UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Improving land use and land cover monitoring by integrating optical imagery and synthetic aperture radar in fragmented rural landscapes around Nandoni Dam, Limpopo Province, South Africa.

Name: Mutakusi Bashian Ragimana

Supervisor: Prof. Abel Ramoelo

Student Number: 19248297

Date: November 2024

Submitted in fulfilment of the requirements for the degree of Master of Science in Geoinformatics.

Faculty of Natural & Agricultural Sciences

University of Pretoria

Declaration of Originality

I, Mutakusi Bashian Ragimana, declare that the dissertation titled: “Improving land use and land cover monitoring by integrating optical imagery and synthetic aperture radar in fragmented rural landscapes around Nandoni Dam, Limpopo Province, South Africa." , that I am submitting for the degree of Master of Science in Geoinformatics at the University of Pretoria is a result of my own original work and was not previously submitted by me or any other student for a degree at any other tertiary institution.
The contents of information from other sources were paraphrased and the authors were appropriately referenced with the text and as part of the reference list.

Tables, images, and diagrams that form part of this research were of my original work, and in cases where they were obtained from other sources, the contents were properly referenced and acknowledged.

Name: ……
Sign: ……
Date: ……

As the candidate's sole supervisor, I attest that the preceding declaration is true to the best of my knowledge and that this dissertation is ready for submission.
Name: ……
Sign: ……
Date: ……

Acknowledgments

This project serves as an ode, to my mother, Muofhe Molly Mukhethoni, for her unwavering support and unconditional love throughout the years. I also want to humbly express my gratitude to Prof. Abel Ramoelo, thank you for taking in me and supporting me throughout this journey that seemed impossible at first. Through your expertise, guidance and resourcefulness, this project became a success. Finally, I would like to praise myself for my full devotion and hard work. I am immensely proud of what I achieved with this research.

Draft manuscript from this dissertation

Ragimana, M.B, Ramoelo, A, 2024, Enhancing Land Use Land Cover Classification in Fragmented Rural Landscapes: A Study on SAR and Optical Data Fusion.

Abstract

Globally, Monitoring Land Use Land Cover Change (LULCC) is vital as anthropogenic activities continue to reshape the natural environment leading to biodiversity loss and a reduction in ecosystems services. To date most of the studies on LULCC studies have primarily focused on more developed regions in the Northern hemisphere, with less attention given to landscapes in the Southern hemisphere, particularly those in the rural areas that are interspersed and fragmented. The research study had two main objectives: to quantify and monitor land cover changes before and after the construction of Nandoni Dam over a 20 - year period (2001 to 2021); and secondly, to investigate the use of data fusion (Synthetic Aperture Radar (SAR) and optical remotely sensed data) to improve Land Use Land Cover (LULC) classification in interspersed rural area. For the first objective, optical imagery datasets from Landsat 7 and 8 were utilized. Six LULC classes were identified: water, bare ground, agriculture, vegetation, residential areas, and commercial buildings. The Random Forest 9RF) model was employed to classify land covers for the years 2001 (before dam construction) and 2021 (after dam construction). The model's performance was evaluated using Kappa statistics, The Random Forest (RF) model achieved a Kappa score of 0.82 for 2001 and 0.85 for 2021. A significant decrease in the vegetation class coverage was observed from 270 km² to 210 km² over the two decades, raising concerns about biodiversity loss and reducing ecosystem services for the local communities. This research highlights the challenges of classifying land cover in rural areas such as those surrounding Nandoni Dam, where land cover classes are interspersed within vast areas of vegetation. To address these challenges, the study's second objective focused on integrating Sentinel 1 SAR statistical textures with Sentinel 2 optical imagery. The fusion of 2021 SAR and optical data achieved a kappa score of 0.89. However, the study did not find a statistical significance between the average kappa score of using optical and using optical SAR + optical, with a p value of ≈ 0.9023. This study demonstrates the importance of exploring data fusion in improving LULC classification in rural area settings with complex interspaced landscapes. The findings here, provide a basis for better land cover classifications, policy making, and effective land use management using open - source data and data fusion methodologies. The integration of the multiple data sources (here optical imagery and SAR) proves the be a valuable approach for enhancing traditional LULC studies.

Key Words: Remote Sensing, Random Forest, Landsat 7, Landsat 8, Sentinel 1 SAR Textures, Sentinel 2 A, Model Optimization, Heatmaps, Kappa Scores.

Table of Contents

Declaration of Originality ii
Acknowledgments iii
Draft manuscript from this dissertation iv
Abstract v
Chapter 1: Research Introduction 1
1.1 General Introduction 1
1.2 Research Problem 3
1.3 Aims and objectives. 5
1.4 Research Questions 5
1.5 Scope of the study 6
1.6 Research Outline 6
Chapter 2: Literature Review 7
2.1 Rural Dam Developments. 8
2.2 Satellite-Based Remote Sensing for the Natural Environment. 9
2.3 Land Use Land Cover (LULC) Mapping. 12
2.4 Literature Review General Synthesis 23
Chapter 3: Research Methodology 24
3.1 Study Area. 24
3.2 Data Collection. 26
3.2.1 Acquisition of training and validation data 26
3.2.2 Acquisition of satellite remote sensing data (optical) 28
3.3.3 Acquisition of satellite remote sensing data (active) 34
3.3. Data Analysis. 35
3.3.1. Remote Sensing Image Pre-processing. 35
3.3.2. Image Classification Algorithms. 39
3.3.3. Accuracy Assessments and McNemar Test 40
3.3.4 LULC monitoring and change detection. 41
Chapter 4: Results 41
4.1 LULC mapping using Landsat imagery. 41
4.2 LULC mapping using Sentinel 1(SAR) and 2 (Optical) 49
Chapter 5: Discussion 55
Chapter 6. Conclusion and Recommendation 60
6.1. Conclusions 61

6.2 Recommendations 61
References 62
Appendix. 80

Figures and Tables

List of Figures.

Figure 1: Nandoni Dam, Sentinel 2 A view. 25
Figure 2: Landsat 7 and Landsat 8 Band combination; composite displayed through band 3,2,1 for Landsat 7 (2021); and band 4,3,2 for Landsat 8 (2021). 32
Figure 3: Band Spectral Similarity Comparison of Landsat 7 and 8 bands with Sentinel - 2. 33
Figure 4: First objective Workflow. 36
Figure 5: Second Objective Workflow. 38
Figure 6: 2001 and 2021 classified images Using an Optimized Random Forest Classifier. 43
Figure 7: Percentage Landcover Change (Per Class, From- To). 46
Figure 8: Percentage Landcover Change (Per Class, To- From). 47
Figure 9: Landcover Change Class Change (Per Square Km). 48
Figure 10: SAR Extracted Statistical Textures. 50
Figure 11: SAR Textures and Sentinel 2; composite displayed through band 3 for Sentinel 2; and the entropy and energy band for SAR. 51
Figure 12: SAR, and Sentinel 2 Classified Image. 52
Figure 13: Chi-Square Distribution Table. 81

List of Tables

Table 1:Parametric LULC Techniques. 13
Table 2: Non-parametric LULC Techniques. 15
Table 3: Supervised Classification Techniques. 17
Table 4: Unsupervised Classification Techniques. 19
Table 5: Post Classification Analytic Techniques. 20
Table 6: Land Cover Classification Scheme Description based on the South African National Land Cover Dataset; and Number of Polygons used for model Training, Validation and Testing. 27
Table 7: Metadata of Landsat Images Selected for Classification. 29
Table 8: Landsat 7 and 8 Band Descriptions. 29
Table 9: Spectral Information Sentinel 2A. 33
Table 10: Sentinel 1 SAR Description. 34
Table 11: 2001 Land Cover Classification Confusion Matrix. 44
Table 12: 2021 Land Cover Classification Confusion Matrix. 45
Table 13: SAR, and Sentinel 2 Fusion LULC Confusion Matrix. 52

Acronyms and Abbreviation

ANN - Artificial Neural Network.

AVHRR- Advanced Very High-Resolution Radiometer.

DL- Deep Learning.

ESA - European Space Agency.

FLAASH- Fast Line of Sight Atmospheric Analysis of Spectral Hypercubes.

GIS - Geographical Information System.

GRD- Ground Resolved Distance.

GBM -Gradient Boosting Machine.

KNN -K Nearest Neighbour.

LHWP- Lesotho Highlands Water Project.

LULC - Land Use Land Cover.

LULCC- Land Use Land Cover Change.

ML – Machine Learning.

NASA – National Aeronautical Space Agency.

OA – Overall Accuracy.

PA- Precision Accuracy.

RF - Random Forest.

RS – Remote Sensing.

SANLC- South African National Land Cover.

SPUMLA - Spatial Planning and Land Use Management Act of 2013

SAR - Synthetic Aperture Radar.

UA- User Accuracy.

Chapter 1: Research Introduction

1.1 General Introduction

Globally, monitoring Land Use Land Cover (LULC) changes has been important, as anthropogenic activities have rapidly reshaped the natural environment (Chang *et al*, 2018). LULC changes such as deforestation and urbanisation through infrastructure developments fragments natural habitats and disrupt natural processes, leading to biodiversity loss and a decline in the availability of ecosystem services (Sharma *et al*., 2018; Mekuria *et al*., 2021; Mugari *et al*., 2022).

Geographical Information Systems (GIS) and Remote Sensing (RS) have proven to provide powerful technologically driven tools to derive accurate and timely classification information of LULC changes (Reis, 2008; Hanaraj and Angadi, 2020).To classify LULC we depend on classification methods that are largely grouped into supervised and unsupervised classification methods (Love, 2002) .These classification methods produce variations in levels of accuracy of LULC change detection, variations can be attributed to the specific algorithm behind a specific classification method, for example LULC can be classified using methods such, ISO Data, Maximum Likelihood Classifier (MLC), Support Vector Machines (SVM), and Random Forest. Furthermore, other consideration in the performance of a LULC classifier depends on factors such as, the resolution of imagery used, training data, and quality of ground control points (GCPs) (Rojan and Chen, 2004; Hanaraj and Angadi, 2020).

LULC analysis has been accelerated by the rapid developments in earth observations technology, through space-borne or satellite systems (Ayeni *et al*., 2013; Tayyebi *et al*., 2014; Chang *et al*., 2018). However, significant land cover change monitoring that is both timely and precise has widely been focused in developed urban areas of the global north and, to a lesser extent, in interspersed rural areas in the global south (Ayeni *et al*., 2013).

Furthermore, convergence between the fields of data science and remote sensing over the years, has resulted in Machine Learning algorithms for LULC classification increasingly being deployed for land cover change monitoring in numerous regions globally, which makes it interesting to explore land-transforming developments such as Nandoni Dam in rural South Africa (Abdi, 2020). Here, the study shall explore the use of the Random Forest (RF) Machine Learning algorithm in detecting LULC changes around Nandoni Dam before dam construction and at the present day.

The suitability successes of RF for LULC using optical remotely sensed imagery have been well documented in the literature. Amani *et al.,* 2018 achieved an OA (Overall Accuracy) of 92.34% in mapping LULC classes in Indus Basin, Pakistan, with the RF classifier surpassing the OA results achieved by the Support Vector Machines and MLC classifier. These successes extend to areas that experience rapid land-use changes, Talukdar *et al.,*2020, deployed the RF model in North-Eastern India to classify LULC classes with OA achieved here outperforming the K Nearest Neighbour (KNN) classifier. Similarly, Singh and Pandey (2021) observed that the RF model is superior to Artificial Neural Network (ANN) when carrying LULC studies in the Indian Doon Valley, achieving an OA of 88.5% classifying Landsat 8 imagery. These are some of the studies highlighting RF's consistent success to deliver highly accurate LULC modelling tasks.

Furthermore, the success of RF paves the way for exploring its potential in data fusion studies, where information from multiple sources is combined for improved classification. Kaplan and Avdan; 2018 and Lenco *et al*., 2019 suggest that data fusion is becoming increasingly popular in RS studies. While most RS research focuses on urban areas, the potential benefits for rural land cover monitoring like the case here for Nandoni Dam, where classes tend to be more interspersed, warrants a need for this study. The robust performance of RF in LULC classification using optical data suggests that it can be suitable for data fusion applications, particularly in fragmented rural areas that are not well studied.

Limited studies within the Nandoni Dam area have explored the LULC change in the area before and after the dam construction. The area has been a sight of interest in recent years due to the increased protests by community members as a direct effect

of the lack of service delivery promised through the dam development. The failure to carefully plan and monitor land use in the region is a cause of concern for future land cover and usage in the region. This project intends to try understanding how land cover changes are occurring and how to improve land use land monitoring in fragmented rural landscapes by investigating the spatiotemporal utilisation of land in the region using remote sensing techniques as a foundation for the local municipality and landowners to make informed decisions for future land usage in the areas that surround Nandoni Dam.

Consequently, this will give insight into understanding LULC and its effects on the natural environment and social effects in the communities around Nandoni Dam. The outcomes of this project intend to enable the provision of critical inputs to the decision-making of environmental management and planning the future of land use around the dam by the municipality and stakeholders.

This research was motivated by the need for cost-effective and accurate land cover monitoring techniques as anthropogenic activities continuously shape and reshape the environment. Such studies are significant as they provide information for the future planning and management of land, to prevent further degradation of the biophysical environment that provides ecosystem services which nearby rural communities depend on. This study also hopes to influence policy making in terms of the Spatial Planning and Land Use Management Act 16 of 2013 (SPUMLA) regarding developments of large-scale dams and the influences they have on nearby areas.

1.2 Research Problem

Rural areas in South Africa have experienced growth and transformation, from the apartheid era through the transition to democracy and up to the present day. This dynamic change, over time has posed substantial spatial planning challenges (Nel *et al*., 2011; Jayne *et al*., 2014). These challenges have significantly impacted the delivery of services, development of infrastructure, and environmental degradation,

ultimately leading to minimal to no improvements in the quality of life for those residing in rural areas (Harrison *et al*., 2001; Jayne *et al*., 2014).

There are well-documented concerns in the current management and usage of the Nandoni Dam. This study intends to provide an in-depth review of the use of the land around the dam and future implications (Johnson, 2014). In return, enabling the local municipality and landowners around the dam to make informed decisions for future land use planning and help improve current living conditions of the communities in close proximity to the dam.

Furthermore, most remote sensing LULC studies are conducted in more organised urban areas of the global north, as these systems are completely different to complex rural landscapes which exist in a interspersed manner, that require high-resolution remote sensing data to carry out studies and classify, such as the region of interest in this study near Nandoni Dam, South Africa (Trinsci and Turner, 2014; Kanianska *et al*., 2014).

This study has the potential to explore remote sensing applications using both optical and Synthetic Aperture Radar data to assess LULC changes and contribute methodologies for LULC studies for rural areas to the literature. The use of optical imagery is valuable as the data can be openly sourced; in addition, the methodology of using the data for LULC is well documented with high accuracy within the literature. SAR, on the other hand, provides an opportunity to explore newer, less explored ways of classifying landcover by exploiting its textural components for LULC monitoring that might be better suited for fragmented rural landscape which are complex to classify using only optical imagery (Mishra *et al*., 2018)

1.3 Aims and objectives.

Aim:

To monitor LULC change over a 20-year period (Before and after dam construction) and improve LULC classification in fragmented rural landscapes around Nandoni Dam by integrating optical imagery and SAR data.

Objectives:

1. To quantify and monitor land cover class change over a 20-year period from 2001 through the application of an optimized Random Forest algorithm on Landsat optical imagery.

2. To assess the potential of integrating Sentinel 2A optical imagery and Sentinel 1 SAR remote sensing data in improving LULC in rural interspersed landscapes.

1.4 Research Questions

The research questions for this study, based on the objectives. Are as follows:

1. What are the changes in land cover size across six LULC classes around Nandoni Dam over the past 20 years, comparing periods before and after the dam's construction using Landsat imagery for classification?

2. Can optical and active remote sensing data be effectively integrated to improve the accuracy of LULC mapping in fragmented rural areas or landscapes?

1.5 Scope of the study

This study focused on utilising Landsat 7 and 8 imagery to monitor land use and land cover (LULC) changes in the area surrounding Nandoni Dam near Thohoyandou, Limpopo Province, South Africa between 2001 and 2021. Recognizing the importance of accuracy, the study meticulously prioritized extensive optimization of the chosen classification model parameters before running it on the selected imagery.

However, the growing accessibility of Earth Observation data presented another opportunity. The study went beyond just Landsat, venturing into the exploration of multi-sensor data fusion. Specifically, in investigating the potential of combining Sentinel-1 Synthetic Aperture Radar (SAR) data with Sentinel-2A Multispectral imagery for LULC monitoring. This fusion approach holds promise for areas like interspersed rural landscapes, where traditional methods for LULCC and monitoring tends to be a challenge.

1.6 Research Outline

Chapter one
Chapter one provides an introductory outline to the background of the study, and it also highlights the significance of land cover monitoring in interspersed rural areas which are often overlooked in the literature. Furthermore, this chapter gives insight on the importance of using remote sensing-based techniques to effectively monitor LULC changes.

Chapter two
Chapter two focuses on the review of previous work done on LULC changes using remote sensing. This chapter highlights the challenges that come with dam constructions, the significance of the use of remote sensing of the natural environment, and finally mapping landcover change in rural areas.

Chapter three

Chapter three provides a clear description of the study region and details the methodology used in this research. The chapter explains in detail the acquisition of all the datasets used for the study, the data preprocessing methods, description and optimizing the RF classifier, the integration technique used for SAR and optical fusion, and post classification.

Chapter four

Chapter four gives a presentation of the results obtained linked to objective 1 and objective 2 of the study. The results are presented in the form of tables for the confusion matrices, heatmaps for landcover classes interchanges, bar graphs showing land cover changes between 2001 and 2021. Performances of the RF models were also presented for both the objectives of the research.

Chapter five

In chapter 5 the discussion of the study is presented against the main objectives of the study, the achievements of the objectives and aims of this research are presented in an in-depth approach; by leaning into the literature and conceptualizing the successes and importance of this study to LULC effective monitoring in rural landscape and highlights the contribution this study has in the science community.

Chapter six

Chapter six conclude the study by providing context as to how the research was able to achieve its main objectives and also give recommendations for future studies that can be conducted to improve our understanding of LULC in fragmented rural landscapes.

Chapter 2: Literature Review

This chapter provides a review of rural dam constructions in South Africa. The chapter then transcends to the applications of remote sensing in monitoring LULC changes in rural areas. The literature review chapter also discusses parametric and non-parametric algorithms used in remote sensing applications.

2.1 Rural Dam Developments.

Globally, dam constructions result in a variety of benefits such as electric power generation, flood control, recreational activities, fishing opportunities and other benefits (Whitelaw and MacMullan, 2002). Although dams have a vast number of benefits, they have detrimental effects on local ecosystems and can be devastating to the livelihood in nearby communities where they are constructed (Whitelaw and MacMullan, 2002). Dam developments have been linked to biodiversity loss, geographical hazards (landslides, flooding, soil erosion), forced relocation of local communities and to some extent loss of ecosystem services (Obeng, 1977; Kornijów, 2009; Johnson, 2014).

