

MALAWI UNIVERSITY OF SCIENCE AND TECHNOLOGY

NDATA SCHOOL OF CLIMATE AND EARTH SCIENCE

DEPARTMENT OF EARTH SCIENCES

COMPARATIVE ASSESSMENT OF RANDOM FOREST AND SUPPORT VECTOR MACHINE FOR LULC CLASSIFICATION AND CNN AND CA-MARKOV FOR PREDICTION OF LULC DYNAMICS IN DEDZA BOMA, MALAWI

SUBMITTED BY:

AJASI SAIDI (GIS/027/21)

RESEARCH DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE BACHELOR OF SCIENCE IN GEO-INFORMATION AND EARTH OBSERVATION SCIENCE.

SUPERVISOR: MR. DICKSON MBEYA

20 MAY, 2026

Declaration

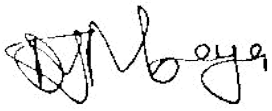
I, AJASI SAIDI, hereby declare this research project is entirely our original work and has never been submitted to this institution or any other institution for any purpose. All the sources of information used in this report have been appropriately acknowledged.

Signatures:

AJASI SAIDI

RESEARCH DISSERTATION APPROVED BY:

SUPERVISOR:

MR. DICKSON MBEYA Signature: 

ASSESSOR:

_____ Signature: _____

Disclaimer

This document describes work undertaken as part of a study in Bachelor of Science in Geo-information and Earth Observation Science at Ndata School of Climate and Earth Sciences of the Malawi University of Science and Technology. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the School.

Copyright Statement

This copy of the dissertation has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with its author and that no quotation from the dissertation and no information derived from it may be published without the acknowledgement of the author.

Acknowledgements

First and foremost, I give all glory and honor to the Almighty God for the precious gift of life, strength, opportunities, countless blessings, and His unfailing grace and love throughout these four years of study; without Him, this milestone would have remained a dream, and all praise and thanksgiving belong to Him. To my family members, and to all my classmates, friends, and lecturers woven between the lines of this dissertation are the imprints of your care and encouragement: your smiles, your financial backing, and your simple everyday check-ins have meant more than words could ever capture, so from the bottom of my heart, thank you.

A deeply special acknowledgment goes to my supervisor, Mr. Dickson Mbeya, for his unwavering commitment to my project. Thank you for the thought-provoking questions, the wise guidance, the mentorship, and for those reassuring "no pressure" moments that always came when I needed them most; may you receive the recognition you truly deserve. Working with you has never been a burden but a genuine pleasure, and I admit I always looked for the smallest excuse to meet you, knowing I would walk away from every conversation feeling sharper and more inspired than before.

Abstract

Land Use and Land Cover (LULC) change is a major environmental concern in rapidly developing towns such as Dedza Boma, Malawi. Understanding how land cover has changed over time is essential for informing sustainable urban planning and natural resource management. This study conducted a comparative assessment of machine learning approaches for LULC classification and prediction in Dedza Boma for the years 2015, 2020, and 2025. Sentinel-2 satellite imagery was classified using two algorithms Random Forest (RF) and Support Vector Machine (SVM) on the Google Earth Engine (GEE) platform. Accuracy assessment revealed that RF outperformed SVM across all time periods, achieving overall accuracies of 90.0% (2015), 91.0% (2020), and 92.0% (2025) with Kappa coefficients of 0.87, 0.88, and 0.89 respectively, compared to SVM accuracies of 86.0% (2015), 87.0% (2020), and 88.0% (2025). RF was therefore selected as the best-performing classification algorithm.

Post-classification change detection revealed notable land cover transitions over the study period. Major changes included the conversion of farmland and vegetated areas into built-up land, indicating rapid urban expansion within the town. Additional shifts from vegetation to bare land and farmland were also observed, demonstrating increasing pressure on natural resources. For future LULC prediction, two approaches were comparatively assessed: a Convolutional Neural Network (CNN) deep learning model and the Cellular Automata–Markov (CA–Markov) model. CNN outperformed CA–Markov, achieving an overall accuracy of 89.0% (Kappa: 0.87) compared to CA–Markov's 85.0% (Kappa: 0.82). Future LULC scenarios produced using CNN showed continued growth of built-up areas and further declines in farmland and vegetation by 2030 and 2035.

The overall findings provide up-to-date and reliable spatial information that can guide planners, policymakers, and environmental managers in making informed decisions. The study concludes that RF is the best-performing classification algorithm and CNN is the superior prediction model for LULC dynamics in Dedza Boma. The predicted LULC scenarios offer valuable insights for sustainable land use planning and future development in the study area.

Table of Contents

Declaration I

Disclaimer III

Copyright Statement IV

Acknowledgements V

Abstract VI

List of Figures IX

List of Tables 2

List of Equations 2

Abbreviations 3

INTRODUCTION 4

1.1 Background of the Study 4

1.2 Problem Statement 6

1.3 Objectives 7

1.3.1 Main Objective 7

1.3.2 Specific Objectives 7

1.4 Research Questions 7

1.5 Significance of the Study 8

Chapter 2: Literature Review 9

2.1 Land Use and Land Cover 9

2.2 Drivers of Land Use and Land Cover Change 10

2.3 Impacts of LULC Change 11

2.4 Remote Sensing and GIS in LULC Mapping 12

2.5 Machine Learning Approaches for LULC Classification 13

2.5.1 Overview of Supervised Classification 13

2.5.2 Random Forest (RF) 13

2.5.3 Support Vector Machine (SVM) 15

2.6 Change Detection in LULC Studies 16

2.7 Predictive Modelling of LULC Changes 17

2.7.1 Overview of Predictive Modelling Approaches 17

2.7.2 Cellular Automata-Markov (CA-Markov) Model 18
2.7.4 Machine Learning for LULC Prediction 18
Chapter 3: Methodology 21
3.1 Research Design 21
3.2 Study Area 22
3.3 Materials and Methods 23
3.4 LULC Classification 25
Chapter Four: Results and Discussion 30
4.2.1 LULC Classification Results 30
4.2.2 Accuracy Assessment Results 33
4.2.3 LULC Change Detection Results 35
4.2.4 Predicted LULC Results 37
5.1.1Classification Performance: Random Forest vs Support Vector Machine 41
5.2.2 LULC Change Dynamics (2015–2025) 42
5.3 Prediction Model Performance: CNN vs CA-Markov 44
5.4 Future LULC Scenarios (2030 and 2035) 44
5.5 Comparison with Previous Studies 45
5.6 Limitations of the Study 46
Chapter Five: Conclusion and Reccommendations 48
REFERENCES 49

List of Figures

Figure 3.1: Map of Dedza Boma (Study Area)

Figure 4.1: 2015 Land Use and Land Cover (LULC) Map of Dedza Boma

Figure 4.2: 2020 Land Use and Land Cover (LULC) Map of Dedza Boma

Figure 4.3: 2025 Land Use and Land Cover (LULC) Map of Dedza Boma

Figure 4.4: Bar Chart showing LULC Class Areas (km²) for 2015, 2020, and 2025

Figure 4.5: SVM LULC Classification Maps for 2015, 2020, and 2025

Figure 4.5: Change Detection Map for Dedza Boma (2015–2025)

