study period has been associated with the human intrusion. Hence, we recommend that the authorities establish the policy demarcation limits on urban expansion and adopt the density residential areas rather than the sparse customary layout.

The forecasting of the possible future land cover state is a pivotal tool in the management of natural resources as it allows foreseeing the likely environmental threats and enables minimizing or avoiding those threats. The findings of the LULC prediction for 2032 revealed that the grassland and the forest land cover will decline at the expense of the built-up land use class. While other classes, such as cropland, barren, water, mines, and plantation show insignificant changes. The trends of projected LULC generally are more similar to what was observed during the 2003 and 2023 periods. Which is the alert to the water resource managers to divert or mitigate these alarms.

The rich archives of Landsat imagery provided feasible temporal resolution data, enabling the temporal land use analysis from 2003 to 2023. In addition, giving an opportunity to get insight into the historical land use patterns enabled the prediction of future land cover scenarios. With L8-OLI 30 m resolution, all the identified classes in the uMngeni River Catchment were successfully delineated with good accuracy metrics. Similar findings have been reported in other studies focused on LULC analysis in the catchment areas (Mandal et al., 2023; Mugo et al., 2020; Namugize et al., 2018; Obaid et al., 2023). Some other studies utilise produced classifications to map the possible future LULC leveraging the MLPNN-Markov (Abbas et al., 2023).

The half-decadal and projected spatial information unpacked in this study will equip the water resource managers with quantitative information to troubleshoot water quality-related issues. The unprecedented urban expansion in the catchment with its related water, ecosystem, and climate issues should be contained with a set of recommendations presented in this work.

References

Abbas, H., Tao, W., Khan, G., Alrefaei, A. F., Iqbal, J., Albeshr, M. F., and Kulsoom, I. (2023). Multilayer perceptron and Markov Chain analysis based hybrid-approach for predicting land use land cover change dynamics with Sentinel-2 imagery. *Geocarto International*, *38*(1). https://doi.org/10.1080/10106049.2023.2256297

Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience and Remote Sensing*, *57*(1), 1–20. https://doi.org/10.1080/15481603.2019.1650447

Abdulkareem, J. H., Sulaiman, W. N. A., Pradhan, B., and Jamil, N. R. (2018). Long-term hydrologic impact assessment of non-point source pollution measured through Land Use/Land Cover (LULC) changes in a tropical complex catchment. *Earth Systems and Environment*, *2*(1), 67–84. https://doi.org/10.1007/s41748-018-0042-1

Adam, E., Masupha, N. E., and Xulu, S. (2023). Spatial assessment and prediction of urbanization in Maseru using earth observation data. *Applied Sciences*, *13*(10), 5854. https://doi.org/10.3390/app13105854

Aghsaei, H., Mobarghaee Dinan, N., Moridi, A., Asadolahi, Z., Delavar, M., Fohrer, N., and Wagner, P. D. (2020). Effects of dynamic land use/land cover change on water resources and sediment yield in the Anzali wetland catchment, Gilan, Iran. *Science of the Total Environment*, *712*. https://doi.org/10.1016/j.scitotenv.2019.136449

Ahmed, B., and Ahmed, R. (2012). Modeling urban land cover growth dynamics using multi-temporal satellite images: A case study of Dhaka, Bangladesh. *ISPRS International Journal of Geo-Information*, *1*(1), 3–31. https://doi.org/10.3390/ijgi1010003

Akar, O., and Tunc Gormus, E. (2022). Land use/land cover mapping from airborne hyperspectral images with machine learning algorithms and contextual information. *Geocarto International*, *37*(14), 3963–3990. https://doi.org/10.1080/10106049.2021.1945149

Aneseyee, A. B., Soromessa, T., and Elias, E. (2020). The effect of land use/land cover changes on ecosystem services valuation of Winike watershed, Omo Gibe basin, Ethiopia. *Human and Ecological Risk Assessment*, *26*(10), 2608–2627. https://doi.org/10.1080/10807039.2019.1675139

Ashiagbor, G., Asare-Ansah, A. O., Amoah, E. B., Asante, W. A., and Mensah, Y. A. (2023). Assessment of machine learning classifiers in mapping the cocoa-forest mosaic landscape of Ghana. *Scientific African*, *20*, e01718. https://doi.org/10.1016/j.sciaf.2023.e01718