Controversies around dam are attributed to their environmental and social impacts (Scudder, 2001, Altinbilek, 2002; Égré and Senécal 2003; Siciliano *et al*., 2015). It is important to note that ongoing controversies surrounding dam construction globally do not dispute the benefits of these developments, but rather revolve around carefully planning and mitigating the social and environmental consequences of dam constructions (Bird and Wallace, 2001). Such instances were clearly observed around the construction of the LHWP (Lesotho Highlands Water Project), where local rural community members were resistant to the dam development, because they would have minimal benefits from the dam at the expense of the land they depend on for farming and of natural resources (Keketso, 2003).

Bhalla and Mookerjee (2001) suggest that despite the long-standing bad reputation of dams, the primary objectives for dam development should be to combat poverty and improve livelihoods for the people, by ensuring water security. More importantly, improving livelihoods should be a priority for large dams that are constructed near rural areas with limited long-standing lack of basic services provision, such as Nandoni Dam. Furthermore, studies done by Bhalla and Mookerjee (2001) and Ansar (2014) acknowledge that big dam developments in rural areas of India had the potential to be productive from an economic point of view and contribute to agricultural growth and poverty reduction in rural areas, if responsibly managed. As they promote job creation and create fishing opportunities and subsistence farming.

South Africa is fundamentally regarded as a water-scarce country; this is mainly attributed to the climate, as it only receives 492 millimetres of rainfall annually, which by global comparison is close to half of the global average at 985 millimetres of rainfall annually (Colvin *et al*., 2016). This puts pressure on activities that are the backbone of the economy, such as mining and agricultural activities that heavily depend on sufficient and consistent water provision (Writer,2015; Colvin *et al*., 2016).

Population growth, poor maintenance of existing water infrastructure, and economic development are some of the major factors that have led to the national water shortage crisis (Weiss and Slobodian, 2014; Liu *et al*., 2017). To counter the water crisis, the South African government continues to invest money in major dam constructions (Biswas and Tortajada, 2001; Bird and Wallace, 2001).

However, LULC around large dam sites in rural communities needs to be monitored to assess the influence of climate, support rural sustainable development, and influence localised policy decisions. By monitoring LULC in rural areas near dams, we can better quantify the land use changes taking place and promote rural sustainable development.

Remote sensing serves as a watchful eye in large-scale rural dam projects, where environmental and socioeconomic changes have far-reaching consequences. It records land-use changes, tracks water resources, and assesses community well-being (Biswas and Tortajada, 2001; Bird and Wallace, 2001). From deforestation to reservoir levels, its data guides decisions that assure sustainability, reduce environmental damage, and promote the livelihoods of those most impacted. Remote sensing fundamentally directs us toward a future in which dams coexist peacefully with the landscapes they modify (Bird and Wallace, 2001).

2.2 Satellite-Based Remote Sensing for the Natural Environment.

Remote sensing is defined as the acquisition of data from the natural and built environment without any physical touch. It focuses on specialized sensors that detect and record the electromagnetic energy emitted or reflected by objects of interest (Barrett and Curtis, 1999; Campbell and Wynne, 2011).

Different type of remote sensing sensors exists, and can be largely subdivided into passive and active, passive remote sensing relies on solar illumination, where optical sensors then pick up the reflected energy from the objects of interest, this method is mainly used in the acquisition of earth observation remotely sensed data from space - borne satellite systems (Ulaby *et al*., 1981). Active remote sensing, on the other hand, is independent of solar illumination as the sensors transmit energy towards objects of interest and then record the backscatter energy (Ulaby *et al*., 1981; Tsang *et al*., 1985; Chuvieco, 2019).

The rapid advancements in earth observation came in the early 1970s with the launch of the Landsat 1 mission by NASA (National Aeronautical Space Administration), dedicated to global repetitive and continuous earth monitoring through satellite-based remote sensing (Lauer *et al*., 1997). Since its inception in the early 1970s, the Landsat mission has provided the longest continuous global coverage of satellite-based remotely sensed data for close to 50 years (Wulder *et al*., 2008; Chuvieco, 2019).

Since the introduction of Landsat, several commercial and non-commercial space-borne satellites have been launched with varying levels of resolutions and applications. Examples include the Satellite Pour l'Observation de la Terre (SPOT), IKONOS, Quick Bird, RapidEye, Sentinel constellation, Landsat, and Advanced Very High-Resolution Radiometer (AVHRR). Of these, a notable non-commercial satellite-based was the introduction of the Sentinel 2 satellite in 2015 by ESA (European Space Agency) as part of their Copernicus mission, capable of providing remotely sensed data at 10 m resolution as compared to the latest Landsat remotely sensed data at 30 m resolution (ESA, 2022).

The introduction of Sentinel 2A open-source data revolutionized the access to medium resolution in earth observation. However, similarities in Landsat and

Sentinel 2A imagery made both valuable in earth observation applications, as they capture similar bandwidths within the electromagnetic spectrum (Chuvieco, 2019; ESA, 2022). However, the enhanced resolution makes Sentinel data more favourable in remote sensing applications requirements of higher accuracy, such as land use land cover mapping (Kapan and Avdan, 2018). A study carried out by Ghayour *et al*., 2021 found that Sentinel 2 imagery was able to produce an overall LULC mapping accuracy of 94 % and compared to using a Landsat 8 imagery which only had an overall accuracy of 80 %.

Synthetic Aperture Radar (SAR) is an active spaceborne sensor that emits radio and microwave frequencies towards the Earth's surfaces and records the return frequencies (Jan and Ban, 2017; Dino *et al.*, 2019). SAR systems allow for much finer resolution remote sensing data, as they are not limited by atmospheric conditions such as cloud cover and solar illumination (Lenco *et al.,* 2019). Synthetic Aperture Radar (SAR) offers unique capabilities for remote sensing, particularly in challenging conditions. Unlike optical sensors, SAR can penetrate cloud cover, enabling year-round monitoring (Jan and Ban, 2017).

SAR is sensitive to surface roughness and moisture content, providing valuable information for land cover discrimination. In fragmented rural landscapes such as Nandoni Dam, where diverse land cover types are often intermixed, SAR's ability to capture textural information can significantly improve classification accuracy (Al-Ruzouq *et al*., 2019; Nicolau *et al.,* 2021). Sentinel 1 SAR has the primary advantage of combining the radar sensor's all-weather conditions from space, as this is a freely accessible data source with global coverage, the intention of this study is to explore the use of Sentinel 1A SAR fused with Sentinel 2A optical imagery, in LULC around fragmented rural landscapes of Nandoni Dam.

Sentinel 2A has spectrally rich information in the visible-near infrared spectrum, and the short revisit period of both satellites makes it worthwhile to investigate the fusion of the two data sources for LULC studies (Al-Ruzouq *et al.,* 2019). Because SAR and optical remote sensing data provide complementary information, land cover classification tasks can benefit from the fusion of both data types, resulting in

increased mapping accuracy overall (Haas and Ban, 2017; Kaplan and Avdan, 2018; Lenco *et al.,* 2019).

Combining SAR and optical data offers a synergistic approach to LULC mapping, leveraging the strengths of both data sources. Optical data provides detailed spectral information about land cover types, while SAR offers complementary textural and structural information (Kaplan and Avdan, 2018; Al-Ruzouq *et al*., 2019; Lenco *et al.,* 2019). By fusing these data sources, it is possible to improve classification accuracy, particularly in complex and heterogeneous landscapes. However, challenges such as spatial and radiometric differences between the two data sources must be carefully addressed through appropriate preprocessing and fusion techniques (Al-Ruzouq *et al.,* 2019;)

Using remote sensing, successful large-scale accurate models of LULC of the natural and built environment can be achieved with simple straight forward techniques (Wulder *et al*., 2008). Application of remotely sensed data can be applied in numerous fields, as mentioned above, this is largely dictated by the resolution of the data in relation to the specific field of interest, for example, meteorology, water quality, land surveying, geology, aviation, ecology, engineering, oceanography, and military are some of the fields that employ remote sensing techniques (Hutchinson, 1991; Barrett and Curtis, 1999; Schumann *et al*., 2009; Horning *et al*., 2010).

2.3 Land Use Land Cover (LULC) Mapping.

A key application of remote sensing has been monitoring land-use change and land cover; this has largely been attributed to the rise in demand for land use conversion as a direct impact of the rise in global population (Hanaraj and Angadi, 2020; Mansour *et al*., 2020). The definition of LULC is often confusing to understands, as it is made ups of two separate terms, firstly “Land Use” and secondly “Land Cover” (Sewenet and Abebe, 2017). “Land Use” refers to the utilisation and modification of land cover through anthropogenic activities, while “Land Cover” Land cover refers to the physical material that covers the Earth's surface, such as vegetation (forests,

grasslands, shrubs), water bodies (oceans, lakes, rivers), bare soil, ice/snow, and human-made structures (cities, roads, agricultural fields). (Sewenet and Abebe, 2018; NOAA, 2024).

Monitoring LULC changes through remote sensing in a specific area is necessary for the context of implementing sustainable land-use practices and future land use management planning policies (Balakeristanan and Said, 2012; Sewenet and Abebe, 2018; Mansour *et al*., 2020).

The monitoring of LULC plays a significant role in the study of global environmental change as more land is needed to accommodate humans for needs such as housing, industries, recreational activities, and agriculture (Ligtenberg, 2017). On the other hand, the alteration of the natural state of the land through human modifications have resulted in large-scale deforestation, biodiversity loss and global warming, leaving those in the poorest of regions more vulnerable than ever (Reis, 2008; Liping and Yujun, 2018). Such threats have led to an increased interest in the development of effective technological tools for timely, effective, and accurate LULC monitoring and management.

LULC models fall into parametric and non-parametric data analysis techniques. Parametric techniques are those that use distributional assumptions to estimate the parameters of the distribution assumed for the data (Jensen, 2005; Sterk *et al*., 2007). Parametric LULC classifiers examples and description of these are shown in Table 1 below:

Table 1:Parametric LULC Techniques.

Name	**Description**	**Reference**
Iterative Self-Organizing Data Analysis Technique (ISOData)	A clustering algorithm that uses an iterative process to distribute input	(Tou and Gonzalez, 1974; Jensen, 2005; Enderle

	data into similar clusters based on their position in the nth dimension of spectral space.	and Weih, 2005; Al Doski *et al*., 2013)
K means	Aims to divide data into k clusters so that data points in the same cluster are similar and data points in different clusters further apart.	(Jensen, 2005; Hamerly and Elkan,2003; Likas *et al*., 2003; Steinley, 2006; Al Doski *et al*., 2013)
Maximum Likelihood	Defines decision boundaries between clusters in input data using probability of inputs belonging to a specific cluster.	(Sisodia *et al*., 2014; Bouaziz *et al*.,2017)
Parallelepiped	Determines and defines a distinguishing hyperbole statistic within each land cover class.	Jensen, 2005
Minimum Distance to Mean	Calculates the distance between a pixel's spectral signature and the mean per class.	(Schowengerdt, 2006; Congalton and Green, 2019)

On the other hand, non-parametric data analysis techniques are defined as data analysis techniques that do not necessitate making prior distributional assumptions about the data (Cortijo and De La Blanca, 1997; Sterk *et al*., 2007; Altman and Bland, 2009; Al-Doski *et al*., 2013; Tayyebi *et al*., 2014). The examples for non-parametric are given with descriptions in Table 2 below:

Table 2: Non-parametric LULC Techniques.

Name	**Description**	**References**
Support Vector Machines (SVM)	A type of Machine Learning technique, aimed to tackle and resolve binary classification problems by separating data into various classes using a plane.	(Cortes and Vapnik, 1995; Pal, 2005; Bouaziz *et al*.,2017; Abdi, 2020; Binbusayyis and Vaiyapuri, 2020; Eilertsen *et al*., 2020; Talukdar *et al*., 2020b)
Random Forest (RF)	A type of Machine Learning data analysis technique that is based on an ensemble of decision trees to classify and find meaningful patterns within input data.	(Ma *et al*., 2004; Millard and Richardson, 2015; Belgiu and Drăguţ 2016; Veronesi and Schillaci, 2019; Talukdar *et al*., 2020b)
Artificial Neural Networks (ANN)	Functions like the human brain or nervous system where nerve fibres are interconnected allowing the ability for the model to learn and produce patterns and results from given examples.	(Tayyebi, *et al*., 2014; Talukdar *et al*., 2020b)
Classification and Regression Trees (CART)	Partitions data into decision trees, based on a given set of binary splits.	(Pal and Mather, 2003; Pedregosa *et al*., 2011)
Boosted Tree	Combines weak decisions	(De'Ath, 2006; Pedregosa

	trees to form a enable of stronger splits within the data.	et al., 2011; Friedman 2020)
Nearest Neighbors with Rejection (NNR)	Functions by rejecting pixels with high k-nearest neighbor distance, which potentially representing outliers.	(Aha and Kibler, 1991; Cortijo *et al*., 1995)

While non-parametric methods have the advantage of not requiring assumptions about the distribution of the data like parametric techniques, they have the disadvantage of being primarily suited to hypothesis testing and yielding no useful estimates (Cortijo and De La Blanca, 1997; Altman and Bland, 2009; Tayyebi *et al*., 2014).

Remote sensing has made monitoring of LULC more effective through the deployment of various land cover classification techniques. These techniques can be grouped into supervised (See Table 3) and unsupervised (See Table 4) classifications. In supervised image classification, pixel categorization is based on user-defined numerical descriptors (known as training areas) of the various land cover types (Sisodia *et al*., 2014).

The training areas describe the spectral attributes for each feature type of interest. Each pixel is then compared to each category in the training area and assigned the category that “it almost looks like”, Supervised learning takes place when the training examples are labelled with their known outputs (Ma *et al*., 2004; Millard and Richardson, 2015; Belgiu and Drăguţ, 2016; Veronesi and Schillaci, 2019; Talukdar *et al*., 2020b).

Maximum likelihood classification (MLC) is the most widely used traditional type of supervised classification (Sisodia *et al*., 2014; Bouaziz et al., 2017). In MLC, a pixel is assigned to a class according to its probability of belonging to a particular class,

which its mean, and covariance is modelled as forming a normal distribution in multispectral features (Sisodia *et al*., 2014).

The main advantage of supervised over unsupervised is how there is a prior distinction of useful information and then spectral separation, while in unsupervised classification, spectral separation occurs before the required information can be defined (Lillie and Keifer, 2014).

In recent years there has been a consideration of use from traditional forms of land cover classification techniques to those based on Machine Learning (Rajasekaran and Saravanan, 2020). Machine Learning methods are a computerised modelling approach capable of learning patterns from the data to make automatic decisions without explicit programming rules by an individual (Lantz, 2019; Rajasekaran and Saravanan, 2020). Machine Learning utilizes experiences or example scenarios to discover underlying structures, similarities, or dissimilarities present in data to explain or classify a new experience properly (Ma *et al*., 2004; Singh *et al*., 2016). Types of machine learning classifications include Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting Machine (GBM), and Deep Learning (DL).

Table 3: Supervised Classification Techniques.

Technique	**Description**	**Advantages**	**Disadvantages**	**References**
Support Vector Machines (SVM)	Locates the optimal hyperplane that distinguishes classes in high dimensional data.	Performs well for small datasets, which are high dimensional. Produces clear class boundaries.	Sensitive to data scaling, tends to be computationally expensive.	Cortes and Vapnik, 1995.
Random Forest (RF)	Integrates decision trees	Highly accurate	Computationally expensive to	Breiman, 2001.

	where in each tree selects on the class of a data point. The final prediction is the most frequent tree selection.	outputs, handles mixed data types and high dimensional dataset with ease.	train the mode. Need prior hyperparameters tuning, making it prone to overfitting.	
Gradient Boosting Machine (GBM)	Constructs decision trees that focus of correcting errors of the previous tree.	Can manage complex non-linear relationships within data.	Prone to mode overfitting. Computationally expensive to train.	Friendman, 2001.
Convolutional Neural Networks (CNNs) (Deep Learning)	Utilizes artificial neural networks within multi-dimensional data to learn complex patterns that exist within the data.	Highly equipped to manage complex higher dimensional data.	Requires large amount of data to train, extremely computational expensive.	LeCun *et al.,* 2015.

In unsupervised image classification, the algorithm running stage and the allocation of class names to the resulting clusters are dependent on similarities in pixel attributes (See Table 4). The algorithms used for classification include ISODATA and K-means clustering techniques (Al Doski *et al*., 2013). The unsupervised techniques (ISO Data and K means) were used in classifying land cover in Abbotabad Pakistan, where result indicated that both techniques have the potential of successfully

classifying land cover depending on the complex nature of the study area and data used (Abbas *et al., 2016)*

Table 4: Unsupervised Classification Techniques.

Technique	**Description**	**Advantages**	**Disadvantages**	**References**
ISO Data	A clustering algorithm that uses an iterative process to distribute input data into similar clusters based on their position in the nth dimension of spectral space.	No need to predefined classes, adapts data distribution.	Require post class assigning. Difficult to interpret.	(Enderle and Weih, 2005; Jensen, 2005; Al Doski *et al*., 2013; Hemalatha and Anouncia, 2017)
K Means	Aims to divide data into k clusters so that data points in the same cluster are similar and data points in different clusters further apart.	Fast and computationally efficient.	Require post class assigning. Difficult to interpret.	(Likas *et al*., 2003; Steinley, 2006; Cheng and Wei, 2019; Al Doski *et al*., 2013)

The defining distinction between the two types of classification is that for supervised image classification, ground truth data or prior knowledge of the area is needed, while for unsupervised classification, no prior knowledge is needed (Hasmadi *et al*., 2009; Al Doski *et al*., 2013).

Post LULC classification analysis can be carried out to understand LULCC, where classified images from different dates compared to identify changes per class this is achieved by area calculation to determine the proportion of change, change detection to locate areas of change, and spatial analysis to identify hotspots (Jensen, 2005, Mi *et al*., 2019). Time series analysis can also be utilized to recognize patterns and trends and understanding the dynamics of change can be aided by land cover transitions.

Possible post classification analyses (See Table 5) include accuracy assessment, which focused on understanding the level of agreement between a classified output and referenced data; fragmentation analysis; which tries to understand the spatial arrangements of the desired classes between each other; land cover diversity, which interrogate the range of classes that exist within a given landscape; land use suitability analysis, that tries to analyse the potential of the land use activities that can exist on a specific land; and landscape change trajectory analysis, focuses on the temporal change and driver of change on a landscape (Jensen, 2005; Thonfeld, 2020).

Table 5: Post Classification Analytic Techniques.

Technique	**Description**	**Reference**
Accuracy Assessment	Evaluates the agreement between classified LULC and ground truth/ reference data. Ensure reliability of the LULC results but identifying potential biases within the	(Cheng *et al.,* 2019; Congalton and Green, 2019; Wickham *et al.,* 2023)

	model when separating the classes. The main indicators in an Accuracy Assessments are Overall accuracy (OA), which represents the total number of pixel that were correctly classified; User's Accuracy (UA), indicates the proportion of pixels within a reference class (e.g., vegetation) that were actually classified correctly as that class; Producer's Accuracy, This represents the proportion of pixels classified as a particular class (e.g., vegetation) that actually belong to that reference class. These indicator play a role in this study to evaluating the effectiveness of the chosen classification methods used; and lastly Cohen's Kappa score, which is a measure of agreement between two raters or observers who classify items into mutually exclusive categories.	