Figure 4.7: Predicted LULC Maps and Validation (2030 & 2035)

List of Tables

Table 3.1: Summary of datasets used in the study
Table 4.1: LULC class areas (km² and %) for Dedza Boma (2015, 2020, and 2025)
Table 4.2: Accuracy metrics for RF and SVM classifications (2015–2025)
Table 4.3: Confusion matrix for 2015 Random Forest classification
Table 4.4: Confusion matrix for 2020 Random Forest classification
Table 4.5: Confusion matrix for 2025 Random Forest classification
Table 4.6: Net LULC change (2015–
2025)..... 43
Table 4.7: Change detection categories (2015–2025)
Table 4.8: LULC transition matrix (2015–2025) (km²)
Table 4.9: Change detection validation results
Table 4.10: Comparison of CNN and CA–Markov prediction accuracy
Table 4.11: CNN Projected LULC Area (2030 & 2035)
Table 4.12: CA-Markov Projected LULC Area (2030 & 2035)

List of Equations

Equation (1): Random Forest Majority Voting Rule (RF Classification)
Equation (2): SVM Decision Boundary (Support Vector Machine Classification)

Abbreviations

ANN Artificial Neural Network

CA Cellular Automata

CA–Markov Cellular Automata–Markov Chain Model

CNN Convolutional Neural Network

CRS Coordinate Reference System

DN Digital Number

GEE Google Earth Engine

GIS Geographic Information Systems

LSTM Long Short-Term Memory

LULC Land Use and Land Cover

MLC Maximum Likelihood Classification

NIR Near-Infrared

OA Overall Accuracy

PA Producer's Accuracy

RBF Radial Basis Function

RF Random Forest

RS Remote Sensing

SSA Sub-Saharan Africa

SVM Support Vector Machine

SWIR Short-Wave Infrared

UA User's Accuracy

INTRODUCTION

1.1 Background of the Study

Land Use and Land Cover (LULC) change is a major global environmental challenge, driven largely by population growth, rapid urbanisation, agricultural expansion, and economic development. Worldwide, natural landscapes continue to be transformed at an unprecedented pace, affecting ecological balance, biodiversity, climate regulation, and the availability of natural resources (Lambin et al., 2001). Over the past decades, global studies have demonstrated that changes in land cover significantly influence carbon storage, water cycles, and surface temperatures, making LULC monitoring an essential component of environmental management and climate change adaptation (Foley et al., 2005).

Sub-Saharan Africa (SSA) is experiencing some of the fastest rates of LULC change globally, with rural and peri-urban areas under increasing pressure from population growth, demand for agricultural land, fuelwood collection, and unplanned settlement development (Kamwi et al., 2024; Sanogo et al., 2022; Yiran et al., 2020).. The region continues to face persistent challenges related to deforestation, soil degradation, and urban sprawl, all of which threaten food security, water resources, and ecological stability (Veldkamp & Verburg, 2004). Furthermore, weak land governance systems, limited enforcement of land use policies, and heavy reliance on natural resources for rural livelihoods further accelerate land transformations across SSA.

Malawi, like many countries in the region, is undergoing rapid demographic and socio-economic changes that place significant pressure on land resources. The country's population has continued to rise steadily, increasing demand for agricultural land, housing, and infrastructure. As a result, many urban and semi-urban centres in Malawi are experiencing notable shifts in land cover, including the expansion of built-up areas, reduction in vegetation cover, and conversion of farmland into settlements. These changes have profound implications for watershed health, biodiversity conservation, urban planning, and the long-term sustainability of natural resources (Government of Malawi, 2019).

Dedza District, located in the Central Region of Malawi, is among the areas experiencing these pressures. Dedza Boma, the district's administrative and commercial hub, lies along the M1 road and serves as a key transportation corridor connecting Malawi's northern and southern regions. Its strategic location has contributed to rapid urbanisation, increased economic activity, and significant population growth in recent years. As a result, the town has witnessed substantial land use transitions, including the expansion of built-up areas, reduction of vegetated land, and increased demand for land for housing, agriculture, and commercial purposes. These transformations highlight the

urgent need for reliable spatial information to support sustainable planning and development.

Remote Sensing (RS) and Geographic Information Systems (GIS) have become essential tools for assessing LULC dynamics, providing multispectral and multi-temporal data that enable accurate and cost-effective monitoring of land changes (Guerschman et al., 2003). The emergence of cloud-based platforms such as Google Earth Engine (GEE), combined with machine learning algorithms like Random Forest (RF) and Support Vector Machine (SVM), has significantly improved the speed, accuracy, and accessibility of LULC classification.

These technologies allow researchers to analyse large volumes of satellite imagery, assess spatial patterns, quantify land transformations, and generate high-resolution maps that inform evidence-based decision-making.

For LULC prediction, various modelling approaches have been developed, ranging from traditional transition-based models such as Cellular Automata–Markov (CA–Markov) to advanced deep learning techniques such as Convolutional Neural Networks (CNNs). While CA–Markov models effectively capture historical transition probabilities and project future land cover based on past trends (Eastman, 2012), CNNs offer the ability to learn complex spatial patterns and non-linear relationships directly from satellite imagery, potentially yielding more accurate predictions in heterogeneous landscapes (Kamusoko & Gamba, 2015; Mungai et al., 2022). Despite the potential of these advanced techniques, their application and comparative assessment in local-scale LULC prediction studies within Malawi remain limited.

Despite increased interest in LULC studies within Malawi, most analyses have focused on broad district or regional scales, with limited attention given to smaller urban centres such as Dedza Boma. Yet, these areas are among the most rapidly transforming landscapes, and the lack of detailed spatial data makes it difficult for planners and policymakers to manage land resources effectively. Understanding past and present LULC dynamics at the local scale is therefore essential for anticipating future changes and guiding sustainable land use planning. Furthermore, there is a notable gap in comparative assessments of different prediction models specifically between traditional CA–Markov and deep learning CNN approaches for LULC forecasting in Malawian urban contexts.

This study addresses these gaps by analysing LULC changes in Dedza Boma from 2015 to 2025 using geospatial techniques and machine learning classification approaches, comparing the performance of Random Forest and Support Vector Machine for LULC classification and comparing the predictive accuracy of CA–Markov and Convolutional Neural Network (CNN) models for forecasting future LULC scenarios. The insights

gained from this assessment are expected to assist local authorities, urban planners, and environmental managers in making informed decisions regarding land allocation, natural resource protection, and sustainable urban development.

1.2 Problem Statement

Dedza Boma has experienced notable Land Use and Land Cover (LULC) changes due to population growth, urban expansion, and demand for agriculture and infrastructure. Malawi's population is growing at approximately 2.7% annually, placing pressure on land resources (National Statistical Office, 2018). Consequently, natural landscapes are being converted into built-up areas and farmland, contributing to land degradation, vegetation loss, and ecosystem pressure (Lambin et al., 2001; Foley et al., 2005).