Bayas, S., Sawant, S., Dhondge, I., Kankal, P., and Joshi, A. (2022). *Land Use Land Cover Classification Using Different ML Algorithms on Sentinel-2 Imagery* (pp. 761–777). https://doi.org/10.1007/978-981-19-0840-8_59

Belgiu, M., and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bhaga, T. D., Dube, T., Shekede, M. D., and Shoko, C. (2023). Investigating the effectiveness of Landsat-8 OLI and Sentinel-2 MSI satellite data in monitoring the effects of drought on surface

water resources in the Western Cape Province, South Africa. *Remote Sensing Applications: Society and Environment*, *32*, 101037. https://doi.org/10.1016/j.rsase.2023.101037

Bhungeni, O., Ramjatan, A., and Gebreslasie, M. (2024). Evaluating Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal. *Remote Sensing*, *16*(12), 2219. https://doi.org/10.3390/rs16122219

Biau, G., and Scornet, E. (2016). A random forest guided tour. *Test*, *25*, 197–227.

Boakye, E., Odai, S. N., Adjei, K. A., & Annor, F. O. (2008). Landsat images for assessment of the impact of land use and land cover changes on the Barekese catchment in Ghana. European Journal of Scientific Research, 22(2), 269-278. http://www.eurojournals.com/ejsr.htm

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. Random Forests (springer.com)

Burnham, B. O. (1973). Markov intertemporal land use simulation model. *Journal of Agricultural and Applied Economics*, *5*(1), 253–258. https://doi.org/10.1017/S0081305200011110

Camargo, F. F., Sano, E. E., Almeida, C. M., Mura, J. C., and Almeida, T. (2019). A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. *Remote Sensing*, *11*(13). https://doi.org/10.3390/rs11131600

Campbell, J. B., and Wynne, R. H. (2011). *Introduction to remote sensing*. Guilford press.

Carlotto, M. J. (2009). Effect of errors in ground truth on classification accuracy. *International Journal of Remote Sensing*, *30*(18), 4831–4849. https://doi.org/10.1080/01431160802672864

Chen, H., Lin, X., Sun, Y., Wen, J., Wu, X., You, D., Cheng, J., Zhang, Z., Zhang, Z., Wu, C., Jian, H., and Guan, X. (2023). Performance assessment of four data-driven machine learning models: A case to generate sentinel-2 albedo at 10 meters. *Remote Sensing*, *15*(10). https://doi.org/10.3390/rs15102684

Chen, H., Yang, L., and Wu, Q. (2023). Enhancing land cover mapping and monitoring: An interactive and explainable machine learning approach using Google Earth Engine. *Remote Sensing*, *15*(18), 4585. https://doi.org/10.3390/rs15184585

Clark Labs (2016). Geospatial and monitoring and modelling system: Tutorial. Accessed 15 July 2024. TerrSet Tutorial (clarklabs.org)

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Costache, R., Tin, T. T., Arabameri, A., Crăciun, A., Costache, I., Islam, A. R. Md. T., Sahana, M., and Pham, B. T. (2022). Stacking state-of-the-art ensemble for flash-flood potential assessment. *Geocarto International*, *37*(26), 13812–13838. https://doi.org/10.1080/10106049.2022.2082558

Cuypers, S., Nascetti, A., and Vergauwen, M. (2023). Land Use and Land Cover Mapping with VHR and multi-temporal Sentinel-2 Imagery. *Remote Sensing*, *15*(10), 2501. https://doi.org/10.3390/rs15102501

Dahiya, N., Gupta, S., and Singh, S. (2023). Qualitative and quantitative analysis of artificial neural network-based post-classification comparison to detect the earth surface variations using hyperspectral and multispectral datasets. *Journal of Applied Remote Sensing*, *17*(3). https://doi.org/10.1117/1.JRS.17.032403

Dash, P., Sanders, S. L., Parajuli, P., and Ouyang, Y. (2023). Improving the accuracy of land use and land cover classification of Landsat data in an agricultural watershed. *Remote Sensing*, *15*(16). https://doi.org/10.3390/rs15164020

Delogu, G., Caputi, E., Perretta, M., Ripa, M. N., and Boccia, L. (2023). Using PRISMA hyperspectral data for land cover classification with artificial intelligence support. *Sustainability*, *15*(18), 13786. https://doi.org/10.3390/su151813786

Department of Forestry, Fisheries and the Environment. (2024). *South African National Land Cover Datasets*. Pretoria: Department of Forestry, Fisheries and the Environment. Accessed 11 May 2024. https://egis.environment.gov.za/sa_national_land_cover_datasets.