Fragmentation Analysis	Evaluates the spatial configuration of the classes in the LULC output, by understanding the patch sizes of the classes, shapes complexities, and interconnectedness.	(Gulinck and Wagendorp, 2002; Caranza *et al.,* 2023)
Land Cover Diversity Analysis	Evaluates the variety and abundance of different LULC class type within a landscape.	Walz, 2015
Land Use Suitability Analysis	Evaluates the suitability of the land for specific use, e.g., urban development's environmental protected areas.	(Joerin *et al.,* 2001; Luan *et al.,* 2021)
Landscape Change Trajectory Analysis	Evaluates the temporal patterns and trends of LULC in a specified landscape.	Käyhkö and Skånes, 2006
Time Series Analysis	Evaluates the temporal patterns and trends of LULC in a specified landscape, by understanding the rate, drivers, and trend of change.	Matosak *et al.,* 2023

LULC classification in remote sensing relies heavily on the ability to extract meaningful information from spectral data. However, more that often individual sensors often have limitations, such as coarse spatial resolution or limited spectral

bands (Joshi *et al*., 2016; Chen *et al*., 2017). To counter this data fusion can be applied, which is a powerful technique that integrates information from multiple sources, to create a more comprehensive dataset for remote sensing applications (Du *et al*., 2021)

Data fusion can significantly enhance LULC classification processes by offering several advantages. Firstly, it increases the information content, providing a more complete picture of the LULC. Secondly, it improves classification accuracy by overcoming limitations that are often inherent in individual datasets. Finally, data fusion can even create products with higher spatial resolution while preserving spectral details (Zhong *et al*., 2017; Mo *et al*., 2019; Zhang *et al*., 2020; Chachondhia *et al.*, 2021).

However, Data fusion does come with challenges. Fusing data requires specialized techniques and software, increasing the complexity of the analysis. Additionally, ensuring compatibility between datasets from different sources with varying formats, resolutions, and acquisition times requires careful preprocessing (Joshi *et al*., 2016; Chen *et al*., 2017). Furthermore, complex fusion algorithms can be computationally expensive, especially for large datasets (Chen *et al*., 2017). Despite these challenges, data fusion remains a valuable tool, offering the potential to significantly improve the quality and information content of LULC classifications derived from remote sensing data (Joshi *et al*., 2016; Mo *et al*., 2019; Zhang *et al*., 2020)

2.4. Literature Review General Synthesis

The literature review section for this study examined the multifaceted impacts of dam development, the potential of remote sensing in LULC monitoring, and the unique challenges faced in rural areas. While acknowledging the benefits of dams, such as water supply and power generation, the review highlights their potential negative consequences, including biodiversity loss, displacement, and disruption of ecosystem services.

Remote sensing, particularly satellite imagery, emerges as a valuable tool for tracking these changes and informing sustainable decision-making. The integration of optical and SAR data is emphasized, especially in fragmented rural landscapes like Nandoni Dam, where SAR's cloud-penetrating capabilities and surface texture information complement optical data's spectral information.

This literature review recognizes the complexity of rural land cover and the need for tailored methodologies to address the challenges posed by interspersed land types. Accurate LULC monitoring is crucial for understanding the environmental impact of human activities, implementing sustainable land management practices, and informing future land use planning and policy-making to prevent further degradation.

Chapter 3: Research Methodology

This chapter starts off by introducing the Nandoni Dam study area, and then proceeds to explain the data collection techniques and the analysis methods used. Figure 4 and 5, describe the overall workflows followed to achieve the 2 objectives of this research.

3.1 Study Area.

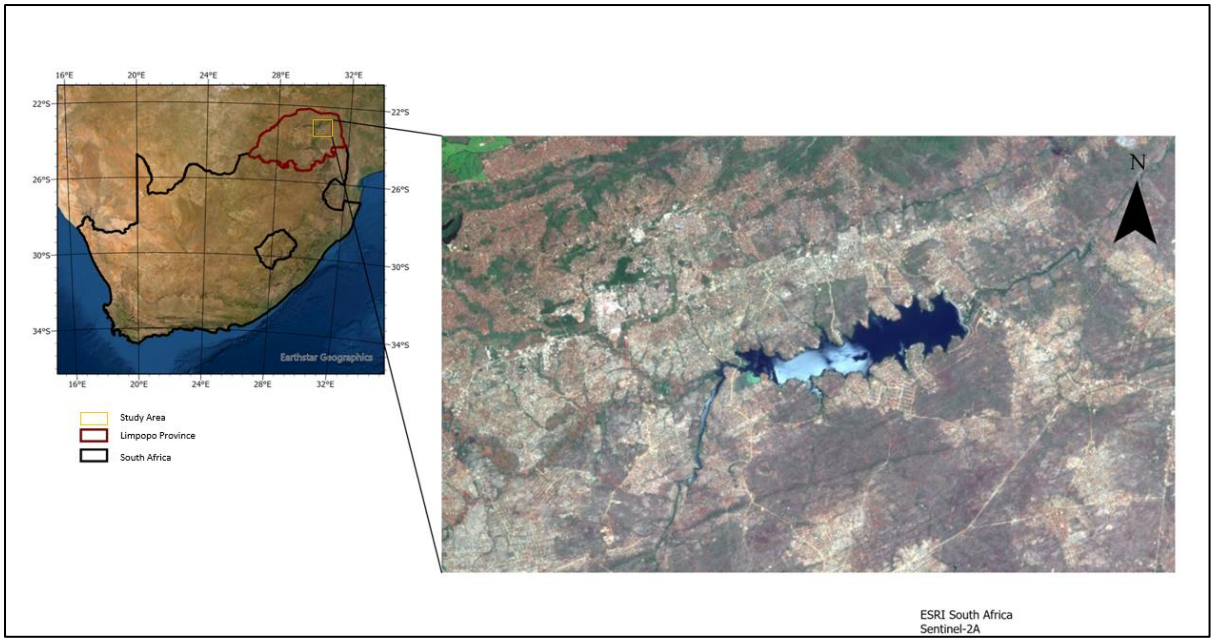


Figure 1: Nandoni Dam, Sentinel 2 A view.

Nandoni Dam is situated in the Limpopo Province of South Africa, with geographic coordinates of 22°59′20″S 30°36′27″E (See Figure 1). The regional vegetation falls within the Granite Lowveld (code: SVI3) vegetation type, within the Savanna biome, the vegetation features are described as tall shrubland with a dominance of *Terminalia sericea, Combretum zeyheri* and *Combretum apiculatum* (Mucina and Rutherford, 2006). The region is underlain by a geology of stable granite rocks of the ancient Kaapvaal craton (Mucina and Rutherford, 2006). The region occurs in altitudes ranging between 250 - 700 m above sea level and is found within the tropics characterised by summer rainfall (400-900 mm mean annual rainfall) and dry Winter season (Mucina and Rutherford, 2006).

Situated on the Luvuvhu River, Nandoni Dam was previously known as Mutoti Dam (Tempelhoff, 2007). The Thulamela Local Municipality region, where Nandoni Dam is located, is praised for its abundant biological diversity of flora and fauna. It borders the famous Kruger National Park and other numerous nature reserves in the area. This rich biodiversity can be attributed to the geographical location and diverse topography (Nghonyama, 2011).

Thulamela Local Municipality covers a vast tract of land that is tribal, and Thohoyandou is its political, administrative, and commercial centre (Nghonyama, 2011). Thohoyandou is the closest town to Nandoni Dam and lies 10 km away. According to the 2011 national census Thulamela Local Municipality had a population of 618 462 people, majority of the population are Black Africans aged between 15-34 years (Nghonyama, 2011). The municipality boasts around 130321 households living mostly in a rural setting (Nghonyama, 2011).

The village communities that surround Nandoni Dam form part of South Africa previously known as the Venda homeland, which was created by the South African apartheid government in 1962 for Tshivenda speaking people (South African History Online, 2011). Homelands or locally known as The Bantustans, were areas created by the South African apartheid government wherein people belonging to Black ethnicity groups were moved to avoid them from living in more developed areas of South Africa (South African History Online, 2011).

Venda was the smallest of all homelands created. Majority of people from Venda were migrant workers in major industrial areas of South Africa. The economy of the Venda homeland was dependent on coal mining and agricultural activities (South African History Online, 2011). Elevated levels of poverty, government maladministration of resources, contestations of chieftainship in rural areas of Venda such as neighbouring communities of Nandoni Dam, Limpopo are direct legacies of Bantustans (Ally and Lissoni, 2012, Julian and Wilson, 2013).

3.2 Data Collection.

3.2.1 Acquisition of training and validation data

For this study, a combination of high-resolution satellite imagery and local knowledge of the Nandoni Dam area was used to gather training and validation data for performing the classification. Utilizing the ESRI Sentinel Explorer, Sentinel 2 imagery was accessed and analysed to identify and digitize representative samples for each

LULC class, including water, bare ground, residential, commercial buildings, agriculture, and natural vegetation (Ramanamurthy and Victorbabu, 2021; Basheer *et al*., 2022; Ryzhok, 2023; kusumaningrum and Priyana, 2024). Familiarity with the area facilitated accurate delineation of these classes directly on the satellite images. The interactive capabilities of the ESRI Sentinel Explorer allowed for precise digitization of training data, ensuring a detailed and comprehensive set of training data (Ryzhok, 2023). The breakdown of the number of polygons drawn for training, testing, and validation the chosen classifier model is shown in Table 6 below. This approach provided high-quality spatial and spectral information, enhancing the performance and reliability of the LULC classification model while maintaining the low costs in acquiring data, as the study aims to use only open-source data sources (Ryzhok, 2023; kusumaningrum and Priyana, 2024).

This study used six LULC classes around Nandoni Dam (agriculture, vegetation, residential, commercial buildings, and water). The SANLC description of these selected land cover classes and their corresponding number of training and validation polygons is presented in Table 6. According to Wolf *et al*., 2005: Sohn *et al*.,2020, limiting the number of classes when performing LULC is important to enhance the classifier's performance. So, for this reason, the six classes, namely bare ground, water, residential, Commercial buildings, agriculture, and Natural vegetation, were used for this study and described in Table 6 below.

Table 6: Land Cover Classification Scheme Description based on the South African National Land Cover Dataset; and Number of Polygons used for model Training, Validation and Testing.

Land cover classes	**Description**	**Number of polygons used for training.**	**Number of polygons used validation**	**Number of polygon used for testing**

Vegetation	Areas consisting of untouched vegetation, grasslands, savanna shrublands.	**7**	**3**	**3**
Bare land	Areas with no vegetation cover, brick making quarries, uncultivated agricultural lands.	**26**	**6**	**6**
Water	Artificial Dams, natural rivers, streams, and ponds.	**2**	**2**	**2**
Residential	houses, apartments, and associated features.	**8**	**2**	**3**
Agriculture	Irrigated area with distinct patches of designated farming activities (subsistence and commercial).	**7**	**6**	**6**
Commercial buildings	Commercial services rendering places and associated features.	**41**	**11**	**10**

3.2.2 Acquisition of satellite remote sensing data (optical)

Landsat 7 and 8 images were selected for this study as the mission contains a catalogue of the longest running global earth observation data spanning over 50 years (since 1973) (Wulder *et al*., 2008). Six bands from Landsat 7 and Landsat 8 were matched according to the similarity of the bandwidth they captured in the separate sensors (See Figure 2, of the completed band combination composites). A description of the specific bands used is given below (Table 7 and Table 8). It is important to note that the selection of dates from separate satellites is a result of no Landsat 8 images before the year 2013, and although Landsat 7 satellite still delivers earth observation, it has left most of the study area with a slicing issue resulting in

some areas with no observations when exploring its datasets in recent times in the study area of this project (See Figure 12 in the addendum for the example of this over the study area).

A description of the dates chosen for acquiring the data, is shown in table 7, These two dates were deemed appropriate for this study as they cover a period before the dam was constructed and a period after the dam was constructed. The month of May was chosen because it represents the best, nationally applicable, single-window period for landscape interpretation in South Africa, in terms of separating natural vegetation covers (i.e. woody from herbaceous) and cultivation activities (Fairbanks *et al*., 2000). It was important to select dates that fall close to each other to reduce the effects of seasonality on surface reflectance on vegetation specifically when conducting land cover change analysis (Fairbanks *et al*., 2000).

After selecting the dates, the data was downloaded for the two dates from the USGS Earth Explorer site (https://earthexplorer.usgs.gov/). Table 7 describes the cloud cover, sun elevation angle, sun azimuth, and scene collection time. And Table 8 describes the chosen bands used for this study for Landsat 7 and 8.

Table 7: Metadata of Landsat Images Selected for Classification.

	Landsat 7 (21 May 2001)	Landsat 8 (20 May 2021)
Cloud cover	0	0,1
Sun elevation angle	36,52	38,03
Sun azimuth	38,4	36,18
Scene time	07:45 AM	07:54 AM

Table 8: Landsat 7 and 8 Band Descriptions.

Attributes	Specifications (Landsat 7)	Specifications (Landsat 8)
Spectral Resolution (µm) (Per Band)	Band 1, 0.45-0.52 Band 2, 0.53-0.61 Band 3, 0.63-0.69 Band 4, 0.78-0.90 Band 5, 1.55-1.75 Band 6, 2.09-2.35	Band 2, 0.45 – 0.51 Band 3, 0.53 – 0.59 Band 4, 0.64 – 0.67 Band 5, 0.85 – 0.88 Band 6, 1.57 – 1.65 Band 7, 2.11 – 2.29
Spatial Resolution (metre)	30 x 30	30 x 30
Temporal Resolution (revisit in days)	16	16
Spatial Coverage (swath) (km)	183 x 170	185
Altitude (km)	705	705
Inclination	98.2°	98.2°

21 May 2001 (Landsat 7)

21 May 2001 (Landsat 7)	N 0 1 3 5 8 10 Km Scale: 1:152,600

20 May 2021(Landsat 8)

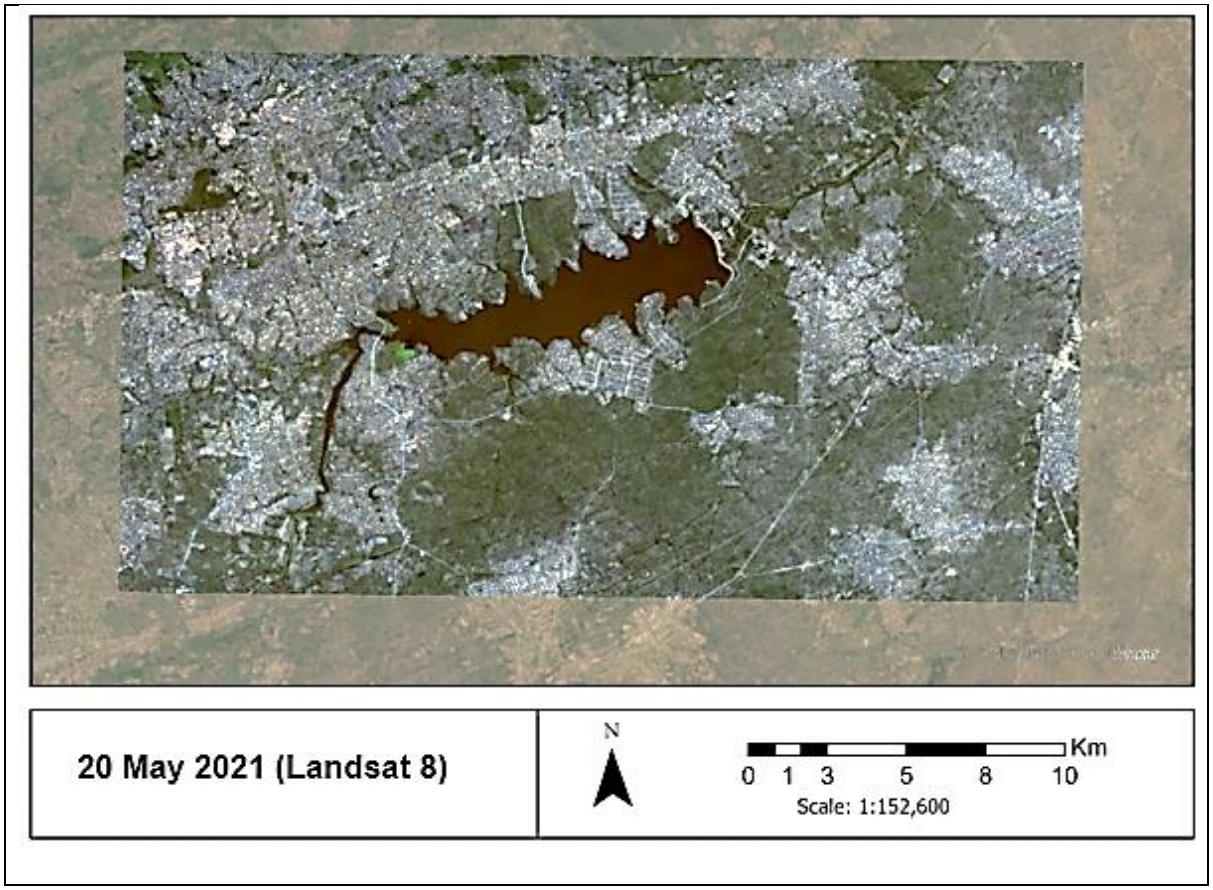


Figure 2: Landsat 7 and Landsat 8 Band combination; composite displayed through band 3,2,1 for Landsat 7 (2021); and band 4,3,2 for Landsat 8 (2021).

Sentinel 2A remotely sensed data were collected for the date of May, with the extraction of bands 8, 4,3,2 with spatial resolutions of 10 m, from the Sentinel 2A optical imaging dataset was used (See Table 9). These 4 bands were selected as they coincided summarily to the bands spectral information used for Landsat 7 and 8 imagery for the first objective of the study. This comparison can be seen on USGS spectral information comparison between the 3 satellites was compared on figure 3 below. Sentinel 2A data datasets are freely accessible and downloadable to the public through the Copernicus ESA open access hub. See Table 9 below for sentinel 2 A spectral information.

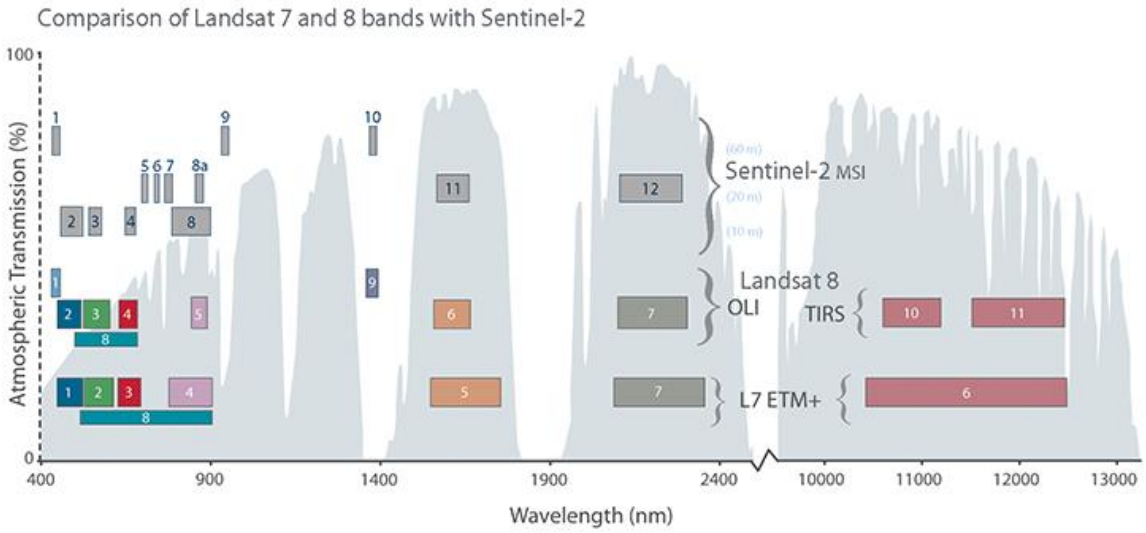


Figure 3: Band Spectral Similarity Comparison of Landsat 7 and 8 bands with Sentinel - 2.