Despite this, up-to-date spatial information on LULC dynamics in Dedza Boma is limited. Most Malawian studies focus on broader regional scales, overlooking smaller urban centres where land transformations are most rapid, limiting effective land use planning. A fundamental gap exists in the comparative evaluation of modelling approaches for LULC analysis in this context. For classification, although Random Forest (RF) and Support Vector Machine (SVM) have been successfully used in various settings (Belgiu & Drăguţ, 2016; Mountrakis et al., 2011), no study has systematically compared their performance for Dedza Boma's heterogeneous landscape. It remains unclear which algorithm yields higher classification accuracy.

For prediction, the gap is even more pronounced. While traditional CA–Markov models have been widely used for LULC forecasting (Eastman, 2012), recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), offer the potential to capture complex spatial patterns that CA–Markov cannot (Kamusoko & Gamba, 2015; Mungai et al., 2022). To date, no study has compared CNN against CA–Markov for LULC forecasting in Dedza Boma, and it is unknown whether deep learning translates into meaningful improvements in prediction accuracy.

The absence of comparative assessments for both classification algorithms and prediction models creates a knowledge gap that hinders effective land use planning. Without knowing which classifier performs best and which prediction model produces the most

reliable future scenarios, planners cannot make informed decisions about model selection.

Therefore, this study addresses these gaps by analysing historical LULC changes in Dedza Boma (2015–2025) comparing classification performance of RF and SVM, and comparing prediction accuracy of CNN and CA–Markov for forecasting future LULC to 2035. The findings will support sustainable development and guide model selection for LULC analysis in similar Malawian urban contexts.

1.3 Objectives

1.3.1 Main Objective

To comparatively assess RF & SVM for LULC Classification and CNN and CA-Markov For prediction of LULC in Dedza Boma.

1.3.2 Specific Objectives

- To compare Random Forest (RF) and Support Vector Machine (SVM) classification accuracy in Google Earth Engine.
- To quantify LULC changes and transitions in Dedza Boma from 2015 to 2025
- To compare Convolutional Neural Network (CNN) and CA–Markov prediction accuracy for forecasting future LULC patterns
- To generate future LULC scenarios for Dedza Boma for 2030 and 2035

1.4 Research Questions

i. Which algorithm (RF or SVM) performs better for LULC classification in Dedza Boma?

ii. How have LULC patterns changed in Dedza Boma from 2015 to 2025?

iii. Which prediction model (CNN or CA–Markov) produces more accurate LULC forecasts for Dedza Boma?

iv. What future LULC changes are expected in Dedza Boma by 2030 and 2035?

1.5 Significance of the Study

This study provides a systematic comparative assessment of machine learning and deep learning models for analysing and predicting Land Use and Land Cover (LULC) changes using Remote Sensing (RS) and Geographic Information Systems (GIS). By comparing Random Forest (RF) and Support Vector Machine (SVM) for classification, and Convolutional Neural Network (CNN) and CA–Markov for prediction, the study identifies the best-performing models for LULC analysis in Dedza Boma. This offers a reliable and cost-effective approach for monitoring land dynamics compared to traditional methods.

The results will help planners, policymakers, and local authorities better understand how land use is changing in Dedza Boma and support informed decision-making in areas such as urban planning, land allocation, and environmental management. The comparative findings will guide model selection, ensuring that future LULC assessments use the most accurate algorithms and prediction tools available for this landscape. This can contribute to improved management of natural resources and sustainable urban development.

Chapter 2: Literature Review

Introduction

This chapter reviews existing literature on Land Use and Land Cover (LULC) change, its drivers and impacts, and the application of Remote Sensing (RS) and Geographic Information Systems (GIS) for LULC analysis. It examines machine learning algorithms for classification, including Random Forest (RF) and Support Vector Machine (SVM), as well as prediction models such as Convolutional Neural Networks (CNN) and CA–Markov. The chapter concludes by identifying research gaps in LULC studies in Dedza Boma and Malawi, which this study aims to address.

2.1 Land Use and Land Cover

Land use and land cover are closely related concepts and are often used interchangeably, although they describe different aspects of the Earth's surface (Fairbanks et al., 2000; Ellis, 2013). Land cover refers to the physical and biological materials on the Earth's surface, such as vegetation, water bodies, bare soil, and built-up areas (Di Gregorio & Jansen, 2000; Brown & Duh, 2004). In contrast, land use describes how humans utilize and manage land for activities such as agriculture, settlement, and forestry (Turner et al., 1995).

While natural scientists tend to define land cover based on biophysical characteristics, social scientists often link it to socio-economic functions. Similarly, land use emphasizes human intentions and activities that modify land cover. According to Di Gregorio and Jansen (2000), land use involves the management and transformation of land cover to meet human needs, often influencing ecological processes such as hydrology, biodiversity, and soil quality.

LULC change refers to both qualitative and quantitative transformations in land systems (Seto et al., 2002). These changes can occur through conversion, where one land cover type is replaced by another, or modification, where the structure or function of land changes without altering its overall class (Lambin et al., 2003). Understanding these processes is essential, as LULC changes significantly affect environmental sustainability, natural resource management, and socio-economic development

2.2 Drivers of Land Use and Land Cover Change

Land use and land cover (LULC) changes are influenced by a combination of socio-economic, environmental, and institutional drivers (Lambin et al., 2003; Geist & Lambin, 2002). Understanding these drivers is essential for explaining spatial and temporal patterns of land transformation.

One of the major drivers of LULC change is population growth, which increases demand for land for agriculture, settlement, and infrastructure development (Lambin et al., 2003; Foley et al., 2005). As population increases, natural landscapes are often converted into cultivated and built-up areas.

Agricultural expansion is another critical driver, particularly in developing countries where subsistence farming is dominant. Expansion of cropland often leads to deforestation, soil degradation, and loss of biodiversity (Geist & Lambin, 2002; Brink & Eva, 2009).

Urbanization also contributes significantly to LULC change. Rapid urban growth leads to the conversion of agricultural and natural land into built-up areas, resulting in environmental degradation and increased pressure on resources (Seto et al., 2002; Veldkamp & Verburg, 2004).

In Sub-Saharan Africa, poverty and dependence on natural resources are key drivers of land degradation. In Malawi, rural communities rely heavily on forests for fuelwood and livelihoods, which accelerates deforestation and land cover change (Kamanga et al., 2009; Sedano et al., 2005).

Environmental factors such as climate variability and change also influence LULC dynamics by affecting vegetation patterns and land productivity (Lambin et al., 2003; Foley et al., 2005). Additionally, weak land management policies and inadequate planning systems contribute to unsustainable land use practices (Geist & Lambin, 2002). These drivers often interact in complex ways. For example, population growth may lead to agricultural expansion, which in turn accelerates deforestation and environmental

degradation. Such interactions highlight the need for integrated approaches in LULC analysis and modeling (Lambin et al., 2003).

2.3 Impacts of LULC Change

LULC changes have significant environmental and socio-economic impacts, particularly in developing regions. Environmentally, these changes contribute to biodiversity loss, soil degradation, and disruption of hydrological systems (Turner et al., 2007).

For instance, deforestation reduces vegetation cover, leading to decreased groundwater recharge and increased risk of water scarcity (Dwivedi et al., 2005). Similarly, urban expansion results in habitat fragmentation and increased pollution, which negatively affects ecosystems (Veldkamp & Verburg, 2004).