Deshpande, A. R., and Emmanuel, M. (2016). Context based Recommendation Methods: A Brief Review. *International Journal of Computer Applications*, 14–19.

Fairbanks, D., Thompson, M., Vink, D., Newby, T., Van den Berg, H., and Everard, D. (2000). The South African land-cover characteristics database: a synopsis of the landscape. *South African Journal of Science*, *96*(2), 69–82. https://hdl.handle.net/10520/AJA00382353_7429

Floreano, I. X., and de Moraes, L. A. F. (2021). Land use/land cover (LULC) analysis (2009–2019) with Google Earth Engine and 2030 prediction using Markov-CA in the Rondônia State, Brazil. *Environmental Monitoring and Assessment*, *193*(4), 239. https://doi.org/10.1007/s10661-021-09016-y

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley, R. D., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Joseph Hughes, M., and Laue, B. (2017). Cloud detection algorithm comparison and

validation for operational Landsat data products. *Remote Sensing of Environment*, *194*, 379–390. https://doi.org/10.1016/j.rse.2017.03.026

Foody, G. M. (2010). Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of Environment*, *114*(10), 2271–2285. https://doi.org/10.1016/J.RSE.2010.05.003

Foody, G. M. (2015). Valuing map validation: The need for rigorous land cover map accuracy assessment in economic valuations of ecosystem services. *Ecological Economics*, *111*, 23–28. https://doi.org/10.1016/J.ECOLECON.2015.01.003

Gandhi, U. (2021). End-to-End Google Earth Engine Course. *Spatial Thoughts*. https://courses.

spatialthoughts.com/end-to-end-gee.html (accessed on 2 February 2024).

Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gyamfi-Ampadu, E., Gebreslasie, M., and Mendoza-Ponce, A. (2020). Mapping natural forest cover using satellite imagery of Nkandla forest reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment*, *18*. https://doi.org/10.1016/j.rsase.2020.100302

Gyamfi-Ampadu, E., Gebreslasie, M., and Mendoza-Ponce, A. (2021). Evaluating multi-sensors spectral and spatial resolutions for tree species diversity prediction. *Remote Sensing*, *13*(5). https://doi.org/10.3390/rs13051033

Gyamfi-Ampadu, E., Gebreslasie, M., and Mendoza-Ponce, A. (2022). Multi-Decadal Spatial and Temporal Forest Cover Change Analysis of Nkandla Natural Reserve, South Africa. *Journal of Sustainable Forestry*, *41*(10), 959–982. https://doi.org/10.1080/10549811.2021.1891441

Hafeez, S., Wong, M. S., Ho, H. C., Nazeer, M., Nichol, J., Abbas, S., Tang, D., Lee, K. H., and Pun, L. (2019). Comparison of machine learning algorithms for retrieval of water quality indicators in case-ii waters: A case study of hong kong. *Remote Sensing*, *11*(6). https://doi.org/10.3390/rs11060617

Hosseiny, B., Abdi, A. M., and Jamali, S. (2022). Urban land use and land cover classification with interpretable machine learning – A case study using Sentinel-2 and auxiliary data. *Remote Sensing Applications: Society and Environment*, *28*, 100843. https://doi.org/10.1016/j.rsase.2022.100843

Hua, A. K. (2017). Land Use Land Cover Changes in Detection of Water Quality: A Study Based on Remote Sensing and Multivariate Statistics. *Journal of Environmental and Public Health*, *2017*. https://doi.org/10.1155/2017/7515130

Huang, C., Davis, L. S., and Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323