Table 9: Spectral Information Sentinel 2A.

Attributes	Specifications - Sentinel 2A
Spectral Resolution (nm) (Per Band)	Band 2, 458 - 523 Band 3, 543 - 578 Band 4, 650 - 680 Band 8, 785 - 899
Spatial Resolution (metre)	10 ✕ 10
Temporal Resolution (revisit in days)	5 days
Spatial Coverage (swath)	290 km

(km)	
Altitude (km)	786 km
Inclination	98.62°

3.3.3 Acquisition of satellite remote sensing data (active)

A Sentinel 1 SAR Ground Range Detected (GRD) Interferometric Wide Swath (IW) image with 5 m ✕ 20 m spatial resolution datasets is freely accessible and downloadable to the public through the Copernicus ESA open access hub (https://dataspace.copernicus.eu/). The dataset's dates within May 2021 coincide with the dates used to acquire Landsat 7 and 8 and Sentinel 2A optical imagery (See Table 10 for Sentinel 1 SAR data description).

Table 10: Sentinel 1 SAR Description.

Attributes	Specifications (Sentinel 1 SAR, Interferometric Wide Swath (IW))
Spatial Resolution (metre)	5 ✕ 20
Temporal Resolution (revisit in days)	12 days
Spatial Coverage (swath) (km)	250 km
Altitude (km)	693 km

Inclination	98.18°

3.3. Data Analysis.

3.3.1. Remote Sensing Image Pre-processing.

For the first objective (See Figure 4 for a complete workflow), Landsat (Level 2 from Landsat 7 and Landsat 8) remotely sensed data was radiometrically and spectrally calibrated to rectify and remove the geometric and radiometric errors that come from motions of satellite aircrafts housing these sensors. This process is important and needs to be done before the image analysis (here LULC classification) to reduce data uncertainty (Tansock *et al*., 2015). ENVI software for image analysis was used to calibrate the images. Furthermore FLAASH (Fast Line of sight Atmospheric Analysis of Spectral Hypercubes) on ENVI was used to correct atmospheric effects on spectral radiance of the satellite imagery due to absorption and scattering from atmospheric particles, this process is critical in reducing the influence of the atmosphere on the quality of the data where small reflectance differences might lead to significant effects/errors on image classification analysis (Yuan and Niu, 2008; Owojori and Xie, 2005; Nazeer, 2014; López-Serrano, 2016).

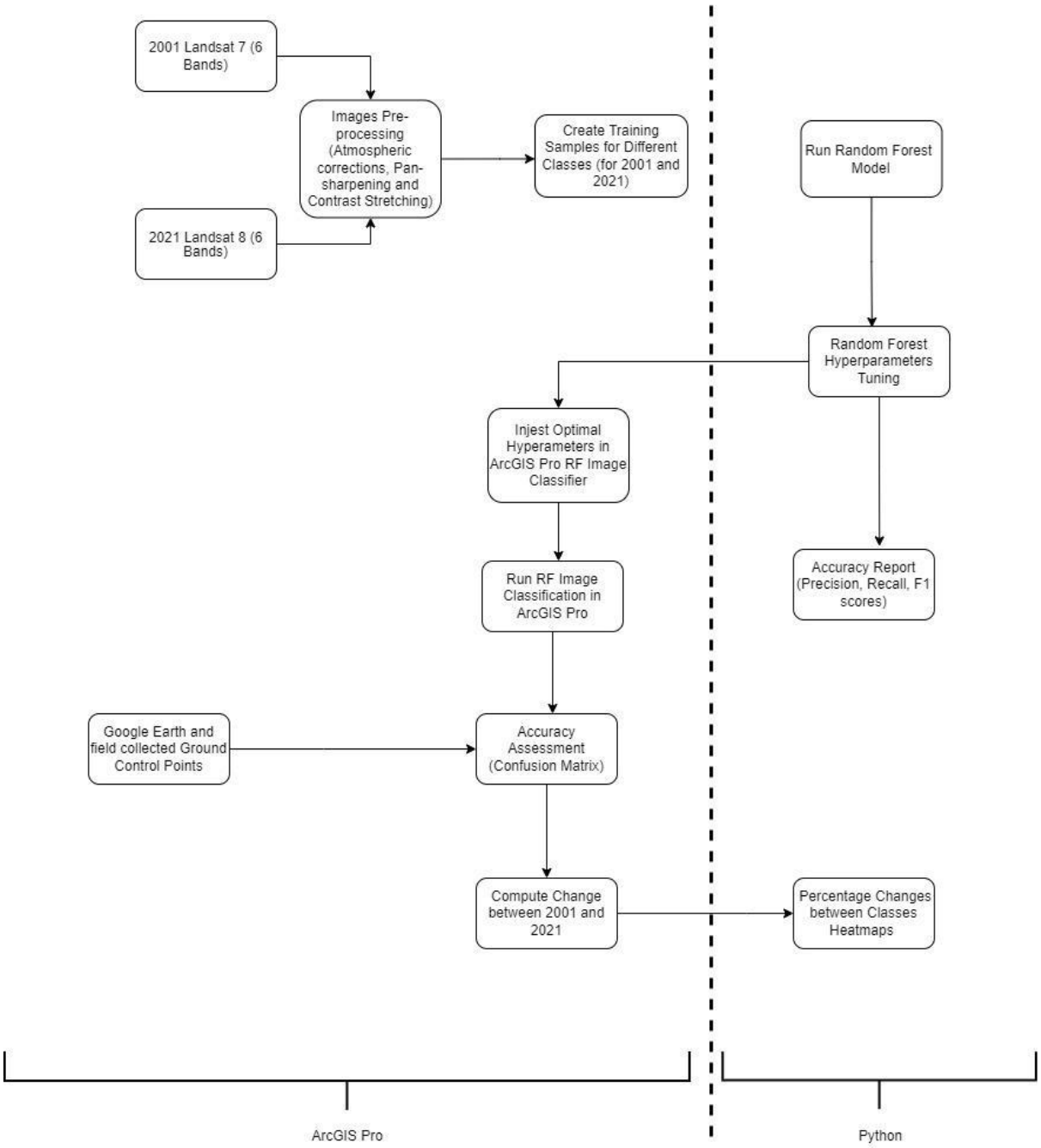


Figure 4: First objective Workflow.

Two image appearance enhancement techniques (pan-sharpening and contrast stretching) were explored for the two Landsat images (Level 2 from Landsat 7 and Landsat 8), this process is fundamental in satellite remotely sensed data to improve visual interpretation and image analysis (Sharma *et al*., 2021). Pan sharpening integrates a finer resolution panchromatic image with a medium multispectral image to give an improved resolution of the multispectral image (shah *et al*., 2008;

Rahaman *et al*., 2017; Özay and Tunga, 2021). Pan sharpening was done by using a Sentinel 2A panchromatic image band (10 m) on the Landsat imagery (30 m) where in the Landsat imagery. Contrast stretching, on the other hand, is useful for improving the illumination of the remotely sensed images; this process improves differences in brightness over a range of grey values for better visualization (Rahaman *et al*., 2017).

For the second objective (See Figure 5 for complete workflow), fusing Sentinel 1 SAR and Sentinel 2A, the data covering the entire study area was downloaded from the ESA's Sentinel Scientific Hub (ESA, 2022). SAR data pre-processing of the Level-1 GRD image was performed using the SNAP Sentinel Application Platform toolbox (ESA, 2022) standard processing methods (calibration, thermal noise removal, TOPSAR deburst, Speckle filtering), as well as resampling to 10 × 10 m spatial resolution and multilooking parametrization for both VV and VH polarization images and extraction of the statistical textural components inherent within SAR Remotely Sensed data (Clerici *et al*.,2017; Haas and Ban, 2017; Kaplan and Avdan, 2018; Lenco *et al*., 2019).

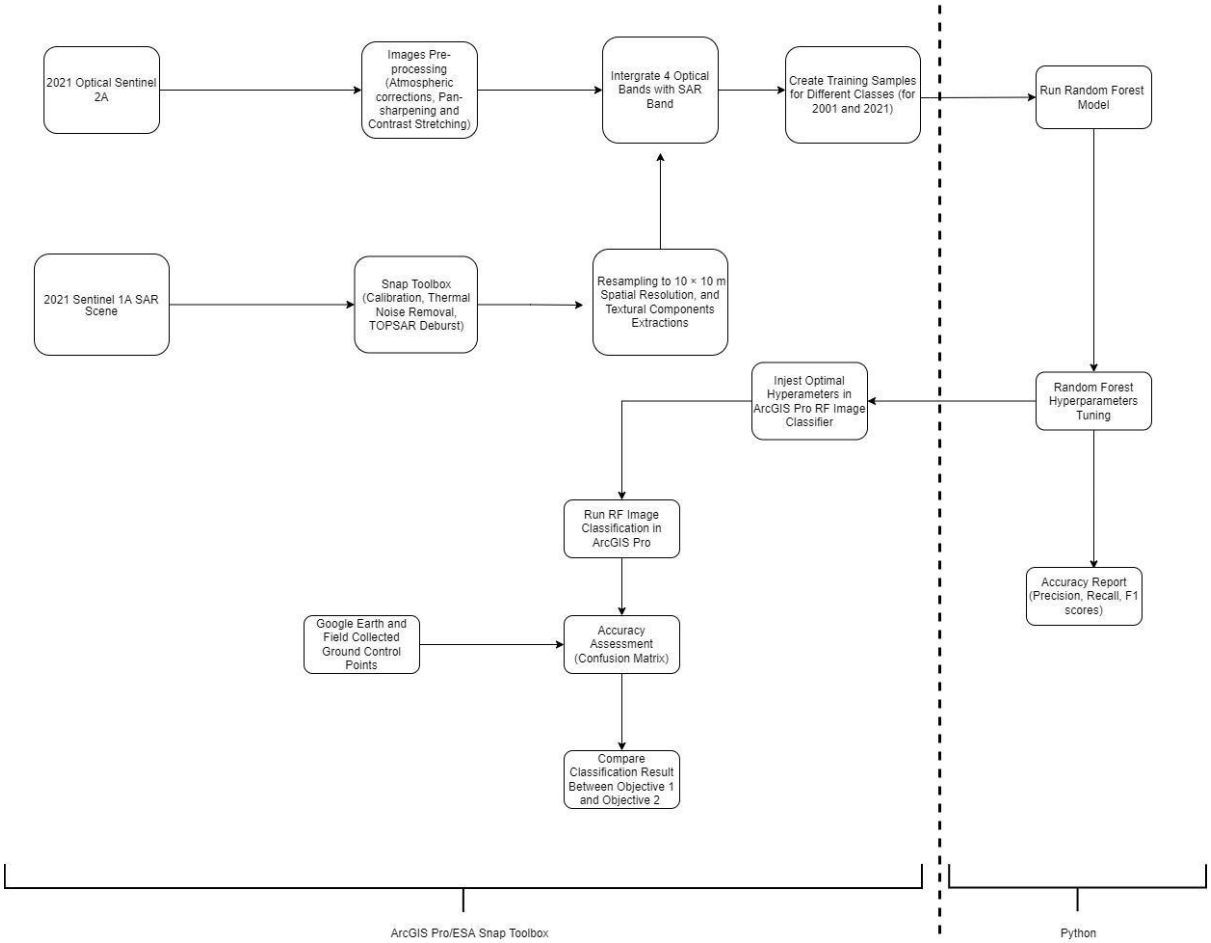


Figure 5: Second Objective Workflow.

SAR images offer unique insights into surface properties by capturing information about the scattering of radar waves, these are referred to as SAR textural components. SAR textural components provided valuable information of the spatial arrangement of pixel intensities within the SAR imagery data (ESA, 2022).

Speckle effects, characteristic of radar images that influence the radiometric information, were compensated using a Refined Lee low-pass filter with a 5 × 5 kernel Using the SNAP Sentinel Application Platform toolbox (Clerici *et al*.,2017). Atmospheric and terrain correction of the Sentinel 2A optical imagery was done in ArcGIS Pro. Bands that were deemed appropriate in Sentinel 2A were 2, 3, 4 and 8, with a spatial resolution of 10 × 10 m, these bands were chosens as they match spectral signatures in bands in Landsat 7 and 8 (See Figure 3) (Clerici *et al*.,2017). Two image appearance enhancement techniques (pan-sharpening and contrast

stretching) were performed on the Sentinel 2A image to match the enhancement process conducted in objective 1 for the Landsat Imagery.

3.3.2. Image Classification Algorithms.

Machine learning algorithms have increasingly been deployed for image classification in the remote sensing sector as they have favourable success for image classifications (Du *et al*., 2015; Belgiu and Drăguţ 2016). The algorithm that was used for image classification is the Random Forest (RF) classifier; this classifier is more favourable as it removes spectral confusion compared to more traditional classifiers like the Maximum Likelihood classifier (Sterk et al., 2007; Benediktsson and Ghamisi, 2015).

A complete random sample dataset was collected/digitized as a polygon of the 6 different landcover classes of interest on Google Earth, 70% of the data sets were used in training and optimizing the RF Classifier and 30 % was used for model validation; the breakdown was shown on Table 6. RF generates multiple random decisions on the original training dataset by using boot-strapping simulations, therefore capable of explaining complex non-linear relationships (Ma *et al*., 2004; Millard and Richardson, 2015; Belgiu and Drăguţ 2016; Veronesi and Schillaci, 2019).

RF requires two hyperparameters to be set when performing RF models, firstly the number of decision trees to be generated (*Ntree)* and secondly the number of variables that will be used to grow and split the trees (referred to as *Mtry*). These were determined using the Python programming language Scikit-learn Machine Learning library to optimize the RF model before conducting the classifier on the imagery using Esri ArcGIS Pro Software (Scikit-learn: Machine Learning in Python, 2011; Belgiu and Drăguţ 2016). After optimizing the hyperparameters for the RF model, the optimal *Ntree* value was 9; and the *Mtry* value was also 9.

3.3.3. Accuracy Assessments and McNemar Test

Accuracy assessments for the RF classifier were done by producing a confusion matrix and accuracy report (Kappa Scores, User Accuracy, Producer Accuracy, Overall Accuracy). From original dataset samples, 30 % of the data were used to evaluate model performance. The Sklearn library for machine learning in Python and ArcGIS Pro Accuracy assessment tools were used.

Accuracy assessments are critical for land cover classifications generated from any remote sensing data, as they help understand the success of the classification algorithm (Reis, 2008; Rwanga and Ndambuki, 2017; Adbi, 2020). This step was vital in assessing percentages of land cover changes for each land cover class between the two dates around the dam. Thereafter, the total areas of land cover change, gain or loss were derived using the image analyst on ArcGIS Pro for objective 1. They also inform the success of fusing Sentinel 1 SAR and Sentinel 2A optical data for monitoring land cover changes in the future in line with objective 2.

This research prioritised Kappa statistics scores, as they are widely used for assessing agreement between categorical variables. But Kappa scores have several limitations that can compromise their reliability and validity. Notably, Kappa's dependence on prevalence can lead to misleading interpretations, as its value can fluctuate with changes in category prevalence (Pontius and Millones, 2011). In unbalanced situations, Kappa can exhibit undesirable behaviour, making it unreliable for performance comparison. The quality of measurement, which Kappa does not account for, can also impact its reliability. Furthermore, the number of categories can influence Kappa's value, potentially leading to misleading interpretations (Pontius and Millones, 2011).

McNemar's test was done on the kappa score achieved between optical and also optical + SAR. This statistical test is used to compare the proportions of two paired binary outcomes The test is based on a 2x2 contingency table and assesses whether there is a significant difference between the two paired proportions (Here optical only Kappa score and also optical + SAR Kappa score) (Stuart, 1955; Sun and Yang, 2008). The test statistic follows a chi-squared distribution with one degree of freedom. It's important to ensure that the sample size is sufficient and that the

assumptions of the test are met. While McNemar's test is a valuable tool, it has limitations, such as low power for small sample sizes and its restriction to two related groups (Stuart, 1955; Sun and Yang, 2008).

3.3.4 LULC monitoring and change detection.

To quantify land cover change, the study converted raster land cover maps from 2001 and 2021 into separate polygons for each class. This made it possible to calculate the total area occupied by each land cover class at each time, providing a clear picture of how much land cover changed within each class in square kilometres between the 2001 LULC map and the 2021 LULC map.

Beyond simply measuring total change, the study also aimed to understand how different land cover classes transitioned between each other. To achieve this, a compute change raster detection tool in ArcGIS Pro was used. This tool analysed the 2001 and 2021 LULC maps and calculated the percentage of land area that interchanged between specific land cover classes over the 20-year study period.

Chapter 4: Results

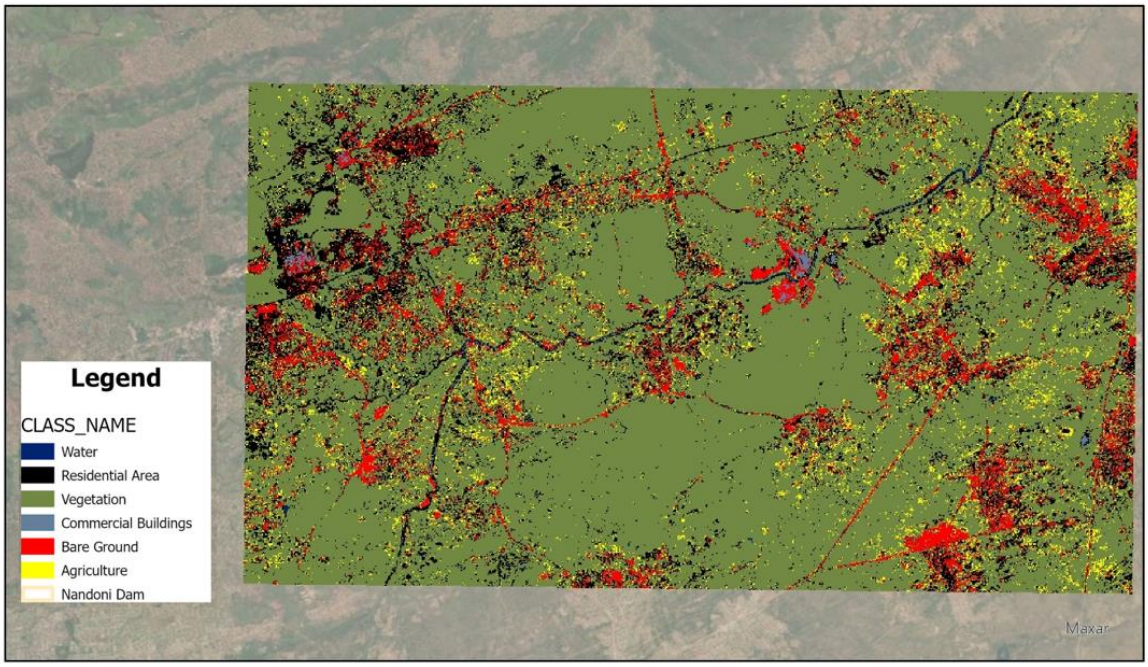
The following section collates and describes the results achieved for both objective 1 and 2. This chapter focused on describing the trends visible using classified maps, heatmaps, bar graphs, and confusion matrices to synthesize findings of the RF model's performance.