The socio-economic impacts of LULC change are especially pronounced in Sub-Saharan Africa, where rural communities depend heavily on natural resources for their livelihoods. In Malawi, poverty and land insecurity drive the overexploitation of forest resources, leading to environmental degradation and reduced agricultural productivity (Kamanga et al., 2009).

These environmental and socio-economic impacts are closely interconnected. For example, deforestation can lead to soil erosion, which reduces crop yields and ultimately contributes to food insecurity and poverty. This highlights the importance of sustainable land management practices to mitigate the negative consequences of LULC change.

2.4 Remote Sensing and GIS in LULC Mapping

Remote sensing (RS) and Geographic Information Systems (GIS) have become essential tools for analyzing and monitoring land use and land cover (LULC) changes. Remote sensing involves the acquisition of information about the Earth's surface without direct physical contact, typically through satellite or aerial sensors, while GIS provides a framework for storing, analyzing, and visualizing spatial data (Jensen, 2005; Campbell & Wynne, 2011).

Traditionally, LULC mapping relied on field-based methods, including ground surveys, manual observations, and interpretation of topographic maps and aerial photographs. These approaches, although useful, are often time-consuming, expensive, and limited in spatial and temporal coverage (Lillesand et al., 2015). In addition, traditional methods are difficult to apply over large or inaccessible areas and may lack consistency due to human subjectivity.

In contrast, remote sensing provides significant advantages over these traditional approaches. It enables large-scale and repetitive data acquisition, allowing for continuous monitoring of land surface changes over time (Turner et al., 2007). Satellite platforms such as Landsat and Sentinel offer long-term, freely available datasets with moderate to high spatial resolution, making them highly suitable for detecting and analyzing LULC dynamics (Wulder et al., 2018).

Furthermore, GIS enhances the analytical capability of remote sensing data by allowing integration with other spatial datasets, such as population data, elevation models, and land management information. This integration supports more comprehensive analysis and better decision-making in land resource management.

Recent advancements in cloud computing have further transformed LULC analysis through platforms such as Google Earth Engine, which provide access to vast archives of satellite imagery and powerful processing capabilities (Gorelick et al., 2017). GEE enables efficient processing of multi-temporal datasets and supports advanced analytical

techniques, including machine learning, making it particularly valuable for large-scale LULC studies.

Despite these advantages, remote sensing also has limitations. For example, cloud cover can obstruct satellite imagery, particularly in tropical regions, while spatial and spectral resolution constraints may affect the accuracy of classification results (Campbell & Wynne, 2011). Additionally, classification errors may occur in heterogeneous landscapes where different land cover types exhibit similar spectral characteristics. These challenges highlight the importance of using robust classification methods and conducting accuracy assessments to ensure reliable results.

2.5 Machine Learning Approaches for LULC Classification

2.5.1 Overview of Supervised Classification

Supervised classification is the dominant approach for LULC mapping from satellite imagery, involving the training of a classification algorithm on labelled reference data (training samples) and the subsequent application of the trained model to classify the full image into discrete LULC categories (Lillesand et al., 2015).The accuracy of supervised classification depends on the quality and representativeness of the training data, the choice of classification algorithm, and the spectral separability of the LULC classes under consideration.

Traditional supervised classifiers such as Maximum Likelihood Classification (MLC) have been widely used but are limited in their ability to handle high-dimensional feature spaces and non-linear class boundaries (Pal and Mather, 2005). Machine learning classifiers have increasingly superseded traditional approaches by offering superior performance across diverse landscape contexts.

2.5.2 Random Forest (RF)

Random Forest (RF) is an ensemble machine learning algorithm developed by Breiman (2001) that constructs a large number of decision trees during training and outputs the class that is the mode of the individual tree classifications. RF is widely recognized in the

remote sensing literature as one of the most accurate, robust, and computationally efficient classifiers available for LULC mapping, offering several advantages over traditional methods (Belgiu and Drăguț, 2016). These advantages include resistance to overfitting, the ability to handle high-dimensional datasets with many input features, robustness to noise and outliers, and the provision of feature importance estimates that enable the identification of the most informative spectral and spatial variables for classification (Breiman, 2001; Pal, 2005).

In their comprehensive review of RF applications in remote sensing, Belgiu and Drăguț (2016) demonstrated that RF consistently achieves high classification accuracy across a wide range of land cover types, spatial resolutions, and geographic contexts. RF has been successfully applied to LULC mapping using Landsat, Sentinel-2, and other optical sensors, and has shown particular effectiveness in heterogeneous landscapes where class boundaries are complex or overlapping. Importantly, Belgiu and Drăguț (2016) highlighted that RF's performance is relatively insensitive to the size of the training dataset compared to other classifiers, making it well-suited for data-scarce environments such as Dedza Boma, where the collection of large numbers of ground truth samples may be logistically challenging.

The final predicted class for each pixel is determined by aggregating the votes of all decision trees in the forest, as shown in Equation (1):

(1) $\hat{y} = \text{mode}\{T_1(x), T_2(x), \ldots, T_n(x)\}$

Where:

$\hat{y}$ = final predicted class for pixel x

$T_1(x), T_2(x), \ldots, T_n(x)$ = class predictions from individual decision trees

x = input feature vector (spectral values of a pixel)

n = total number of trees in the forest

mode = majority voting function selecting the most frequent class prediction

2.5.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a kernel-based supervised classifier that seeks to identify the optimal hyperplane separating different land cover classes in feature space by maximizing the margin between classes (Vapnik, 1995; Mountrakis et al., 2011). SVM has been extensively applied in remote sensing for LULC classification and has demonstrated consistently high accuracy, particularly in contexts involving limited training samples and complex, high-dimensional feature spaces (Pal and Mather, 2005). SVM's ability to effectively model non-linear class boundaries through the use of kernel functions, particularly the Radial Basis Function (RBF) kernel, makes it a powerful tool for classifying spectrally diverse landscapes.

Numerous comparative studies have evaluated the relative performance of RF and SVM for LULC classification. Pal and Mather (2005) found that both classifiers outperformed Maximum Likelihood Classification, with SVM performing particularly well with small training datasets. Mountrakis et al. (2011) reviewed SVM applications in remote sensing and concluded that SVM offers high accuracy but can be computationally intensive and sensitive to the choice of kernel parameters. The SVM decision boundary is expressed as:

$$w \cdot x + b = 0 \qquad (2)$$

Where:

w = weight vector defining the orientation of the hyperplane

x = input feature vector (spectral values of a pixel)

b = bias term controlling the position of the hyperplane

$w \cdot x$ = dot product between the weight vector and input features

The comparative evaluation of RF and SVM within the GEE environment, in the specific context of a small-scale African urban centre, is a methodological contribution that this study makes to the literature.

2.5.4 Accuracy Assessment

The accuracy of LULC classifications is conventionally assessed using a confusion matrix, also known as an error matrix, which cross-tabulates the classified outputs against reference data collected independently from the training samples (Congalton and Green, 2009).