Hughes, C. J., De Winnaar, G., Schulze, R. E., Mander, M., and Jewitt, G. P. W. (2018). Mapping of water-related ecosystem services in the uMngeni catchment using a daily time-step hydrological model for prioritisation of ecological infrastructure investment - part 1: Context and modelling approach. *Water SA*, *44*(4), 577–589. https://doi.org/10.4314/wsa.v44i4.07

IDP .(2020). Accessed 20 July 2024. eThekwini Integrated Development Plan IDP 2020 - 2021.pdf (durban.gov.za)

Indraja, G., Aashi, A., and Vema, V. K. (2024). Spatial and temporal classification and prediction of LULC in Brahmani and Baitarni basin using integrated cellular automata models. *Environmental Monitoring and Assessment*, *196*(2). https://doi.org/10.1007/s10661-023-12289-0

James, D., Collin, A., Mury, A., and Costa, S. (2020). Very high resolution land use and land cover mapping using pleiades-1 stereo imagery and machine learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B2-2020*, 675–682. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-675-2020

Jewitt, D., Goodman, P. S., Erasmus, B. F. N., O'Connor, T. G., and Witkowski, E. T. F. (2015). Systematic land-cover change in KwaZulu-Natal, South Africa: Implications for biodiversity. *South African Journal of Science*, *111*(9–10). https://doi.org/10.17159/sajs.2015/20150019

Kang, C. S., Kanniah, K. D., and Mohd Najib, N. E. (2021). Google Earth Engine for Landsat image processing and monitoring land use/ land cover changes in the Johor River Basin, Malaysia. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 4236–4239. https://doi.org/10.1109/IGARSS47720.2021.9554768

Kaur, H., Tyagi, S., Mehta, M., and Singh, D. (2023). Time series (2001/2002–2021) analysis of Earth observation data using Google Earth Engine (GEE) for detecting changes in land use land cover (LULC) with specific reference to forest cover in East Godavari Region, Andhra Pradesh, India. *Journal of Earth System Science*, *132*(2). https://doi.org/10.1007/s12040-023-02099-w

Kaval, P. (2019). Integrated catchment management and ecosystem services: A twenty-five year overview. *Ecosystem Services*, *37*, 100912. https://doi.org/10.1016/j.ecoser.2019.100912

Kibena, J., Nhapi, I., and Gumindoga, W. (2014). Assessing the relationship between water quality parameters and changes in landuse patterns in the Upper Manyame River, Zimbabwe. *Physics and Chemistry of the Earth*, *67–69*, 153–163. https://doi.org/10.1016/j.pce.2013.09.017

Kim, Y. H., Im, J., Ha, H. K., Choi, J. K., and Ha, S. (2014). Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GIScience and Remote Sensing*, *51*(2), 158–174. https://doi.org/10.1080/15481603.2014.900983

Kumar, V., and Agrawal, S. (2023). A multi-layer perceptron–Markov chain based LULC change analysis and prediction using remote sensing data in Prayagraj district, India. *Environmental Monitoring and Assessment*, *195*(5), 619. https://doi.org/10.1007/s10661-023-11205-w

Kusangaya, S., Warburton, M., and Archer van Garderen, E. (2017). Use of ACRU, a distributed hydrological model, to evaluate how errors from downscaled rainfall are propagated in simulated runoff in uMngeni catchment, South Africa. *Hydrological Sciences Journal*, *62*(12), 1995–2011. https://doi.org/10.1080/02626667.2017.1349317

Kusangaya, S., Warburton, M. L., Archer, E., and Garderen, V. (2018). Evaluation of uncertainty in capturing the spatial variability and magnitudes of extreme hydrological events for the uMngeni catchment, South Africa. *Journal of Hydrology*, *557*, 931–946. https://doi.org/10.1016/j.jhydrol.2018.01.017

Lantzanakis, G., Mitraka, Z., and Chrysoulakis, N. (2021). X-SVM: An Extension of C-SVM Algorithm for Classification of High-Resolution Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(5), 3805–3815. https://doi.org/10.1109/TGRS.2020.3017937

Lekka, C., Petropoulos, G. P., and Detsikas, S. E. (2024). Appraisal of EnMAP hyperspectral imagery use in LULC mapping when combined with machine learning pixel-based classifiers. *Environmental Modelling and Software*, *173*, 105956. https://doi.org/10.1016/j.envsoft.2024.105956

Lillesand, T., Kiefer, R. W., and Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley and Sons.