4.1 LULC mapping using Landsat imagery.

The six rural landcover types were pre-identified and classified using the RF Machine Learning classifier of 2 Landsat satellite images, dated 2001 for the before dam construction date and 2021 for the date after the date was constructed (See Figure 6). Both Landsat 7 (2001) and the Landsat 8 images of the site of Nandoni Dam and its surroundings under clear weather conditions, were used in this study. From the optimized RF model, the visual analysis displayed in the figure (See Figure

6) below displays a sea of the vegetation class with what looks like clusters of the other classes interspersed within it. This is a particular issue of remote sensing-based landcover classifications of rural areas where other classes that are of interest are hard to classify based on the vastness of natural vegetation compared to the land-conforming classes. This is a polarization compared to urban settings where land-transforming landcover classes tend to dominate against vegetation.

21 May 2001



2001 Classified Image

N

0 1 3 5 8 10 Km

Scale: 1:161,061

20 May 2021

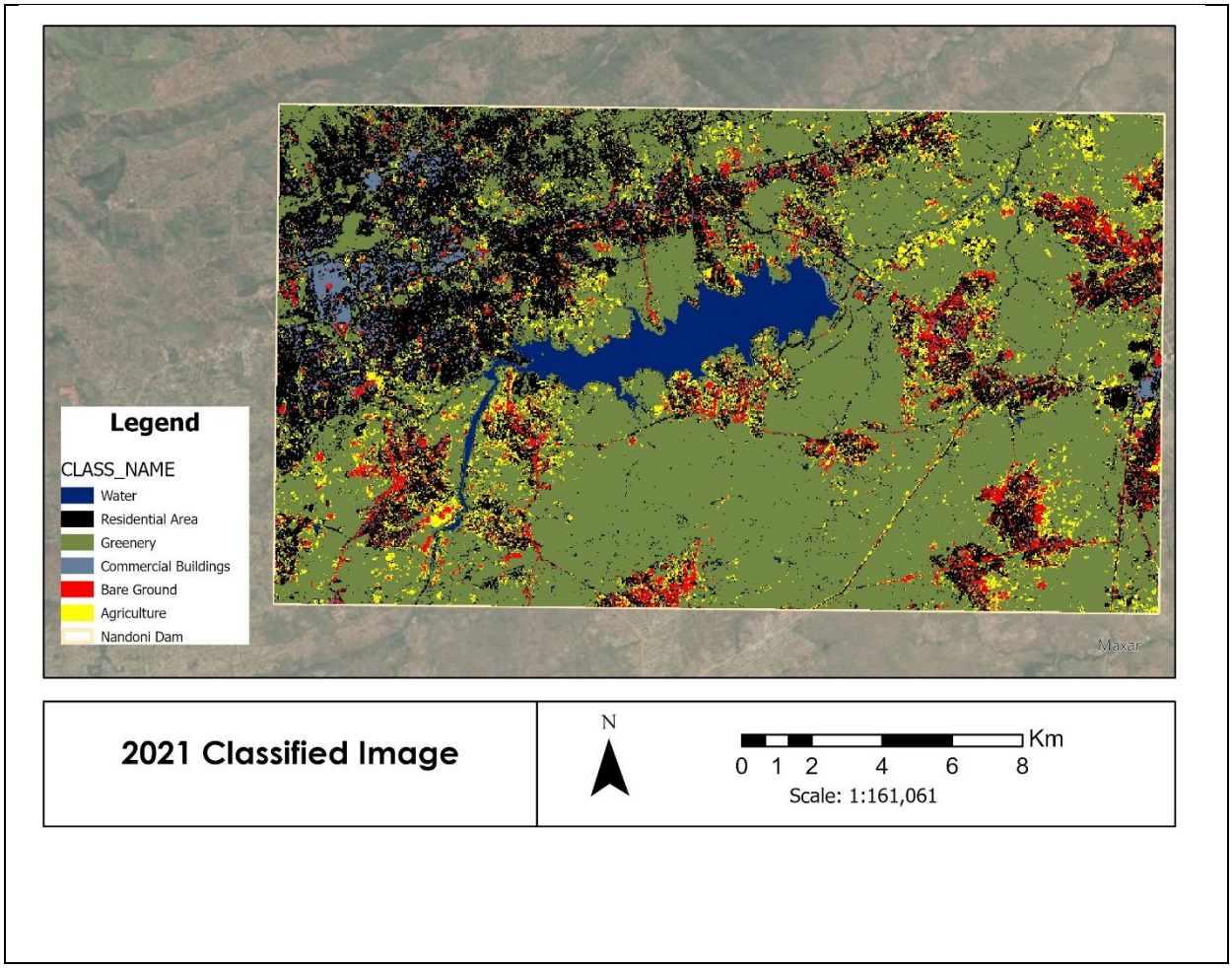


Figure 6: 2001 and 2021 classified images Using an Optimized Random Forest Classifier.

About 15 % of the sample dataset was originally used to train (70% used in model training; 15 for validation), and optimising the classifier was used to validate the success of the model. Confusion matrices (Tables 11 and 12) were used, showing the accuracy of a classification result by comparing the results with GCPs, A total of 500 points were used created by converting the polygons into randomized points in ArcGIS Pro for each class. The accuracy of the categorized images was achieved using randomly distributed ground truth points across the study region. The GCPs used for the accuracy assessment for the historical image dated 2001, were collected from historical google earth imagery. In Tables 11 and 12, a breakdown of the Overall accuracy (OA), which represents the total number of pixel that were correctly classified; The Producer's accuracy (PA), a measure of errors of omission,

was calculated by dividing the number of correctly classified points for a class by the total number of reference points for that class (column total); And the User's accuracy which is a measure of errors of commission, which was calculated by dividing the total number of correctly classified points in a class by the total number of points in that class (Congalton and green, 2019).

Table 11: 2001 Land Cover Classification Confusion Matrix.

2001 Landsat								
	Water	Residential	Vegetation	Commercial Buildings	Bare Ground	Agriculture	Total	U Accuracy
Water	121	0	0	0	0	0	121	1
Residential Areas	0	72	9	1	4	2	88	0.82
Vegetation	0	10	150	0	3	3	166	0.9
Commercial Buildings	0	3	0	37	1	0	41	0.9
Bare Ground	0	1	3	4	21	3	32	0.66
Agriculture	0	8	10	2	2	31	53	0.58
Total	121	94	172	44	31	38	500	
P Accuracy	1	0.77	0.88	0.84	0.68	0.79		
Kappa	0.82							

Table 12: 2021 Land Cover Classification Confusion Matrix.

2021 Landsat								
	Water	Residential	Vegetation	Commercial Buildings	Bare Ground	Agriculture	Total	U Accuracy
Water	131	0	0	0	0	0	131	1
Residential Areas	0	59	1	13	8	10	91	0.65
Vegetation	0	1	150	1	8	6	166	0.9
Commercial Buildings	0	1	0	24	1	0	26	0.92
Bare Ground	0	1	0	1	21	0	23	0.91
Agriculture	0	6	0	0	2	55	63	0.87
Total	131	68	151	39	40	71	500	
P Accuracy	1	0.87	0.99	0.62	0.5	0.77		
Kappa	0.85							

From observing the PA and UA, it can be seen the water and vegetation class seem to be the easiest to classify using the RF model for both dates, this can be attributed to the uniqueness of spectral reflectance of water compared to the other classes and the vastness of natural vegetation within the study area for both dates respectively (See Figure 6). On the other hand, from both matrices (Table 11 and 12), it seems that the hardest classes to classify for the algorithm were the bare ground, agriculture, commercial and residential classes. This is due to the interspersed distribution nature of these classes within the vastness of the vegetation class. In a more specific case, from both tables most of what was agriculture, and the bare ground were mostly confused as vegetation, which is likely due to the gradual transition of these class borders and connections with the dense vegetation class, therefore the RF classifier had a harder time deciding which was the correct class.

A final measure of model performance was the Cohen Kappa coefficient score, which measures the reliability of the classification outputs. The Kappa coefficient ranges from 0 to 1, where a value of 0 suggests similarity to random classification.

A lower value closer to 0 indicates that the classification is significantly poorer than random, while a value approaching 1 signifies accurate classification. The model achieved Kappa scores of 0.82 and 0.85 for the 2001 and 2021 dates respectively, the slight improvements between the kappa of 2001 and 2021 can be attributed to the use of different sensors although both belonged to the same Landsat mission, band matching was done prior (see the methodology section for the detailed process), wherein Landsat 8 used in 2021 boast more modern sensors compared to Landsat 7 used in 2001 yet both sensors deliver the same resolution. Although Landsat 7 is still in orbit and delivering valuable data sets, over the years, the sensors have developed a system glitch that has resulted in some areas globally without adequate coverage (see figure 12 in the appendix for an example of this over the study area).

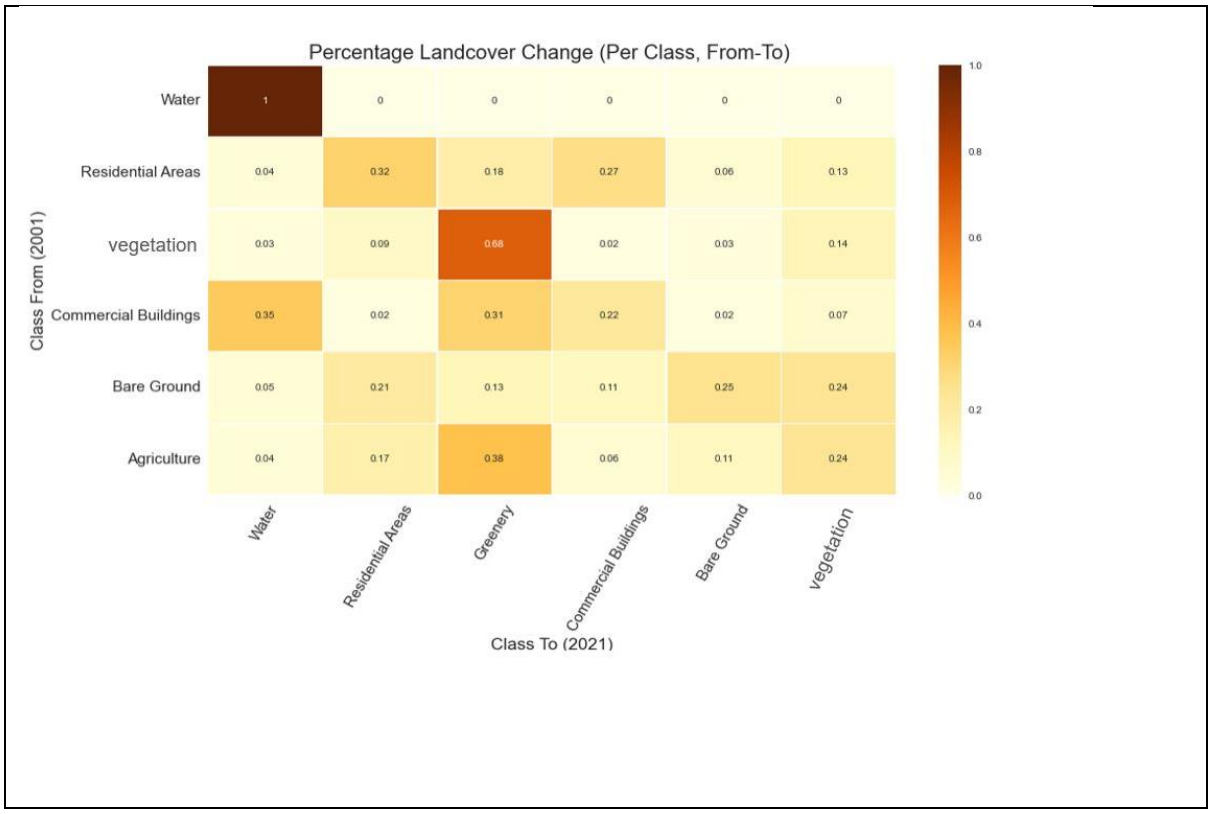


Figure 7: Percentage Landcover Change (Per Class, From- To).

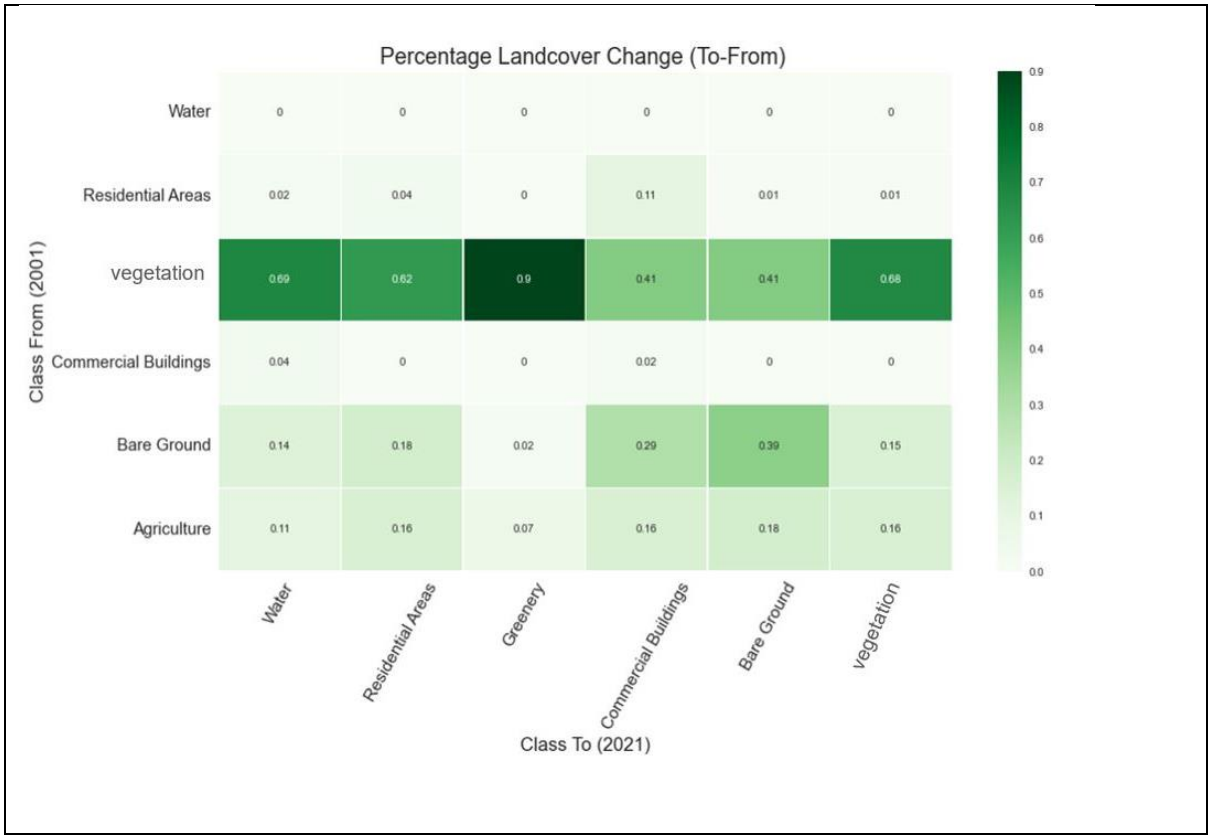


Figure 8: Percentage Landcover Change (Per Class, To- From).

The heatmaps above (see figure 7 and 8) provide a deeper understanding of how the landcover classes are interchanging and feed off each other between 2001 and 2021. This goes beyond simply understanding the performance of the classifier but highlights the changes happening between the classes over 20 years (from 2001 to 2021). The above heat map assigns values between 0 and 1, where values closer to 0 represent lower percentage values and values closer to 1 represent higher percentage values. Figure 7 groups the classes in 2001 as rows against what the class changed to in 2022 as columns; here, we see that 100% of what the model picked up as water in 2001 remained as water in 2022. Furthermore, 21% of what was picked up as part of the bare ground class in 2001 became the residential class in 2022.

Figure 8 assumes a similar interpretation to Figure 7, but this heatmap provides an understanding of where the classes in 2021 come from in relation to the extent of the classes in 2001. This heatmap further makes it clear that 5 of the classes (Water,

Residential Areas, Commercial Buildings, Bare ground and Agriculture) in 2021 have increased as a result of feeding from the vegetation class) this might be because the vegetation class dominants in the study area, which results in the other classes being interspersed as islands within this class; by further assessing the trends, 69 % of what is water in 2021 as a result of the dam construction came from what was the vegetation class.

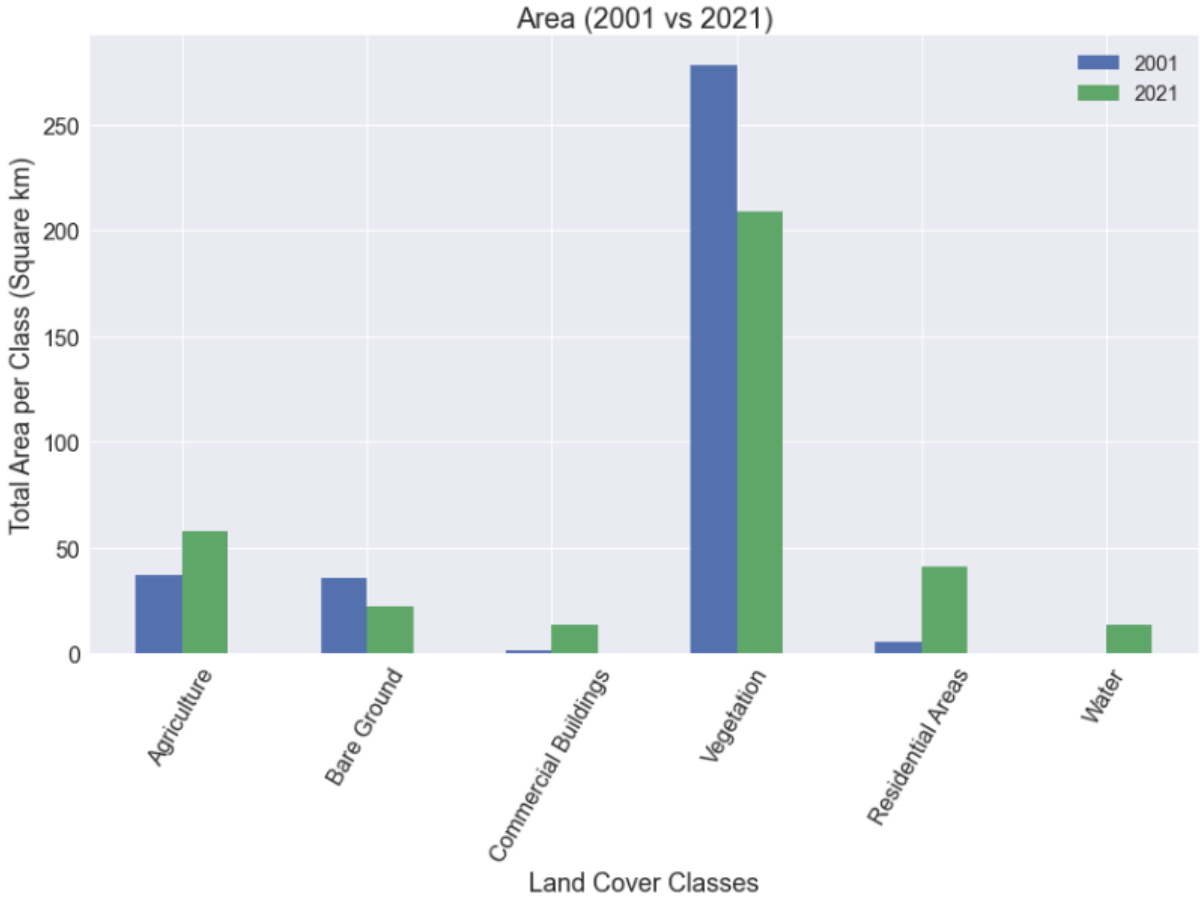


Figure 9: Landcover Change Class Change (Per Square Km).