From the confusion matrix, several accuracy metrics are derived, including Overall Accuracy (OA), which measures the proportion of correctly classified pixels across all classes; Producer's Accuracy (PA), which indicates the probability that a given area is correctly identified; User's Accuracy (UA), which indicates the probability that a pixel classified as a given class actually belongs to that class; and the Kappa Coefficient, which corrects for chance agreement and provides a more conservative estimate of classification performance (Foody, 2002).

The F1-Score, which is the harmonic mean of precision and recall, provides an additional measure of classifier performance particularly useful when class imbalances are present. A minimum Overall Accuracy of 85% is widely accepted as the threshold for reliable LULC mapping (Anderson et al., 1976).

2.6 Change Detection in LULC Studies

Change detection is a fundamental component of Land Use and Land Cover (LULC) analysis, as it enables the identification and quantification of changes in land cover over time. It involves the comparison of multi-temporal datasets to determine the nature, extent, and rate of changes occurring within a given landscape (Lu et al., 2004; Jensen, 2005).

Several change detection techniques have been developed, broadly categorized into pre-classification and post-classification methods. Pre-classification methods involve direct comparison of spectral values between images acquired at different times, while post-classification methods compare classified maps from different time periods. Among these, post-classification comparison is the most widely used approach due to its ability to provide detailed “from–to” change information between land cover classes (Singh, 1989).

Although pre-classification methods are computationally efficient, they are sensitive to atmospheric conditions, sensor differences, and seasonal variations, which can affect the accuracy of results. In contrast, post-classification methods reduce these limitations by analysing already classified images. However, their accuracy depends heavily on the quality of the initial classifications.

Recent advancements in geospatial technologies, particularly the integration of change detection techniques with Geographic Information Systems (GIS) and cloud-based platforms such as Google Earth Engine, have improved the efficiency and accuracy of LULC change analysis. These tools enable automated processing of large datasets and facilitate the generation of change matrices, which summarize transitions between different land cover classes over time.

Despite these advancements, challenges remain in accurately capturing complex land transformations, especially in heterogeneous landscapes (Karakizi et al., 2025; Fromm et al., 2025). Misclassification errors in earlier stages can propagate into change detection results, leading to uncertainties in interpretation (Li et al., 2024; Pitkänen et al., 2024). Therefore, ensuring high classification accuracy is critical for reliable change detection outcomes (Olofsson et al., 2014; Domingues et al., 2026)

In this study, a post-classification comparison approach is adopted to analyse LULC changes between 2015, 2020, and 2025. This method allows for the identification of specific transitions, such as conversion from vegetation to built-up areas, providing a clear understanding of land transformation patterns within the study area.

2.7 Predictive Modelling of LULC Changes

2.7.1 Overview of Predictive Modelling Approaches

Predictive LULC modelling seeks to simulate future land cover conditions based on historical patterns of change, spatial drivers, and transition probabilities. A wide range of modelling approaches has been developed and applied in the literature, including statistical models, agent-based models, cellular automata models, and hybrid approaches that combine elements of multiple frameworks (Veldkamp and Verburg, 2004; Eastman,

2006). The choice of modelling approach depends on the specific research objectives, the availability of input data, the spatial and temporal scale of analysis, and the degree of complexity and mechanistic detail required.

2.7.2 Cellular Automata-Markov (CA-Markov) Model

The Cellular Automata-Markov (CA-Markov) model is one of the most widely used approaches for predictive LULC modelling, combining the spatial simulation capabilities of cellular automata with the stochastic transition probability estimation of Markov Chain analysis (Pontius and Malanson, 2005; Eastman, 2006). The Markov Chain component estimates transition probability matrices from historical LULC data, quantifying the likelihood of each land cover class transitioning to another class over a defined time period. The CA component then applies these probabilities spatially, accounting for neighbourhood influences and spatial contiguity to simulate the spatial distribution of future LULC conditions (Kamusoko and Gamba, 2015).

The CA-Markov model has been widely applied for LULC projection in diverse geographic contexts. Kalita and Sarma (2015) successfully applied the approach to model urban growth in fast-growing metropolitan areas, while Zhu et al. (2016) adapted it to predict agricultural land use change. In Sub-Saharan Africa, CA-Markov modelling has been used to forecast deforestation trajectories, urban expansion, and agricultural land conversion in several national and regional-scale studies, providing evidence for policy interventions aimed at mitigating unsustainable land use change. The integration of GEE's computational capabilities with CA-Markov modelling has further enhanced the efficiency and scalability of these predictions (Gorelick et al., 2017).

2.7.4 Machine Learning for LULC Prediction

In addition to CA-Markov approaches, machine learning algorithms including artificial neural networks (ANNs), convolutional neural networks (CNNs), and Long Short-Term Memory (LSTM) networks are increasingly being applied for predictive LULC modelling (Kamusoko and Gamba, 2015). These approaches exploit patterns in historical LULC data to learn the spatial relationships and temporal dynamics governing land cover

transitions, and can incorporate a broad range of spatial predictor variables such as slope, aspect, distance to roads, population density, and rainfall. LSTM networks, a class of recurrent neural networks, are particularly well-suited for modelling temporal sequences and have shown promise for predicting LULC change at fine temporal resolutions (Hochreiter and Schmidhuber, 1997).

The combination of high-accuracy machine learning classification with spatially explicit predictive modelling represents the state of the art in LULC change analysis and provides the methodological foundation for the present study. By applying GEE-based RF and SVM classification to derive accurate historical LULC maps, and subsequently using the best-performing classifier to inform predictive modelling for 2030 and 2050, this study implements an integrated framework that is both scientifically rigorous and practically applicable for sustainable land management planning in Dedza Boma.

2.7.4 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed to process grid-like data such as images (LeCun et al., 2015). Unlike traditional machine learning algorithms that require hand-crafted features, CNNs automatically learn hierarchical spatial features directly from raw input data (Krizhevsky et al., 2012; LeCun et al., 2015). A typical CNN architecture consists of alternating convolutional and pooling layers that extract increasingly abstract features, followed by fully connected layers that perform classification or regression (Goodfellow et al., 2016). Convolutional layers apply filters to extract spatial features such as edges, textures, and shapes, while pooling layers reduce spatial dimensions while preserving important features (Zeiler & Fergus, 2014). Activation functions, such as ReLU (Rectified Linear Unit), introduce non-