Liu, Y., Meng, Q., Zhang, L., and Wu, C. (2022). NDBSI: A normalized difference bare soil index for remote sensing to improve bare soil mapping accuracy in urban and rural areas. *CATENA*, *214*, 106265. https://doi.org/10.1016/j.catena.2022.106265

Loukika, K. N., Keesara, V. R., and Sridhar, V. (2021). Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India. *Sustainability*, *13*(24), 13758. https://doi.org/10.3390/su132413758

Lukeš, P., Stenberg, P., Rautiainen, M., Mõttus, M., and Vanhatalo, K. M. (2013). Optical properties of leaves and needles for boreal tree species in Europe. *Remote Sensing Letters*, *4*(7), 667–676. https://doi.org/10.1080/2150704X.2013.782112

Lundberg, S. M., Allen, P. G., and Lee, S.-I. (n.d.). *A Unified Approach to Interpreting Model Predictions*. https://github.com/slundberg/shap

Lynch, A. J., Cooke, S. J., Arthington, A. H., Baigun, C., Bossenbroek, L., Dickens, C., Harrison, I., Kimirei, I., Langhans, S. D., Murchie, K. J., Young, N., and Jähnig, S. C. (2023). People need freshwater biodiversity. *Wiley Interdisciplinary Reviews: Water*, *10*(3). https://doi.org/10.1002/wat2.1633

Machado, R. A. S., Oliveira, A. G., and Lois-González, R. C. (2019). Urban ecological infrastructure: The importance of vegetation cover in the control of floods and landslides in Salvador / Bahia, Brazil. *Land Use Policy*, *89*. https://doi.org/10.1016/j.landusepol.2019.104180

Mahmoud, R., Hassanin, M., Al Feel, H., and Badry, R. M. (2023). Machine Learning-Based Land Use and Land Cover Mapping Using Multi-Spectral Satellite Imagery: A Case Study in Egypt. *Sustainability*, *15*(12), 9467. https://doi.org/10.3390/su15129467

Mandal, S., Bandyopadhyay, A., and Bhadra, A. (2023). Dynamics and future prediction of LULC on Pare River basin of Arunachal Pradesh using machine learning techniques. *Environmental Monitoring and Assessment*, *195*(6), 709. https://doi.org/10.1007/s10661-023-11280-z

Mararakanye, N., Le Roux, J. J., and Franke, A. C. (2022). Long-term water quality assessments under changing land use in a large semi-arid catchment in South Africa. *Science of the Total Environment*, *818*. https://doi.org/10.1016/j.scitotenv.2021.151670

Martins, S., Bernardo, N., Ogashawara, I., and Alcantara, E. (2016). Support Vector Machine algorithm optimal parameterization for change detection mapping in Funil Hydroelectric

Reservoir (Rio de Janeiro State, Brazil). *Modeling Earth Systems and Environment*, *2*(3), 138. https://doi.org/10.1007/s40808-016-0190-y

Mashala, M. J., Dube, T., Ayisi, K. K., and Ramudzuli, M. R. (2023). Using the Google Earth Engine cloud-computing platform to assess the long-term spatial temporal dynamics of land use and land cover within the Letaba watershed, South Africa. *Geocarto International*, *38*(1). https://doi.org/10.1080/10106049.2023.2252781

Mather, P. M., and Koch, M. (2011). *Computer processing of remotely-sensed images: an introduction*. John Wiley and Sons.