Figure 9 provides a square km area change of the different individual classes between 2001 and 2021. The vegetation class had an area change of 270 km² to 210 km² between 2001 and 2021, which was the highest decline in all the classes, this can be attributed to how it being the largest classes, the declines will be more severe than in other large extent classes, which then opens up the conversation of biodiversity loss, and its implication to the surrounding rural areas that depend of their natural vegetation, the only other class that saw a decline was the bare ground

class as other classes went up in area coverage between 2001 and 2021. Water, for example, went up to about 30 km² in 2021 because the Nandoni Dam was constructed.

4.2 LULC mapping using Sentinel 1(SAR) and 2 (Optical)

After pre-processed the SAR image as outlined in the methodology, we were interested in extracting SAR textures, which depicts the statistical traits in SAR imagery and can be applied as an aspect to improve the detection of land cover classes in interspersed land cover classes in a rural setting such as Nandoni Dam (Hajeb *et al*., 2020; Bai *et al*., 2017; Gong *et al*., 2016). The textures, namely, Mean, Variance, Contrast, Homogeneity, Dissimilarity, Energy, Angular Second Moment (ASM), and Gray-Level Co-occurrence Matrix (GLCM), (Mean, Variance, Correlation) were used in this study (See Figure 10 below).

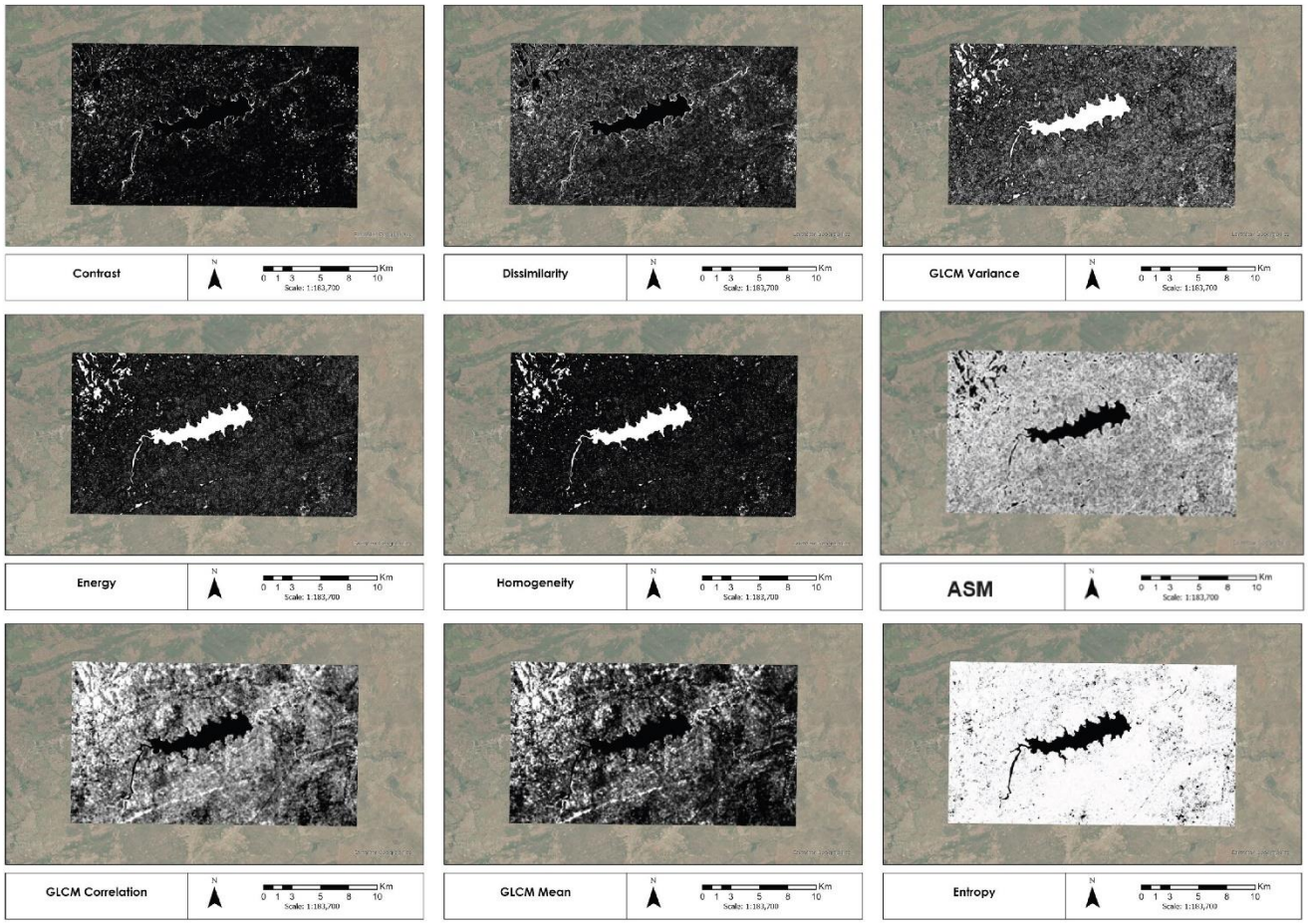


Figure 10: SAR Extracted Statistical Textures.

Figure 11 shows a composite between the different SAR textures outputs and sentinel 2A optical image bands at 10*10 m pixel size. Data fusion between SAR and optical imagery has proven to be of interest in improving land cover classifications (Yuan *et al*., 2020; Abdikan 2018; Forget *et al*., 2018; Uhlmann and Kiranyaz 2014; Tan *et al*., 2018; Huang *et al*., 2008; Michelson, 2000), which was of particular interest to explore this in a rural setting such as Nandoni Dam.

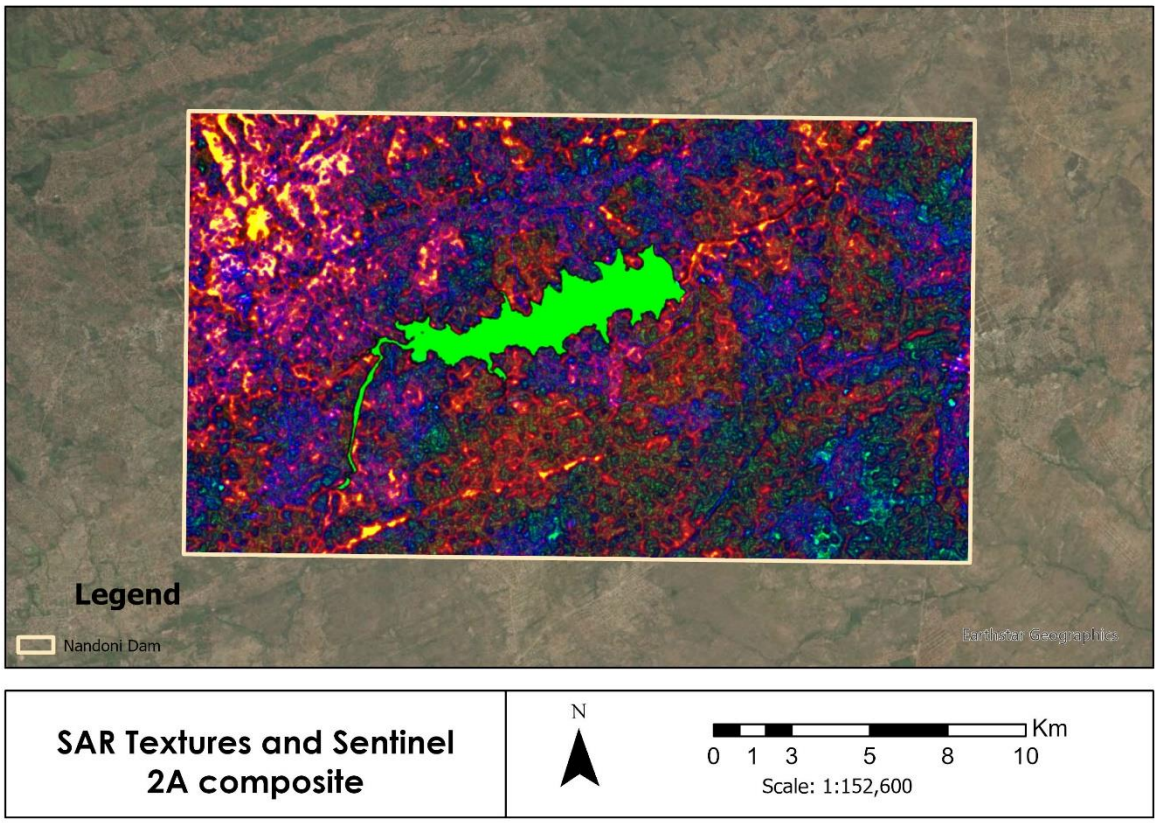


Figure 11: SAR Textures and Sentinel 2; composite displayed through band 3 for Sentinel 2; and the entropy and energy band for SAR.

The 6 rural landcover types were pre-identified and classified through the use of the RF Machine Learning classifier on a composite image of sentinel 1 SAR texture bands and Sentinel 2A Optical bands, at 10*10 m of the site of Nandoni Dam and its surroundings, which was used in this study. From the Optimized RF model, the visual analysis displayed in the Figure 12 below, displays a sea of the vegetation class with what looks like clusters of the other classes interspersed within it. This is a particular issue of remote sensing-based landcover classifications of rural areas appears to be recurring from the results in just using optical Imagery for landcover classification using Landsat imagery for 2001 and Landsat 8 imagery for 2021, whereas other classes that are of interest are hard to classify based on the vastness of natural vegetation compared to the land conforming classes.

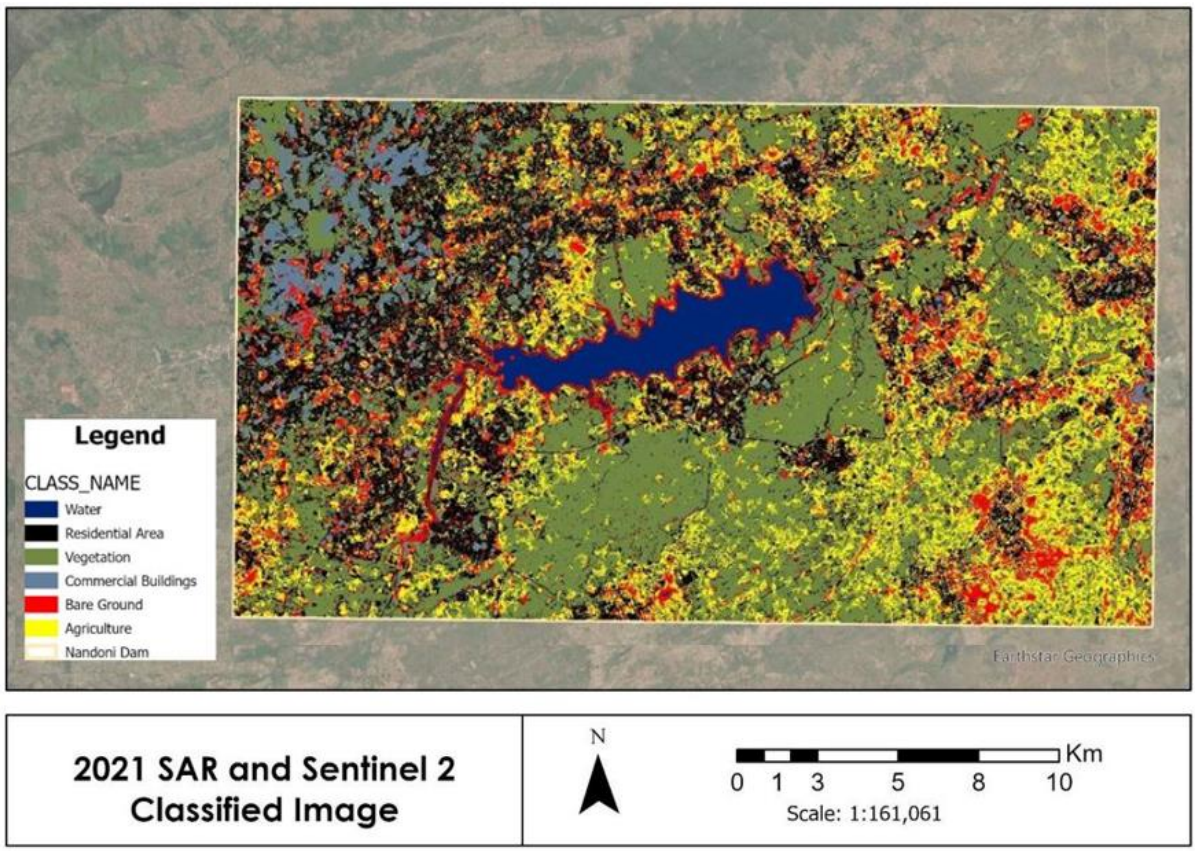


Figure 12: SAR, and Sentinel 2 Classified Image.

From the original sample data, 15 % was originally used to train (70% was used in model training, 15 used for validation), and optimized the classifier was used to validate the success of the model. A confusion matrix was used to validate the accuracy of the RF classification by comparing the classification results with ground-validated points. The accuracy of the categorized image was achieved using randomly distributed ground truth points across the study region. In Table 13, a breakdown of the Producer's accuracy (PA), a measure of errors of omission, was calculated by dividing the number of correctly classified points for a class by the total number of reference points for that class (column total). User's accuracy, a measure of errors of commission, which was calculated by dividing the total number of correctly classified points in a class by the total number of points in that class.

Table 13: SAR, and Sentinel 2 Fusion LULC Confusion Matrix.

SAR + Sentinel								
	Water	Residential	Vegetation	Commercial Buildings	Bare Ground	Agriculture	Total	U Accuracy
Water	174	0	0	0	0	0	174	1
Residential Areas	0	47	0	3	2	0	52	0.9
Vegetation	0	0	124	1	6	19	150	0.83
Commercial Buildings	0	0	0	16	0	0	16	1
Bare Ground	0	2	0	0	24	0	26	0.92
Agriculture	0	7	0	0	0	75	82	0.91
Total	174	56	124	20	32	94	500	
P Accuracy	1	0.8	1	0.8	0.75	0.8		
Kappa	0.89							

From observing the PA and UA, it can be seen that the water and vegetation class seem to be the easiest to classify using the RF model. This can be attributed to the uniqueness of reflectance on water compared to the other classes and the vastness of natural vegetation within the study area (See Figure 12). On the other hand, we see a common trend from just using optical imagery, where, that the hardest classes to classify for the algorithm were the bare ground, agriculture, commercial and residential classes.

Although these classes were not as easy to classify, the results show that there was an improvement in fusing SAR and Optical imagery in their classification compared to just using optical imagery. A study by Forget, (2018), synthesized that data fusion (here sentinel 1 SAR and Sentinel 2) provides improvements in accuracy success in landcover classification in sub-Saharan Africa were misclassifications are often observed due likely to the interspersed distribution nature of these classes within the vastness of the vegetation class.

A final measure of model performance in the land cover classification of the SAR and Optical fusion dataset was the Cohen Kappa coefficient score, which measures the

reliability of the classification outputs. The model achieved a score of 0.89, which shows improved classification compared to the results attained using Landsat imagery only.

4.3 McNemar test on Statistical Significance of Kappa Score Achieved Using Optical vs Optical + SAR.

As a measure of statistical significance between the Kappa Statistics achieved between using only optical imagery; and optical + SAR, A McNemar's test was ran. The test aims to examine the change in binary repeated measurements, by understanding the significant difference between two paired groups, on our case the kappa scores achieved between using only optical imagery; and optical + SAR (Stuart, 1955; Sun and Yang, 2008; Lachenbruch, 2014).

Before we calculate the McNemar's score, we need to determine our null hypothesis. Here we assume a null hypothesis (H0) that, "There is a statistical significance in the Kappa score achieved when intergrating optical + SAR as compared to the average Kappa score achieved from only using optical".

We can then calculate the score based on the Kappa score achieve between using only optical; and using optical + SAR.

Optical Kappa Score (p1): 0.82 + 0.85 = 1.67
= 1.67/2
= 0.835

Optical + SAR (p2) = 0.89

Next, we'll calculate the McNemar test statistic (χ^2) using the following formula:

$\chi^2 = (p2 - p1)^2 / (p1(1-p1) + p2(1-p2))$ ……(eq 1)

Plugging in the values, we get:

$\chi^2 = (0.89 - 0.835)^2 / (0.835(1-0.835) + 0.89(1- 0.835))$
$= 0.0036 / (0.1401 + 0.0979)$
$= 0.0036 / 0.2380$
$= 0.0151$

Now, we need to determine the degrees of freedom (df). Since we're comparing two proportions, df = 1.

Using a chi-squared distribution table (Figure 13, in the adendum), we find the critical value for χ^2 with df = 1 and $\alpha = 0.05$ (standard significance level):

$\chi^2crit = 3.841$

Since our calculated χ^2 (0.0151) is less than χ^2crit (3.841), we fail to reject the null hypothesis (H0).

p-value = $P(\chi^2 \geq 0.0151) \approx 0.9023$ (using a chi-squared distribution table)

The p-value is greater than 0.05, indicating that the difference between kappa score of 0.835 (optical imagery) and 0.89 (Optical + SAR) is not statistically significant at the 5% signficant level.

Chapter 5: Discussion

The following section main focus of this chapter was to explain and synthesis the significance of the findings in this study and relate them with existing knowledge. Furthermore, by exploring the implications of the study in influencing policy and in the broader context of LULC monitoring and managments.

This study explored the use of optical imagery to measure land use/landcover change before and after the construction of Nandoni Dam and leveraged the use of data fusion in improving land use/ landcover classification through integrating SAR and optical imagery. Due to user-friendly programming tools, access to high-end consumer computing power, and the availability of free satellite data, the data science and remote sensing communities have begun to be more integrated and aligned in recent years (Abdi, 2020). More particularly the Random Forest algorithm classifier has proven to be best suitable for land cover classification using optical satellite imagery (Phan *et al*., 2020; Wu *et al*., 2019; Noi and Kappas, 2018).