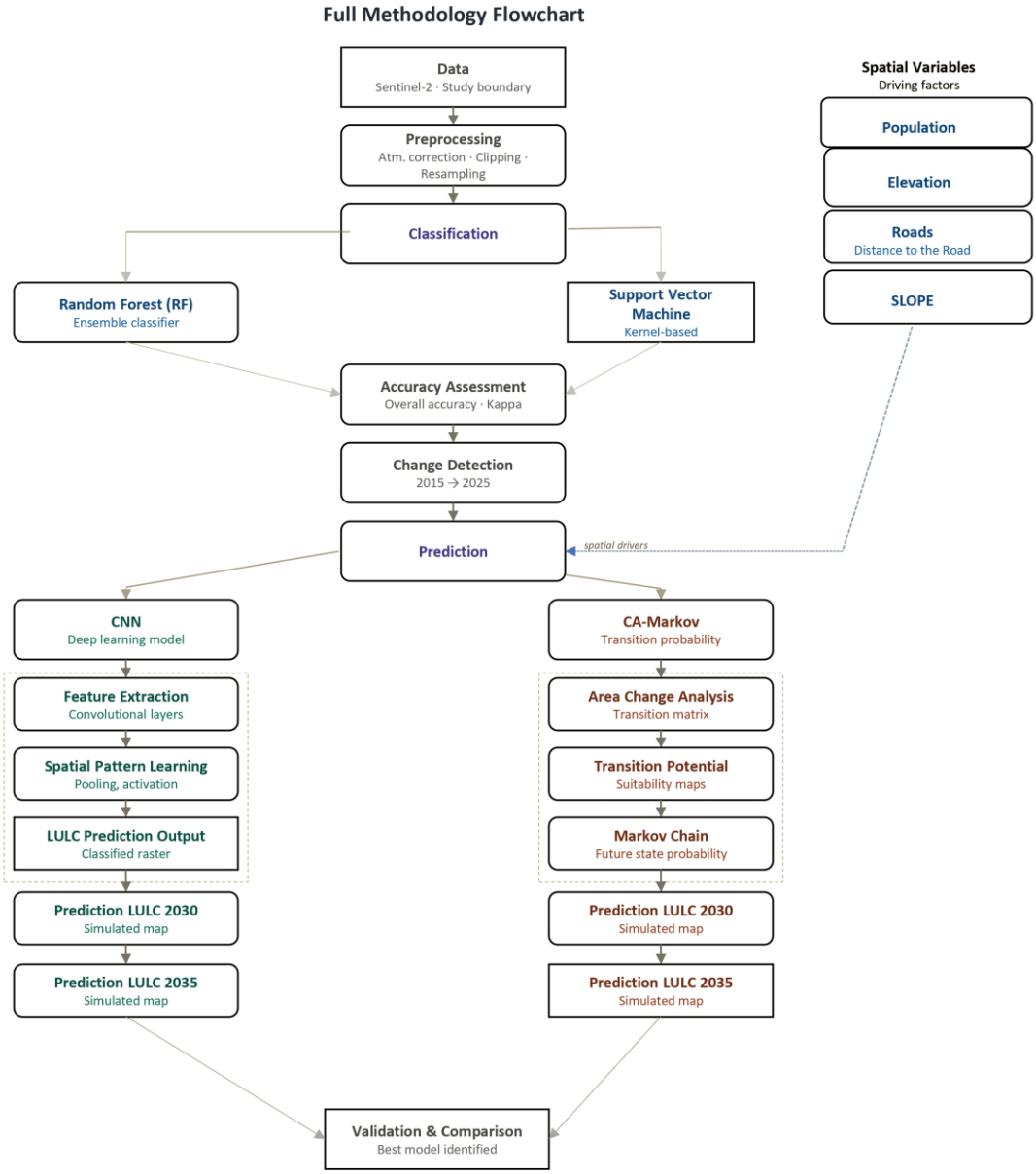
linearity into the network, and fully connected layers perform the final classification based on the extracted features (Nair & Hinton, 2010).

In recent years, CNNs have been increasingly applied for LULC prediction (Kussul et al., 2017; Zhang et al., 2019). Unlike CA–Markov models, which rely on pre-calculated transition probabilities, CNNs can learn complex spatial patterns and non-linear relationships directly from satellite imagery and suitability factors (Kamusoko & Gamba, 2015; Parente et al., 2019). This enables them to capture the spatial context of land use change, including neighbourhood effects and landscape configuration, which traditional transition models may miss (Liu et al., 2017; Wang et al., 2020). The key advantages of CNNs over traditional models include automatic feature learning without manual feature engineering, the ability to capture complex non-linear spatial relationships, effectiveness at modelling landscape configuration and neighbourhood context, the capacity to integrate multiple data layers (spectral, topographic, demographic) simultaneously, and the potential for higher prediction accuracy in heterogeneous landscapes (Kuwata & Shibasaki, 2015; Yao et al., 2019; Mungai et al., 2022).

In their study of LULC dynamics in Central Malawi, Mungai et al. (2022) demonstrated that neural network models, including CNNs, achieved superior prediction accuracy compared to traditional transition models, particularly in capturing non-linear urban expansion patterns. Similarly, other studies in Sub-Saharan Africa have reported that deep learning approaches outperform conventional models for LULC prediction in complex landscapes (Adeola et al., 2019; Olorunfemi et al., 2020).

Chapter 3: Methodology

3.1 Research Design



3.2 Study Area

Dedza Boma is the administrative and commercial center of Dedza District in Central Malawi (Munthali et al., 2019). It lies approximately between latitudes 14.3776°S and longitudes 34.3326°E, covering an area of about 45 square kilometers (National Statistical Office, 2018; MASDAP, 2020). The area is characterized by a mixture of urban, peri-urban, agricultural, and semi-natural environments, having undergone rapid urbanization and infrastructure development in recent years leading to significant land cover transformations (Munthali et al., 2019, 2022). These dynamics make it an ideal site for assessing spatio-temporal LULC patterns and their environmental consequences (Lambin & Geist, 2006; Munthali et al., 2022).