Mauck, B. A., and Warburton, M. (2014). Mapping areas of future urban growth in the Mgeni catchment. *Journal of Environmental Planning and Management*, *57*(6), 920–936. https://doi.org/10.1080/09640568.2013.775062

Maviza, A., and Ahmed, F. (2020). Analysis of past and future multi-temporal land use and land cover changes in the semi-arid Upper-Mzingwane sub-catchment in the Matabeleland south province of Zimbabwe. *International Journal of Remote Sensing*, *41*(14), 5206–5227. https://doi.org/10.1080/01431161.2020.1731001

McCarty, D. A., Kim, H. W., and Lee, H. K. (2020). Evaluation of light gradient boosted machine learning technique in large scale land use and land cover classification. *Environments*, *7*(10), 84. https://doi.org/10.3390/environments7100084

Mekuria, W., Gedle, A., Tesfaye, Y., and Phimister, E. (2023). Implications of changes in land use for ecosystem service values of two highly eroded watersheds in Lake Abaya Chamo sub-basin, Ethiopia. *Ecosystem Services*, *64*. https://doi.org/10.1016/j.ecoser.2023.101564

Moodley, K., Toucher, M. L., and Lottering, R. T. (2023). Simulating future land-use within the uThukela and uMngeni catchments in KwaZulu-Natal. *Scientific African*, *20*. https://doi.org/10.1016/j.sciaf.2023.e01666

Mugo, R., Waswa, R., Nyaga, J. W., Ndubi, A., Adams, E. C., and Flores-Anderson, A. I. (2020). Quantifying land use land cover changes in the lake victoria basin using satellite remote sensing: The trends and drivers between 1985 and 2014. *Remote Sensing*, *12*(17), 1–17. https://doi.org/10.3390/rs12172829

Namugize, J. N., Jewitt, G., and Graham, M. (2018). Effects of land use and land cover changes on water quality in the uMngeni river catchment, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C*, *105*, 247–264. https://doi.org/10.1016/j.pce.2018.03.013

National Planning Commission. (2012). Our future make it work: national development plan 2030. Accessed 15 July 2024. https://www.gov.za/sites/default/files/Executive%20Summary-NDP%202030%20- %20Our%20future%20- %20make%20it%20work.pdf.

Nayak, A., and Bhushan, B. (2022). *Wetland Ecosystems and Their Relevance to the Environment* (pp. 1–16). https://doi.org/10.4018/978-1-7998-9498-8.ch001

Nguyen, C. T., Chidthaisong, A., Kieu Diem, P., and Huo, L.-Z. (2021). A modified bare soil index to identify bare land features during agricultural fallow-period in southeast Asia using Landsat 8. *Land*, *10*(3), 231. https://doi.org/10.3390/land10030231

Nyathi, N. A., Zhao, W., and Musakwa, W. (2020). Land use land cover changes and their impacts on ecosystem services in the nzhelele river catchment, south africa. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*(3), 809–816. https://doi.org/10.5194/isprs-Annals-V-3-2020-809-2020

Obaid, A., Adam, E., and Ali, K. A. (2023). Land use and land cover change in the Vaal Dam catchment, South Africa: A study based on remote sensing and time series analysis. *Geomatics*, *3*(1), 205–220. https://doi.org/10.3390/geomatics3010011

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, *148*, 42–57. https://doi.org/10.1016/j.rse.2014.02.015

Orieschnig, C. A., Belaud, G., Venot, J.-P., Massuel, S., and Ogilvie, A. (2021). Input imagery, classifiers, and cloud computing: Insights from multi-temporal LULC mapping in the Cambodian Mekong Delta. *European Journal of Remote Sensing*, *54*(1), 398–416. https://doi.org/10.1080/22797254.2021.1948356

Ouchra, H., Belangour, A., and Erraissi, A. (2023). Comparison of machine learning methods for satellite image classification: a case study of casablanca using Landsat imagery and Google Earth Engine. *Journal of Environmental and Earth Sciences*, *5*(2), 118–134. https://doi.org/10.30564/jees.v5i2.5928

Ozturk, D. (2015). Urban growth simulation of atakum (Samsun, Turkey) using Cellular Automata-Markov Chain and Multi-Layer Perceptron-Markov Chain models. *Remote Sensing*, *7*(5), 5918–5950. https://doi.org/10.3390/rs70505918

Palanisamy, P. A., Jain, K., and Bonafoni, S. (2023). Machine learning classifier evaluation for different input combinations: A case study with Landsat 9 and Sentinel-2 Data. *Remote Sensing*, *15*(13), 3241. https://doi.org/10.3390/rs15133241