However, two important RF classification parameters, namely the number of trees and the number of variables tried at each split, have an impact on classification accuracy. As a result, optimal parameter selection is an unavoidable issue in RF-based image classification (Ming *et al*., 2016). This study therefore extensively prioritized optimizing these two parameters for RF land cover classification models over Nandoni Dam.

The objectives for this study were two fold, firstly to develop and compare the success of the Random Forest algorithm for landcover classification over a 20-year period on Landsat 8 remotely sensed data for the 2021 (after dam construction) date and the Landsat 7 for the 2001 date (before dam construction) over Nandoni Dam. And the second objective was to explore the data fusion between SAR and Optical Imagery to try improving land cover classification in fragmented rural areas.

The Landsat 8 (2021 date) trained model produced a marginally better Kappa score compared to the Landsat 7 (2001 date) trained model (0.85 and 0.82) respectively; however, this was not unexpected as Landsat 7 sensors precede Landsat 8 which captures a narrower bandwidth, therefore the quality of sensors used might have an impact on the quality of the remotely sensed data produced regardless of the high degree of similarities between both sensors (Holden and Woodcock, 2016; He *et al*., 2015; Li *et al*., 2014). Nonetheless, when comparing both models, they were able to produce similar results in the ability to classify the identified landcover classes of

interest, this is represented by the Confusion Matrices Accuracy assessments done for each date (See Table 11 and 12), There is little information in the literature to compare the value of conducting landcover change over a period using separate satellites, but due to lack of quality remotely sensed data covering a 20-year period from each satellite, prompted the study to use two separate satellite imagery.

Landsat 7 and 8 have bands that capture similar spectral ranges, this requires meticulous matching that was maintained when ensuring subjectivities and biases are removed when developing the RF landcover classification models for the two satellites (See methodology section). Some environmental factors such as moisture and vegetation cover may have changed in study region over the 20-year period, having an influence on the quality of models produced for the separate dates.

Rural communities around Nandoni Dam are still heavily dependent on firewood, medicinal plants and grazing pastures. Therefore, this study suggests an urgency in conducting in-depth studies in the structure and overall health of the natural vegetation surrounding Nandoni Dam.This was not the central focus of this study, but the vegetation class is the most changed out of the 6 classes. Therefore, this observation forms basis of applying other RS techniques such as Canopy Height Measurements to assess continuous biomass changes within the region.

Furthermore, the results of LULC monitoring between 2001 and 2021 influences decisions by local municipalities regarding the value of rapid development in the region. The analysis considers the potential benefits of rapid development in commercial, agricultural, and housing sectors against the risks to the natural environment. The concern here is that rapid development might put a strain on the natural vegetation around the dam, which provides essential ecosystem services to local communities (see heatmaps in Figures 7 and 8 for reference). Striking a balance between natural vegetation preservation and advances in commercial development is key in securing the overall livelihoods of those in the region.

The results of the RF model were satisfactory in the classification of land use and land cover, with a mean of 0.85 Cohen kappa score between only using optical and data fusion (here SAR and Optical imagery) (See Table 11, 12, and 13) This is in line with recent studies, that outline the improved success rates and reliability of using the RF machine learning-based model in land cover classification compared to traditional classification techniques (Dong *et al*., 2022; Buhvald *et al*., 2022; Alzahrani *et al*., 2022; Lingwal *et al*., 2022; Marlina, 2022; Shen *et al*., 2022).

The success and quality of the outputs of the model were of utmost importance in the rural setting of Nandoni Dam where land use types such as commercial buildings, residential areas, agricultural land uses, and water are interspersed in a sea of natural vegetation.This rural setting is not the focus of most literature as most studies are focused on urban areas.

The results achieved in this study provide us with insight that the natural vegetation around Nandoni Dam has been on a decline for the past 20 years (between 2001 and 2021). As a result of the increase in area coverage of the other classes (mainly Commercial and Residential).This can be attributed to the presence of water due to the presence of the dam in recent years fuelling the rapid increase in these classes that require adequate water to thrive. But it is important to note that these findings raise potential threats to the biodiversity and ecosystem services that the natural vegetation provides to the people in the region. Thus, this research is important in using data and recent technological advancements in methodologies in remote sensing to influence policy-making that prioritizes appropriate land use and future land use in rural areas.

As remote sensing data becomes more widely available, data fusion techniques have become increasingly popular in land use/land cover monitoring (Yuan *et al.*, 2020; Abdikan, 2018; Forget *et al*., 2018; Uhlmann and Kiranyaz, 2014). Therefore, despite the visible trends and success in classifying land cover change between 2001 and 2021 using only optical Landsat imagery, it was important to explore if

integrating SAR and optical imagery would improve rural land cover classification to better understand interspersed landcover classes in rural settings.

The outcomes of this study alluded that fusion of SAR and optical imagery can provide a comprehensive better understanding of the land cover in rural areas, as the model Cohen Kappa score was at 0.89, a significant improvement from only using optical imagery. Although when running the Mcnemar test, we did not find any statistical significance between the average optical kappa score and that of optical + SAR; However, we argue that the outcome here does not make the findings of this study less important. “The Absence of evidence is not evidence of absence”, just because we did not find a statistically significant difference or effect, it doesn't mean it doesn't exist (Altman and Bland, 1995; Alderson, 2004). The statistical insignificance here might be due to limitations in the study design and other factors such as sample size used in conducting the Accuracy Assessment, expertise or the chosen classifer itself. Using optical + SAR to conduct LULC classification in fragmented rural landscapse is still important to investigate.The overall improvements should be celebrated and higlighted on the Kappa scores achieved, and offers a basis on data fusion for improved LULC classfication and monitoring in fragmaneted rural landscapes. This is in line with studies from Yuan et al., 2020 who found significant improvements in landcover classification accuracies when integrating SAR and Multispectral optical imagery.

The success can be attributed to SAR data being able to provide information about the surface even under adverse weather conditions, while optical imagery can provide detailed information about the colour of the surface in the study area. By combining the two data sources, the limitations of each data source were overcomed to an extent, resulting in a nearly accurate and complete classification of land cover around Nandoni Dam. Moreover, optical data provided spectral information while SAR data offers information about the statistical textures, roughness, and the backscattered energy, which was useful in identifying different land cover features and overcome the traditional limitations of just relying on optical imagery.

This approach coincided with how trends suggest that land use and land cover change are among the major drivers of biodiversity loss and ecosystem degradation worldwide (Bimenyimana *et al*., 2022; Chen *et al*., 2022). Thus, effective monitoring and management of land use change are critical for maintaining biodiversity and ecosystem services, as well as for achieving sustainable development goals (Schneider *et al*., 2022). Remote sensing technologies provide powerful tools for monitoring land use and land cover change at regional and global scales, enabling us to detect and analyze changes in land use over time, and to assess the impacts of these changes on biodiversity and ecosystem services for rural settings like the Nandoni Dam study area.

More so, remote sensing based land use and land cover studies have played a significant part in shaping conservation and sustainable development efforts, as well as aiding regulations at the national and international levels in recent years. The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) report, for example, emphasizes the necessity of tracking biodiversity changes to inform conservation efforts and policy decisions. Similarly, the Living Planet Report stresses the pressing need for conservation action in the face of dramatic declines in global wildlife populations (which in turn have declined by 68% since 1970), and the CBD global biodiversity framework calls for the integration of biodiversity considerations into the development of policies and planning (Schneider *et al*., 2022).

Chapter 6. Conclusion and Recommendation

Chapter six conclude the study by providing context as to how the research was able to achieve its main objectives, charpter five also highlights the contribution this study has in the scientific and also give community and recommendations for future studies that can be conducted to improve our understanding of LULC in fragmented rural landscapes.

6.1. Conclusions

Remote sensing-based land use and land cover studies have emerged as a feasible alternative in assessing and managing the impacts of land use change on biodiversity and ecosystem services, particularly in rural settings such as Nandoni Dam. This calls for continuous improvements in the LULC monitoring methods that are focused on fragmented rural landscape. This study therefore not only illustrate the land cover change happening over a span of 20 year around Nandoni Dam, in line with the first objective of montioring land cover before and after the dam was constructed, but further portrays a case of potential in success where data fusion between SAR and optical data leads to increased success of the land cover classification modelling, as an alternative to only relying on optical imagery for LULC classification for the second objective.

6.2 Recommendations

The study successfully achieved its two main objectives. Which were:

1. To quantify and monitor land cover class change over a 20-year period from 2001 through the application of an optimized Random Forest algorithm on Landsat optical imagery.
2. To assess the potential of integrating sentinel 2 optical imagery and Sentinel 1 SAR remote sensing data in improving LULC in rural interspersed landscapes.

Building on the study's findings, we recommend that urgent land use framework and policies should be drafted and enforced for the rural areas in proximity to the Nandoni Dam, in acccordance to the development principles; and norms and standards of Act No. 16 of 2013: Spatial Planning and Land Use Management Act . Furthermore, we acknowledge that there are still areas to research that tie in closely to this study. We therefore, recommend the following possible areas for future studies in the region:

1. Exploring the use of other machine learning algorithms and comparing the success of those models in classifying land cover in interspersed rural settings.
2. In-depth understanding of new fusion techniques that might be better at integrating SAR and optical imagery to improve LULC monitoring in interspersed rural settings.
3. Conduct research on biomass changes over time, as the vegetation class appears to be the most changing.
4. Conduct qualitative studies in the rural communities to understand, how LULC changes has affected their lives over time.

References

Abbas, A., Minallh, N., Ahmad, N., Abid, S., Khan, M. (2016). K-Means and ISODATA Clustering Algorithms for Landcover Classification Using Remote Sensing. Sindh University Research Journal, 48.

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.

Abdikan, S. (2018). Exploring image fusion of ALOS/PALSAR data and LANDSAT data to differentiate forest area. *Geocarto International*.

Al-Doski, J., Mansorl, S. B., Shafri, H. Z. M. (2013). Image classification in remote sensing. Department of Civil Engineering, Faculty of Engineering, University Putra, Malaysia, 3(10).

Ally, S., Lissoni, A. (2012). ‘Let's Talk About Bantustans. *South African Historical Journal*, 64, 1-4.

Altinbilek, D. (2002). The role of Dams in development. *Water Sci Technol*,45,169–180.

Altman, D. G., Bland, J. M. (2009). Parametric v non-parametric methods for data analysis. *Bmj*, 338.

Alzahrani, A., Kanan, A. (2022). Machine Learning Approaches for Developing Land Cover Mapping. *Applied Bionics and Biomechanics*.

Al-Ruzouq, R., Shanableh, A., Gibril, M., Kalantar, B. (2019). Multi-scale correlation-based feature selection and random forest classification for LULC mapping from the integration of SAR and optical Sentinel images. , 11157, 111570B - 111570B-13. https://doi.org/10.1117/12.2533123.

Amani, M., Salehi, B., Mahdavi, S., Brisco, B. (2018). Spectral analysis of wetlands using multi-source optical satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 144, 119-136.

Ayeni, A. O., M. A. Cho, A. Ramoelo, R. Mathieu, A. S. O. Soneye, and J. O. Adegoke. "Could local perceptions of water stress be explained by LULCC?" *Geoinfor Geostat:* An Overview 2013, S1.

Bai Y, Adriano B, Mas E, Koshimura S. (2017) Building Damage assessment in 2015 Gorkha, Nepa, earthquake using only post-event dual polarization synthetic aperture radar imagery. *Earthq Spectral*. 33 (s1): 185-195.

Barrett, E. C., Curtis, L. F. (1999). Introduction to environmental remote sensing. Psychology Press.

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

Benediktsson, J. A., Ghamisi, P. (2015). Spectral-spatial classification of hyperspectral remote sensing images. Artech House.

Bimenyimana, T., Bugenimana, E., Habineza, E., Bushesha, M., Ali, M. (2022). Impact of Urbanization on Land use and Land Cover Changes in Growing Cities of Rwanda. *Journal of Korean Society of Environmental Engineers*.

Binbusayyis, A., Vaiyapuri, T. (2020). Comprehensive analysis and recommendation of feature evaluation measures for intrusion detection. *Heliyon*, 6(7), e04262.

Bird, J., Wallace, P. (2001). Dams and Development – An Insight to the Report of the World Commission on Dams. *Irrigation and drain*, 50, 53-64.

Biswas, A.K., Tortajada, C. (2001). Development and Large Dams: A Global Perspective. *International Journal of Water Resources Development*, 17, 9-21.

Bouaziz, M., Eisold, S., Guermazi, E. (2017). Semiautomatic approach for land cover classification: a remote sensing study for arid climate in southeastern Tunisia. *Euro-Mediterranean Journal for Environmental Integration*, 2(1), 1-7.

Breiman, L. 2001. Random Forests. *Machine Learning*, 45, 5–32.

Buhvald, A., Racic, M., Immitzer, M., Oštir, K., Veljanovski, T. (2022). Grassland Use Intensity Classification Using Intra-Annual Sentinel-1 and -2 Time Series and Environmental Variables. *Remote Sensing.*

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

Caranza, J., Calderon, M., Ancog, R., Carandang, M.,Rebancos, C. (2023). Fragmentation Analysis of Capisaan Surface Karst Landscape through Changes in Land Use and Land Cover using FRAGSTATS. *Mindanao Journal of Science and Technology*. https://doi.org/10.61310/mndjstecbe.1049.23.

Chachondhia, P., Shakya, A., & Kumar, G. (2021). Performance evaluation of machine learning algorithms using optical and microwave data for LULC classification. *Remote Sensing Applications: Society and Environment*.

Chang, Y., Hou, K., Li, X., Zhang, Y., Chen, P. (2018). Review of land use and land cover change research progress. *In IOP Conference Series: Earth and Environmental Scienc*e (Vol. 113, No. 1, p. 012087). IOP Publishing.

Chen, B., Chen, B., Huang, B., Xu, B., Xu, B., & Xu, B. (2017). Multi-source remotely sensed data fusion for improving land cover classification. *Isprs Journal of Photogrammetry and Remote Sensing*, 124, 27-39. https://doi.org/10.1016/J.ISPRSJPRS.2016.12.008.

Chen, H., Chen, Y., Chen, X., Zhang, X., Wu, H., Li, Z. (2022). Impacts of Historical Land Use Changes on Ecosystem Services in Guangdong Province, China. *Land*.

Cheng, G., Wei, J. (2019). Color Quantization Application Based on K-Means in Remote Sensing Image Processing. *Journal of Physics: Conference Series*, 1213. https://doi.org/10.1088/1742-6596/1213/4/042012.

Cheng, K., Ling, J., Lin, T., Liu, Y., Shen, Y., Kono, Y. (2019). A NEW THINKING OF LULC CLASSIFICATION ACCURACY ASSESSMENT. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* https://doi.org/10.5194/ISPRS-ARCHIVES-XLII-2-W13-1207-2019.

Chuvieco, E. (2019). Fundamentals of satellite remote sensing. CRC press.

Clerici, N., Valbuena Calderón, C. A., Posada, J. M. (2017). Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. *Journal of Maps*, 13(2), 718-726.

Congalton, R. G., Green, K. (2019). Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

Cortes, C., Vapnik, V. (1995). Support vector machine. *Machine Learning*, 20(3), 273-297.

Cortijo, F. J., De La Blanca, N. P. (1997). A comparative study of some non-parametric spectral classifiers. Applications to problems with high-overlapping training sets. *International Journal of Remote Sensing,* 18(6), 1259-1275.

Cortijo, F. J., Perez de la Blanca, N., Molina, R., & Abad, J. (1995, April). On the Combination of Nonparametric Nearest Neighbor Classification and Contextual

Correction. In Pattern Recognition and Image Analysis*. Proceedings of the VI Spanish Symposium on Pattern Recognition and Image Analysis, C ordoba, Spain* (pp. 503-510).

De'Ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, 88(1), 243-251.

Dong, X., Cao, W., Cheng, H., Yin, X. (2022). Random Forest Reliability Evaluation Based on Combination Feature Selection. 2022 *IEEE 10th Joint International Information Technology and Artificial Intelligence Conference* (ITAIC).

Du, P., Samat, A., Waske, B., Liu, S., Li, Z. (2015). Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 38-53.

Du, X., Zheng, X., Lu, X., Doudkin, A. (2021). Multisource Remote Sensing Data Classification with Graph Fusion Network. *IEEE Transactions on Geoscience and Remote Sensing*, 59, 10062-10072. https://doi.org/10.1109/TGRS.2020.3047130.

Égré, D., Senécal, P. (2003). Social impact assessments of large Dams throughout the world: lessons learned over two decades*. Impact Assessment and Project Appraisal*, 21, 215-224.

Eilertsen, G., Jönsson, D., Ropinski, T., Unger, J., Ynnerman, A. (2020). Classifying the classifier: dissecting the weight space of neural networks. arXiv preprint arXiv:2002.05688.

Enderle, D. I., Weih Jr, R. C. (2005). Integrating supervised and unsupervised classification methods to develop a more accurate land cover classification. *Journal of the Arkansas Academy of Science*, 59(1), 65-73.

Fairbanks, D, Thompson, MW, Vink, DE, Newby, TS, Van den Berg, HM, Everard, D. (2000). The South African land-cover characteristics database: a synopsis of the landscape. *South African Journal of Science*, 96(2), 69-82.

Forget, Y, Shimoni, M., Gilbert, M., Linard, C. (2018). Complementarity Between Sentinel-1 and Landsat 8 Imagery for Built-Up Mapping in Sub-Saharan Africa (preprints).

Friedman, J. H. (2020). Contrast trees and distribution boosting. Proceedings of the *National Academy of Sciences*, 117(35), 21175-21184.

Friedman, J.H. (2000). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*. 29. 10.1214/aos/1013203451.

Ghayour, L., Neshat, A., Paryani, S., Shahabi, H., Shirzadi, A., Chen, W., Al-Ansari, N., Geertsema, M., Amiri, M., Gholamnia, M., Dou, J., Ahmad, A. (2021). Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/Use Classification Using a Comparison between Machine Learning Algorithms. *Remote. Sens*., 13, 1349.

Gulinck, H., Wagendorp, T. (2002). References for fragmentation analysis of the rural matrix in cultural landscapes. *Landscape and Urban Planning*, 58, 137-146. https://doi.org/10.1016/S0169-2046(01)00216-X.

Haas, J., Ban, Y. (2017). Sentinel-1A SAR and sentinel-2A MSI data fusion for urban ecosystem service mapping. *Remote Sensing Applications: Society and Environment*, 8, 41-53.

Hajeb M, Karimzadeh S, Fallahi A (2020) Seismic Damage assessment in Sarpole-Zahab town (Iran) using synthetic aperture radar (SAR) images and texture. *Nature Hazards.*

Hamerly, G., Elkan, C. (2003). Learning the k in k-means. *Advances in neural information processing systems*, 16.

Hanaraj, K., Angadi, D. P. (2020). Land use land cover mapping and monitoring urban growth using remote sensing and GIS techniques in Mangaluru, India. *GeoJournal*, 1-27.

Harrison, P., Todes, A. (2001). The Use of Spatial Frameworks in Regional Development in South Africa. *Regional Studies,* 35, 62-72 *https://doi.org/10.1080/00343400120025682.*

Hasmadi, M., Pakhriazad, H. Z., Shahrin, M. F. (2009). Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografia: Malaysian Journal of Society and Space*, 5(1), 1-10.