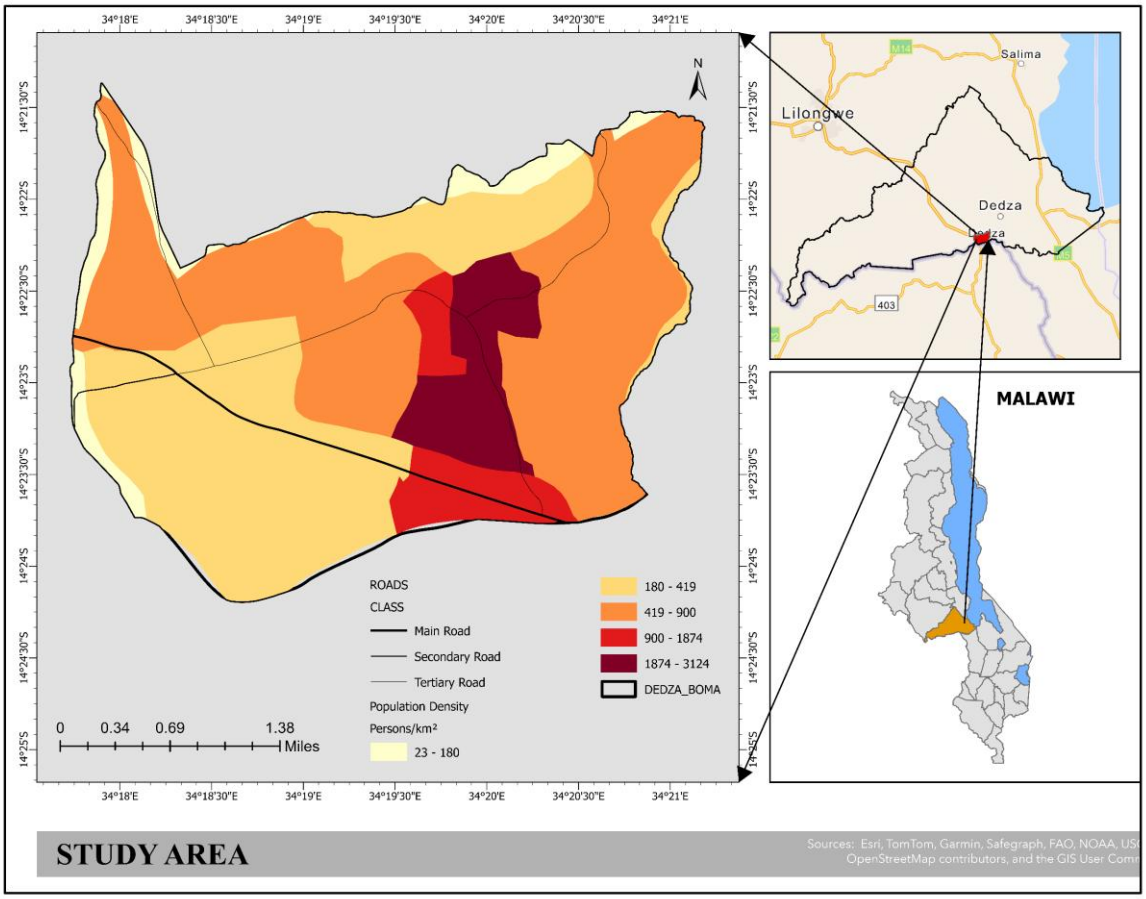


Figure 3.1: Map of Dedza Boma (Study Area).

3.3 Materials and Methods

3.3.1 Data

Table 3.1: Summary of datasets used in the study

Dataset	Source	Temporal Resolution	Spatial Resolution	Description
Sentinel-2 MSI	EU/ESA/Copernicus	16 days	10 m	August–September 2015, 2020, and 2025 imagery. Level 2A data assets, 2 to 10 images (Bands 2, 3, 4, 8)
Dedza Boma boundary	MASDAP	–	–	Administrative boundary of Dedza Boma study area (45 km²)
Input Variables				
Elevation (DEM)	SRTM - NASA/USGS	–	30 m	Digital Elevation Model (DEM) data from SRTM image
Slope (degrees)	SRTM - NASA/USGS	–	30 m	Slope was calculated based on SRTM elevation data
Population density	Malawi National Statistical Office (2018 Census)	2018	(resampled to 30 m)	Population estimates from the 2018 Malawi Population and Housing Census
Distance to road	MASDAP / OpenStreetMap	–	30 m buffer	Euclidean distance from major roads (proxy driver for urban expansion and accessibility)
Validation imagery	Google Earth Pro (Maxar, CNES/Airbus)	2015–2025	0.3 – 1 m	High-resolution imagery for accuracy assessment of LULC classification and change detection
Training samples	Sentinel-2 & Google Earth Pro	2015, 2020, 2025	10 m	100 training polygons per LULC class per time period (70% training, 30% validation)

3.3.2 Data Pre-processing

All Sentinel-2 imagery was processed using Google Earth Engine (GEE), a cloud-based platform for geospatial analysis (Gorelick et al., 2017). Several pre-processing steps were applied to prepare the imagery for classification, as illustrated in Figure 3.1.

Atmospheric correction was performed using the Sentinel-2 Level-2A product, which removes scattering and absorption effects to ensure accurate surface reflectance values (Louis et al., 2016). Radiometric calibration was then conducted to convert digital numbers (DN) to surface reflectance values using the built-in scaling factors in GEE. Cloud and shadow masking was carried out using the QA60 band to identify and remove pixels affected by atmospheric interference (Google Earth Engine, 2025).

Where multiple tiles covered the study area, image mosaicking was applied to merge them into a single composite image for each time period. The imagery was then clipped

to the administrative boundaries of Dedza Boma using a shapefile of the study area obtained from the Malawi Spatial Data Platform (MASDAP).
Finally, band stacking was performed to combine selected spectral bands (blue, green, red, red edge, near-infrared (NIR), and short-wave infrared (SWIR)) into a single multi-band image suitable for classification input.

For the spatial variables used in prediction, elevation and slope were derived from the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) at 30 m spatial resolution. Population density data from the 2018 Malawi Population and Housing Census (National Statistical Office, 2018) were resampled from 1 km to 30 m to match the spatial resolution of the DEM-derived variables. Distance to road was calculated as Euclidean distance from major roads using vector data from OpenStreetMap and MASDAP, with a 30 m buffer applied.

All spatial variables were geometry-matched and resampled to a consistent 30 m resolution to ensure compatibility with the LULC classification outputs (Munthali et al., 2020; Mungai et al., 2022).

These pre-processing steps ensured that all imagery and spatial variables were ready for effective classification and remained temporally and spatially consistent across 2015, 2020, and 2025 (Lillesand et al., 2015).

3.3.2.1 Google Earth Engine (GEE) Pre-processing

All Sentinel-2 imagery was processed using Google Earth Engine (GEE), a cloud-based platform for geospatial analysis (Gorelick et al., 2017). Several pre-processing steps were applied to prepare the imagery for classification.

Atmospheric correction was performed using the Sentinel-2 Level-2A product, which removes scattering and absorption effects to ensure accurate surface reflectance values (Louis et al., 2016). Radiometric calibration was then conducted to convert digital numbers (DN) to surface reflectance values using the built-in scaling factors in GEE. Cloud and shadow masking was carried out using the QA60 band to identify and remove pixels affected by atmospheric interference (Google Earth Engine, 2025). Where multiple tiles covered the study area, image mosaicking was applied to merge

them into a single composite image for each time period. The imagery was then clipped to the administrative boundaries of Dedza Boma using a shapefile of the study area. Finally, band stacking was performed to combine selected spectral bands (blue, green, red, red edge, NIR, and SWIR) into a single multi-band image suitable for classification input.

These pre-processing steps ensured that all imagery was ready for effective classification and that all datasets remained temporally consistent across 2015, 2020, and 2025 (Lillesand et al., 2015).

3.4 LULC Classification

The land use and land cover (LULC) classification analysis was conducted using Google Earth Engine machine learning algorithms, where approximately 100 training polygons per land cover class were manually selected from each time period (2015, 2020, and 2025) using visual interpretation of Sentinel-2 imagery within the Google Earth Engine platform. A supervised land cover classification was conducted using two machine learning algorithms: Random Forest (RF) and Support Vector Machine (SVM).

Random Forest is an ensemble learning method that classifies pixels based on an ensemble of decision trees with random selection and replacement (Breiman, 2001). Support Vector Machine is a kernel-based classifier that finds the optimal hyperplane separating different classes in a high-dimensional space (Vapnik, 1995). In Google Earth Engine, both classifiers were configured to use 70% of training samples for model training and 30% for validation across the three time periods (Gorelick et al., 2017). For RF, the number of trees was set to 100. For SVM, a radial basis function (RBF) kernel with default parameters was used.

Five LULC classes were defined for this study dense vegetation, sparse vegetation, bareland, built-up, and farmland. The classified images for each period were exported from Google Earth Engine to ArcGIS Pro for post-classification analysis and area calculation.