Parashar, D., Kumar, A., Palni, S., Pandey, A., Singh, A., and Singh, A. P. (2024). Use of machine learning-based classification algorithms in the monitoring of Land Use and Land Cover practices in a hilly terrain. *Environmental Monitoring and Assessment*, *196*(1). https://doi.org/10.1007/s10661-023-12131-7

Plessis, A. du, Harmse, T., and Ahmed, F. (2014). Quantifying and predicting the water quality associated with land cover change: A case study of the blesbok spruit catchment, South Africa. *Electronic Journal of Theoretical Physics*, *11*(31), 2946–2968. https://doi.org/10.3390/w6102946

Pontius, R. G., and Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, *32*(15), 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Pullanikkatil, D., Palamuleni, L. G., and Ruhiiga, T. M. (2016). Land use/land cover change and implications for ecosystems services in the Likangala River Catchment, Malawi. *Physics and Chemistry of the Earth*, *93*, 96–103. https://doi.org/10.1016/j.pce.2016.03.002

Rikimaru, A., Roy, P. S., and Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical Ecology (India)*, *43*(1).

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*, *351*(1), 309.

Roushangar, K., Aalami, M. T., Golmohammadi, H., and Shahnazi, S. (2023). Monitoring and prediction of land use/land cover changes and water requirements in the basin of the Urmia Lake, Iran. *Water Supply*, *23*(6), 2299–2312. https://doi.org/10.2166/ws.2023.132

Ruppert, J. C., Harmoney, K., Henkin, Z., Snyman, H. A., Sternberg, M., Willms, W., and Linstädter, A. (2015). Quantifying drylands' drought resistance and recovery: the importance of drought intensity, dominant life history and grazing regime. *Global Change Biology*, *21*(3), 1258–1270. https://doi.org/10.1111/gcb.12777

Saraf, N. M., Lokman, M. F., Abdul Rasam, A. R., and Hashim, N. (2022). Assessment of urban growth changes in Klang District using Support Vector Machine by different kernel. *IOP*

Conference Series: Earth and Environmental Science, *1051*(1), 012023. https://doi.org/10.1088/1755-1315/1051/1/012023

Schoeman, F., Newby, T. S., Thompson, M. W., and Den Berg, V. (2013). South African National Land-Cover Change Map. In *South African Journal of Geomatics* (Vol. 2, Issue 2).

Schultz, B., Immitzer, M., Formaggio, A., Sanches, I., Luiz, A., and Atzberger, C. (2015). Self-Guided Segmentation and Classification of Multi-Temporal Landsat 8 Images for Crop Type Mapping in Southeastern Brazil. *Remote Sensing*, *7*(11), 14482–14508. https://doi.org/10.3390/rs71114482

Schütte, S., and Schulze, R. E. (2017). Projected impacts of urbanisation on hydrological resource flows: A case study within the uMngeni Catchment, South Africa. *Journal of Environmental Management*, *196*, 527–543. https://doi.org/10.1016/j.jenvman.2017.03.028

Soni, S., Singh, H., Qurashi, J., Shuja, M., Pandey, M., and Arora, A. (2022). Landuse/landcover monitoring and spatiotemporal modelling using multilayer perceptron and 'multilayer perceptron'-Markov Chain ensemble models: A case study of Dausa City, Rajasthan. *IOP Conference Series: Earth and Environmental Science*, *1032*(1), 012028. https://doi.org/10.1088/1755-1315/1032/1/012028

Sothe, C., Almeida, C., Liesenberg, V., and Schimalski, M. (2017). Evaluating Sentinel-2 and Landsat-8 data to map sucessional forest stages in a subtropical forest in southern Brazil. *Remote Sensing*, *9*(8), 838. https://doi.org/10.3390/rs9080838

Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M., and Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, *50*(9), 1250–1265. https://doi.org/10.1016/j.asr.2012.06.032

Staponites, L. R., Simon, O. P., Barták, V., and Bílý, M. (2022). Management effectiveness in a freshwater protected area: Long-term water quality response to catchment-scale land use changes. *Ecological Indicators*, *144*. https://doi.org/10.1016/j.ecolind.2022.109438

Stats, S. A. (2011). Statistics South Africa. *Formal Census*, *1*, 79–106.

Story, M., and Congalton, R. G. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, *52*(3), 397–399.