He, J., Harris, J. R., Sawada, M., Behnia, P. (2015). A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada's Arctic. *International Journal of Remote Sensing*, 36(8), 2252–2276. https://doi.org/10.1080/01431161.2015.1035410

Hemalatha, S., Anouncia, S. (2017). Unsupervised Segmentation of Remote Sensing Images using FD Based Texture Analysis Model and ISODATA. *Int. J. Ambient Comput. Intell*., 8, 58-75. https://doi.org/10.4018/IJACI.2017070104.

Holden, C. E., Woodcock, C. E. (2016). An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. *Remote Sensing of Environment*, 185, 16-36.

Horning, N., Robinson, J. A., Sterling, E. J., Spector, S., Turner, W. (2010). Remote sensing for ecology and conservation: a handbook of techniques. Oxford University Press.

Huang, M., Shi, Z., Gong, J. (2008). Potential of Multitemporal ERS-2 SAR Imagery for Land Use Mapping in Coastal Zone of Shangyu City, China. Composition. *Remote Sensing*, 12(15), 2411.

Hutchinson, C. F. (1991). Uses of satellite data for famine early warning in sub-Saharan Africa. *International Journal of Remote Sensing*, 12(6), 1405-1421.

Jayne, T., Chamberlin, J., Headey, D. (2014). Land pressures, the evolution of farming systems, and development strategies in Africa: A synthesis. *Food Policy*, 48, 1-17.

Jensen, J. R. (2005). Introductory digital image processing: a remote sensing perspective. Pearson Education India.

Joerin, F., Thériault, M., Musy, A. (2001). Using GIS and outranking multicriteria analysis for land-use suitability assessment. *International Journal of Geographical Information Science*, 15, 153 - 174. https://doi.org/10.1080/13658810051030487.

Johnson, M.C. (2014). Nandoni Dam resettlement: Nkuzi Development Association concerns: Departmental briefing. Department of water and sanitation.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M., Kuemmerle, T., Meyfroidt, P., Mitchard, E., Reiche, J., Ryan, C., & Waske, B. (2016). A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote. Sens.*, 8, 70. https://doi.org/10.3390/RS8010070.

Julian, B, Wilson, S. (2013). A Presumed Equality: The Relationship Between State and Citizens in Post-Apartheid South Africa. *African Studies*, 72, 86-106.

Kanianska, R., Kizeková, M., Nováček, J., Zeman, M. (2014). Land-use and land-cover changes in rural areas during different political systems: A case study of Slovakia from 1782 to 2006. *Land Use Policy,* 36, 554-566.

Kaplan, G., Avdan, U. (2018). Sentinel-1 And Sentinel-2 Data Fusion for Wetlands Mapping: BalikDami, Turkey. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences,* 42(3).

Käyhkö, N., Skånes, H. (2006). Change trajectories and key biotopes: Assessing landscape dynamics and sustainability. *Landscape and Urban Planning*, 75, 300-321.

Kornijów, R. (2009). Controversies around Dam reservoirs: benefits, costs, and future. *Ecohydrology and Hydrology*, 24, 141-148.

Kusumaningrum, R., Priyana, Y. (2024). Spatial Analysis of LULC Changes and Suitability of Ngawi Regency on Spatial Regional Planning 2011-2031 Using Geographic Information System. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1357, No. 1, p. 012029). IOP Publishing.

Lauer, D. T., Morain, S. A., Salomonson, V. V. (1997). The Landsat program: Its origins, evolution, and impacts. *Photogrammetric Engineering and Remote Sensing*, 63(7), 831-838.

LeCun, Y., Bengio, Y. Hinton, G. (1995). Deep learning. *Nature*, 521, 436–444.

Lenco, D., Interdonato, R., Gaetano, R., mMinh, D. H. T. (2019). Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 11-22.

Li, P.; Jiang, L.; Feng, Z. 2014, Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors. *Remote Sensing*. 6, 310-329.

Likas, A., Vlassis, N., Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, 36(2), 451-461.

Lillesand, T. M., & Keifer, R. W. (1994). Remote Sensing and Image Interpretation. United States of America: John Willey & Sons.

Lingwal, S., Bhatia, K., Singh, M. (2022). Semantic segmentation of landcover for cropland mapping and area estimation using Machine Learning techniques. *Data Intelligence*.

Liping, C., Yujun, S., Saeed, S. (2018). Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. *PloS one*, 13(7), e0200493.

López-Serrano, P.M., Corral-Rivas, J.J., Díaz-Varela, R.A., Álvarez-González, J.G. and López-Sánchez, C.A. (2016). Evaluation of radiometric and atmospheric correction algorithms for aboveground forest biomass estimation using Landsat 5 TM data. *Remote Sensing*, 8(5), p.369.

Love, B. (2002). Comparing supervised and unsupervised category learning. Psychonomic Bulletin *and* Review, 9, 829-835. https://doi.org/10.3758/BF03196342.

Luan, C., Liu, R., Peng, S. (2021). Land-use suitability assessment for urban development using a GIS-based soft computing approach: A case study of Ili Valley, China. *Ecological Indicators*, 123, 107333. https://doi.org/10.1016/J.ECOLIND.2020.107333.

Ma, C., Zhang, H.H., Wang, X. (2004). Machine learning for Big Data analytics in plants. *Trends in Plant Science*, December 2014, Vol. 19, No. 12.

Mansour, S., Al-Belushi, M., Al-Awadhi, T. (2020). Monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-Markov modelling techniques. *Land Use Policy*, 91, 104414.

Marlina, D. (2022). Klasifikasi Tutupan Lahan pada Citra Sentinel-2 Kabupaten Kuningan dengan NDVI dan Algoritme Random Forest. STRING (Satuan Tulisan Riset dan Inovasi Teknologi).

Mather, Paul M., and Magaly Koch. Computer processing of remotely sensed images: an introduction. John Wiley and Sons, 2011.

Matosak, B., Fonseca, L., Maretto, R. (2023). Deep Learning and Cloudy Optical Time Series: A Case of Study with LSTM to Map LULC in Pantanal. IGARSS 2023 - 2023 *IEEE International Geoscience and Remote Sensing Symposium*, 7179-7182. https://doi.org/10.1109/IGARSS52108.2023.10282993.

Mekuria, W., Diyasa, M., Tengberg, A., Haileslassie, A. (2021). Effects of Long-Term Land Use and Land Cover Changes on Ecosystem Service Values: An Example from the Central Rift Valley, Ethiopia. *Land*.

Mi, J., Yang, Y., Zhang, S., An, S., Hou, H., Hua, Y., Chen, F. (2019). Tracking the land use/land cover change in an area with underground mining and reforestation via continuous Landsat classification. *Remote Sensing*, 11(14), 1719.

Michelson, D. (2000). Comparison of Algorithms for Classifying Swedish Landcover Using Landsat TM and ERS-1 SAR Data. *Remote Sensing of Environment*.

Millard, K., Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote sensing*, 7(7), 8489-8515.

Ming, D., Zhou, T., Wang, M., Tan, T. (2016). Land cover classification using random forest with genetic algorithm-based parameter optimization. *Journal of Applied Remote Sensing*, 10(3), 035021.

Mishra, V., Prasad, R., Rai, P., Vishwakarma, A., Arora, A. (2018). Performance evaluation of textural features in improving land use/land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data. *Earth Science Informatics*, 12, 71-86.

Mo, Y., Zhong, R., Sun, H., Wu, Q., Du, L., Geng, Y., & Cao, S. (2019). Integrated Airborne LiDAR Data and Imagery for Suburban Land Cover Classification Using Machine Learning Methods. *Sensors (Basel, Switzerland)*, 19.

MohanRajan, S. N., Loganathan, A., Manoharan, P. (2020). Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and Challenges. *Environmental Science and Pollution Research,* 27(24), 29900-29926.

Mucina, L., Rutherford, M. C. (2006). The Vegetation of South Africa, Lesotho, and Swaziland. South African National Biodiversity Institute. Pretoria.

Mugari, E., Masundire, H. (2022). Consistent Changes in Land-Use/Land-Cover in Semi-Arid Areas: Implications on Ecosystem Service Delivery and Adaptation in the Limpopo Basin, Botswana. *Land*.

Nazeer, M., Nichol, J. E., Yung, Y. K. (2014). Evaluation of atmospheric correction models and Landsat surface reflectance product in an urban coastal environment. *International Journal of Remote Sensing*, 35(16), 6271-6291.

Nel, E. (2011). Rethinking Patterns of South African Urban Growth: 1911–2001. *Urban Forum*, 22, 331-342.

Nghonyama, M, G. (2011). The sustainability of local economic development projects in Thulamela projects in the Vhembe District of the Limpopo Province. University of Venda.

Nicolau, A., Flores-Anderson, A., Griffin, R., Herndon, K., Meyer, F. (2021). Assessing SAR C-band data to effectively distinguish modified land uses in a heavily disturbed Amazon forest. *Int. J. Appl. Earth Obs. Geoinformation*, 94, 102214. https://doi.org/10.1016/j.jag.2020.102214.

NOAA. Historical Maps and Charts audio podcast. National Ocean Service website, https://oceanservice.noaa.gov/podcast/july17/nop08-historical-maps-charts.html, accessed on 25/10/24.

Obeng, L. (1977). Should Dams be build? The Volta Lake Example. *Ambio*, 1, 46-50.

Owojori, A., Xie, H. (2005). Landsat image based LULC changes of San Antonio, Texas using advanced atmospheric correction and object-oriented image analysis approaches. In 5th *international symposium on remote sensing of urban areas*, Tempe, AZ.

Özay, E. K., Tunga, B. (2021). A novel method for multispectral image pans sharpening based on high dimensional model representation. *Expert Systems with Applications*, 170, 114512.

Pal, M. (2005) Random Forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 217-222.

Pal, M., Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment*, 86(4), 554-565.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.

Phan, T. N., Kuch, V., Lehnert, L. W. (2020). Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image

Pontius, R., Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, 32, 4407 - 4429.

Rahaman, K. R., Hassan, Q. K., Ahmed, M. R. (2017). Pan-sharpening of Landsat-8 images and its application in calculating vegetation greenness and canopy water contents*. ISPRS International Journal of Geo-Information*, 6(6), 168.

Rajasekaran, K., Saravanan, P. (2020). Conceptual Methodology on Machine Learning and Types of Learning Algorithms. JAC: *A Journal of Composition Theory,* 13, 233-249.

Ramanamurthy, B. V., Victorbabu, N. (2021). Land Use Land Cover (LULC) classification with wasteland demarcation using remote sensing and GIS techniques. *In IOP conference series: materials science and engineering* (Vol. 1025, No. 1, p. 012035). IOP Publishing.

Reis, S, 2008. Analyzing Land Use/Land Cover Changes Using Remote Sensing and GIS in Rize, North-East Turkey. *Sensors*, 8,6188-6202.

Rogan, J., Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, 61(4), 301-325.

Ryzhok, Z. (2023). Method of creating an interactive land use map in the sentinel-2 land cover explorer web application. *International Scientific and Practical Conference.*

Schneider, J., Zabel, F., Schünemann, F., Delzeit, R., Mauser, W. (2022). Global cropland could be almost halved: Assessment of land saving potentials under different strategies and implications for agricultural markets*. PLoS ONE*.

Schowengerdt, R. A. (2006). Remote sensing: models and methods for image processing. *elsevier*.

Schumann, G., Bates, P. D., Horritt, M. S., Matgen, P., Pappenberger, F. (2009). Progress in integration of remote sensing–derived flood extent and stage data and hydraulic models. *Reviews of Geophysics*, 47(4).

Scudder, T. (2001). The World Commission on Dams and the Need for a New Development Paradigm. *International Journal of Water Resources Development*, 3, 329-341.

Sewnet, A., Abebe, G. (2018). Land use and land cover change and implication to watershed degradation by using GIS and remote sensing in the Koga watershed, North-western Ethiopia. *Earth Science Informatics*, 11(1), 99-108.

Shah, V. P., Younan, N. H., King, R. L. (2008). An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets*. IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1323-1335.

Sharma, R., Nehren, U., Rahman, S., Meyer, M., Rimal, B., Seta, G., Baral, H. (2018). Modeling Land Use and Land Cover Changes and Their Effects on Biodiversity in Central Kalimantan, Indonesia. *Land*, 7, 57

Shen, J., Han, W., Zhang, L., Li, H., Lyu, L. (2022). The Research on Quality Inspection of Land Cover Classification Based on Lidar. The International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences.*

Siciliano, G., Urban, F., Kim, S., Lonn, P.D. (2015). Hydropower, social priorities and the rural–urban development divide: The case of large Dams in Cambodia. *Energy Policy*, 86, 73–285.

Singh, A., Ganapathy, B., Singh, A.K., Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, February 2016, Vol. 21, No. 2.

Singh, G., Pandey, A. (2021). Evaluation of classification algorithms for land use land cover mapping in the snow-fed Alaknanda River Basin of the Northwest Himalayan Region. *Applied Geomatics*, 13(4), 863-875.

Sisodia, P. S., Tiwari, V., Kumar, A. (2014) Analysis of Supervised Maximum Likelihood Classification for remote sensing image*. International Conference on Recent Advances and Innovations in Engineering*, 1-4.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Li, C. L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems,* 33, 596-608.

South African History Online, (2011). Accessed on the 21st of October 2022 from https://www.sahistory.org.za/place/venda.

Soyapi, C. B. (2017). Water security and the right to water in Southern Africa: an overview*. Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad,* 20(1).

Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology,* 59(1), 1-34.

Sterk, G., Raju, P. L. N., Amit, B. (2007). Comparing and optimizing land use classification in a Himalayan area using parametric and non-parametric approaches. *Journal of Geomatics,* 1(1), 23-23.

Sun, X., Yang, Z. (2008, March). Generalized McNemar's test for homogeneity of the marginal distributions. In SAS Global forum (Vol. 382, pp. 1-10). Lantz, B.

(2019). Machine learning with R: expert techniques for predictive modelling. Packt publishing ltd.

Sun, X., Yang, Z. (2008, March). Generalized McNemar's test for homogeneity of the marginal distributions. In SAS Global forum (Vol. 382, pp. 1-10). Srinivasa, V., Lambin, E.F., Gorelick, S.M., Thomson, B.H., Rozelle, S. (2012). The nature and causes of the global water crisis: Syndromes from a meta-analysis of coupled human-water studies. Water Resources Research, 48.

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y. A., Rahman, A. (2020a). Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sensing*, 12(7), 1135.

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y. A., Rahman, A. (2020b). Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sensing*, 12(7), 1135.

Tan, X., Jiang, S., Zheng, Z., Zhang, P., Zhu, M., He, Y., Yu, Z., Wang, N., Jiang, L., Zhou, G., Zhang, H., Li, J. (2018). A Manifold Learning Approach of Land Cover Classification for Optical and SAR Fusing Data. IGARSS 2018 - 2018 *IEEE International Geoscience and Remote Sensing Symposium*.

Tansock, J., Bancroft, D., Butler, J., Cao, C., Datla, R., Hansen, S., Helder, D., Kacker, R., Latvakoski, H., Mylnczak, M., and Murdock, T. (2015). Guidelines for radiometric calibration of electro-optical instruments for remote sensing.

Tayyebi, A., Pijanowski, B. C., Linderman, M., Gratton, C. (2014). Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world. *Environmental Modelling & Software,* 59, 202-221.

Tempelhoff, N.J. (2007). Water procurement and the environment: A comparative history of South Africa's Nandoni and De Hoop Dams 1994-2007. School of basic sciences. Northwest University.

Thanh Noi, P., Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 18.

Thonfeld, F., Steinbach, S., Muro, J., Kirimi, F. (2020). Long-term land use/land cover change assessment of the Kilombero catchment in Tanzania using random forest classification and robust change vector analysis. *Remote sensing*, 12(7), 1057.

Tou, J. T., and R. C. Gonzalez. Pattern Recognition Principles, Addison-Wesley Publishing Company, Reading, Massachusetts (1974).

Tsang, L., Kong, J. A., Shin, R. T. (1985). Theory of microwave remote sensing.

Uhlmann, S., Kiranyaz, S. (2014). Classification of dual- and single polarized SAR images by incorporating visual features. *Isprs Journal of Photogrammetry and Remote Sensing*.

Ulaby, F. T., Moore, R. K., Fung, A. K. (1981). Microwave remote sensing: Active and passive. volume 1-microwave remote sensing fundamentals and radiometry.

Veronesi, F., Schillaci, C. (2019). Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecological Indicators*, 101, 1032-1044.

Walz, U. (2015). Indicators to monitor the structural diversity of landscapes. *Ecological Modelling*, 295, 88-106. https://doi.org/10.1016/J.ECOLMODEL.2014.07.011.

Wickham, J., Stehman, S., Sorenson, D., Gass, L., & Dewitz, J. (2023). Thematic accuracy assessment of the NLCD 2019 land cover for the conterminous United States. *GIScience & Remote Sensing*, 60. https://doi.org/10.1080/15481603.2023.2181143.

Wolf, L., Shashua, A., Geman, D. (2005). Feature Selection for Unsupervised and Supervised Inference: The Emergence of Sparsity in a Weight-Based Approach. *Journal of Machine Learning Research*, 6(11).

Writer, S. (2015). Who is using all the water in South Africa? Business tech, Accessed on the 25 of July 2022 from frohttps://businesstech.co.za/news/general/104441/who-is-using-all-the-water-in-south-africa/.

Wu, Q., Zhong, R., Zhao, W., Song, K., Du, L. (2019). Land-cover classification using GF-2 images and airborne lidar data based on Random Forest. *International Journal of Remote Sensing*, 40(5-6), 2410-2426.

Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., Woodcock, C. E. (2008). Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment*, 112(3), 955-969.

Yuan, J., Niu, Z. (2008). Evaluation of atmospheric correction using FLAASH. In 2008 *International Workshop on Earth Observation and Remote Sensing Applications* (pp. 1-6). *IEEE*.

Yuan, L., Zhu, G., Xu, C. (2020). Combining synthetic aperture radar and multispectral images for land cover classification: a case study of Beijing, China. *Journal of Applied Remote Sensing*.

Zhang, R., Tang, X., You, S., Duan, K., Xiang, H., Luo, H. (2020). A Novel Feature-Level Fusion Framework Using Optical and SAR Remote Sensing Images for Land Use/Land Cover (LULC) Classification in Cloudy Mountainous Area. *Applied Sciences*, 10, 2928.

Zhong, Y., Cao, Q., Zhao, J., Ma, A., Zhao, B., & Zhang, L. (2017). Optimal Decision Fusion for Urban Land-Use/Land-Cover Classification Based on Adaptive Differential Evolution Using Hyperspectral and LiDAR Data. *Remote. Sens.*, 9, 868. https://doi.org/10.3390/rs9080868.

Appendix.

Figure 12: SAR, Current slicing Sensor Issue in Landsat 7 Sensors.

Critical values of chi-square (right tail)

Degrees of freedom (*df*)	Significance level (α)							
	.99	.975	.95	.9	.1	.05	.025	.01
1	-------	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379
70	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329
100	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116
1000	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807

Figure 13: Chi-Square Distribution Table.