Stosch, K. C., Quilliam, R. S., Bunnefeld, N., and Oliver, D. M. (2017). Managing multiple catchment demands for sustainablewater use and ecosystem service provision. *Water*, *9*(9), 677. https://doi.org/10.3390/w9090677

Strydom, S., Jewitt, G., Savage, M., and Clulow, A. (2020). Long-term trends and variability in the microclimates of the uMngeni Catchment, KwaZulu-Natal, South Africa and potential impacts on water resources. *Theoretical and Applied Climatology*, *140*(3–4), 1171–1184. https://doi.org/10.1007/s00704-020-03127-1

Tena, T. M., Mwaanga, P., and Nguvulu, A. (2019). Impact of land use/land cover change on hydrological components in Chongwe River Catchment. *Sustainability*, *11*(22). 6415. https://doi.org/10.3390/su11226415

Twesigye, C. K. (2011). The Impact of Land Use Activities on Vegetation Cover and Water Quality in the Lake Victoria Watershed. *The Open Environmental Engineering Journal*, *4*(1). https://doi.org/10.2174/1874829501104010066

Viana, C. M., Girão, I., and Rocha, J. (2019). Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region. *Remote Sensing*, *11*(9), 1104. https://doi.org/10.3390/rs11091104

Vinayak, B., Lee, H. S., and Gedem, S. (2021). Prediction of Land Use and Land Cover Changes in Mumbai City, India, Using Remote Sensing Data and a Multilayer Perceptron Neural Network-Based Markov Chain Model. *Sustainability*, *13*(2), 471. https://doi.org/10.3390/su13020471

Wei, P., Lu, Z., and Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering and System Safety*, *142*, 399–432. https://doi.org/10.1016/j.ress.2015.05.018

Winkler, K., Fuchs, R., Rounsevell, M., and Herold, M. (2021). Global land use changes are four times greater than previously estimated. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-22702-2

Woldemariam, G. W., Tibebe, D., Mengesha, T. E., and Gelete, T. B. (2022). Machine-learning algorithms for land use dynamics in Lake Haramaya Watershed, Ethiopia. *Modeling Earth Systems and Environment*, *8*(3), 3719–3736. https://doi.org/10.1007/s40808-021-01296-0

Wu, J. L., Ho, C. R., Huang, C. C., Srivastav, A. L., Tzeng, J. H., and Lin, Y. T. (2014). Hyperspectral sensing for turbid water quality monitoring in freshwater rivers: Empirical relationship between reflectance and turbidity and total solids. *Sensors*, *14*(12), 22670–22688. https://doi.org/10.3390/s141222670

Yao, J., He, X., Chen, W., Ye, Y., Guo, R., and Yu, L. (2016). A local-scale spatial analysis of ecosystem services and ecosystem service bundles in the upper Hun River catchment, China. *Ecosystem Services*, *22*, 104–110. https://doi.org/10.1016/j.ecoser.2016.09.022

Yilmaz, E. O., and Kavzoglu, T. (2022). Classification of jilin-1 gp01 hyperspectral image using machine learning techniques with explainable artificial intelligence. *Intercontinental Geoinformation Days*, *5*, 145–148.

Yulianto, F., Raharjo, P. D., Pramono, I. B., Setiawan, M. A., Chulafak, G. A., Nugroho, G., Sakti, A. D., Nugroho, S., and Budhiman, S. (2023). Prediction and mapping of land degradation in the Batanghari watershed, Sumatra, Indonesia: utilizing multi-source geospatial data and machine learning modeling techniques. *Modeling Earth Systems and Environment*, *9*(4), 4383–4404. https://doi.org/10.1007/s40808-023-01761-y

Zha, Y., Gao, J., and Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*(3), 583–594. https://doi.org/10.1080/01431160304987

Zhang, Q., Liu, M., Zhang, Y., Mao, D., Li, F., Wu, F., Song, J., Li, X., Kou, C., Li, C., and Man, W. (2023). Comparison of machine learning methods for predicting soil total nitrogen content using Landsat-8, Sentinel-1, and Sentinel-2 Images. *Remote Sensing*, *15*(11), 2907. https://doi.org/10.3390/rs15112907