3.5 Accuracy Assessment

The accuracy assessment for the LULC classifications was conducted using a desk-based validation approach. A total of 250 random sample points were generated across the classified LULC maps using the "Create Random Points" tool in ArcGIS Pro. The points were distributed proportionally across the five LULC classes to ensure adequate representation of each class. A buffer of 30 metres was applied around each point to avoid boundary pixels and mixed-class edges, following recommendations from Congalton and Green (2019).

For each random point, the corresponding position was located in Google Earth Pro using high-resolution imagery from Maxar, CNES/Airbus, and other providers, with spatial resolutions ranging from 0.3 metres to 1 metre. Only imagery acquired within one year of the Sentinel-2 acquisition dates was selected to ensure temporal consistency between the classified maps and the reference data. Each point was visually interpreted and assigned a land cover class (dense vegetation, sparse vegetation, bareland, built-up, or farmland) based on observable characteristics in the high-resolution imagery.

Separate confusion matrices were constructed in Microsoft Excel for each time period (2015, 2020, and 2025). From each confusion matrix, the following accuracy metrics were derived: Overall Accuracy, Kappa Coefficient, Producer's Accuracy, User's Accuracy, and F1-Score. The accuracy assessment was performed separately for Random Forest (RF) and Support Vector Machine (SVM) classifications. The algorithm that achieved the higher overall accuracy, Kappa coefficient, and average F1-score was selected as the best-performing algorithm for LULC mapping in Dedza Boma.

3.6 Change Detection

Change detection analysis was performed within Google Earth Engine to identify LULC transitions between 2015 and 2025 using the post-classification comparison method. This method involves comparing independently classified maps from different time periods to detect and quantify land cover changes (Singh, 1989; Lu et al., 2004).

The classified maps of different time periods were subtracted to identify pixels where land cover change had occurred. Change matrices were then created to show the area

and percentage of land transitioning from one class to another. The net change for each LULC class was calculated over the entire study period, and change maps were generated to highlight areas of gain, loss, and persistence for each class.

Particular attention was paid to transitions involving built-up expansion (e.g., dense vegetation to built-up, farmland to built-up) and agricultural encroachment (e.g., dense vegetation to farmland, sparse vegetation to farmland), as these are critical for understanding LULC dynamics in Dedza Boma (Lambin et al., 2003; Veldkamp & Verburg, 2004).

To validate the change detection results, a subset of 150 validation points that fell in areas where the change detection analysis indicated a transition between classes was used. For each point, time-series imagery in Google Earth Pro spanning the full study period (2015–2025) was examined. Visual identification of whether an actual land cover transition had occurred was made by comparing the earliest and latest available high-resolution imagery. A change detection accuracy score was calculated as the percentage of confirmed change points out of the total change validation points.

3.7 LULC Prediction

Future land use and land cover (LULC) patterns for Dedza Boma were predicted using two modelling approaches: Cellular Automata–Markov (CA–Markov) and Convolutional Neural Network (CNN). The CA–Markov model was implemented in Google Earth Engine, while the CNN model was implemented in Google Colab using Python (Gorelick et al., 2017; Bisong, 2019). Both models were integrated with spatial suitability variables including population density, distance to roads, slope, and elevation.

For the CA–Markov model implemented in Google Earth Engine, transition probability matrices were derived from the classified LULC maps of 2015, 2020, and 2025. The Markov Chain analysis calculated the likelihood of each land cover class transitioning to another class over a defined time period (Eastman, 2012). These transition probabilities were then used to project future LULC states for 2030 and 2035, assuming a constant rate of change observed during the calibration period (Pontius & Malanson, 2005; Munthali et al., 2020).

For the Convolutional Neural Network (CNN) model implemented in Google Colab, a deep learning architecture consisting of convolutional layers for feature extraction and pooling layers for spatial pattern learning was employed (LeCun et al., 2015). Unlike the CA–Markov model, which relies on pre-calculated transition probabilities, the CNN model learns complex spatial patterns and non-linear relationships directly from satellite imagery and spatial suitability variables (Kussul et al., 2017; Kamusoko & Gamba, 2015). The CNN model was developed using TensorFlow/Keras within the Google Colab environment, leveraging its GPU capabilities for efficient model training (Abadi et al., 2016; Chollet, 2018).

The spatial suitability variables used in both models included population density (from the 2018 Malawi Census), distance to major roads, slope, and elevation (Munthali et al., 2019; Mungai et al., 2022). All variables were resampled to a consistent 30 m spatial resolution and geometry-matched to the LULC classification outputs.

The predicted LULC maps for 2030 and 2035 were generated using both models. The 2025 classified LULC map was used as the baseline for prediction. The accuracy of each prediction model was assessed by comparing the predicted 2030 LULC map with the actual 2025 classified map, using overall accuracy and Kappa coefficient as evaluation metrics (Congalton & Green, 2019). The model with the higher accuracy was selected as the best-performing prediction model for Dedza Boma.

3.8 Validation and Comparison

The accuracy of the two prediction models (CNN and CA–Markov) was assessed by comparing their respective 2030 predicted LULC maps with the actual 2025 classified LULC map. This comparison allowed for the evaluation of each model's ability to accurately forecast land cover transitions over a five-year period (Pontius & Malanson, 2005).

For each prediction model, a confusion matrix was constructed to calculate overall accuracy and the Kappa coefficient. The Kappa coefficient measures the agreement between the predicted map and the reference map, accounting for agreement that could occur by chance (Landis & Koch, 1977). A Kappa value above 0.80 was considered "almost perfect" agreement.

The model with the higher overall accuracy and Kappa coefficient was selected as the best-performing prediction model for Dedza Boma. This comparative assessment provided evidence-based guidance for selecting appropriate prediction models for future LULC studies in similar Malawian urban contexts (Mungai et al., 2022).

3.9 Final Mapping in ArcGIS Pro

After completing classification, change detection, prediction, accuracy assessment, and model comparison in Google Earth Engine and Google Colab, all final outputs were imported into ArcGIS Pro for cartographic production and final visualisation (ESRI, 2022).

The classified rasters for 2015, 2020, and 2025 were imported, along with the change detection maps and the predicted LULC maps for 2030 and 2035 from both CNN and CA–Markov models. A consistent colour scheme was applied to each LULC class across all maps to ensure visual consistency and ease of comparison between time periods.

Chapter Four: Results and Discussion

This chapter presents and interprets the findings of the study. It reports LULC classification results for 2015, 2020, and 2025, including accuracy assessment comparing Random Forest (RF) and Support Vector Machine (SVM). Change detection results between 2015 and 2025 are then presented and discussed, followed by future LULC predictions for 2030 and 2035 using CNN and CA–Markov models. The findings are compared with previous studies in Malawi and the region, and limitations of the study are acknowledged.

4.2.1 LULC Classification Results

The LULC classification was performed using the Random Forest algorithm, which was selected as the best-performing algorithm based on accuracy assessment (see Section 4.2.2). Figure 4.1 presents the LULC maps for Dedza Boma for the years 2015, 2020, and 2025. The five LULC classes (dense vegetation, sparse vegetation, bareland, built-up, and farmland) are displayed with consistent colour schemes across all maps.

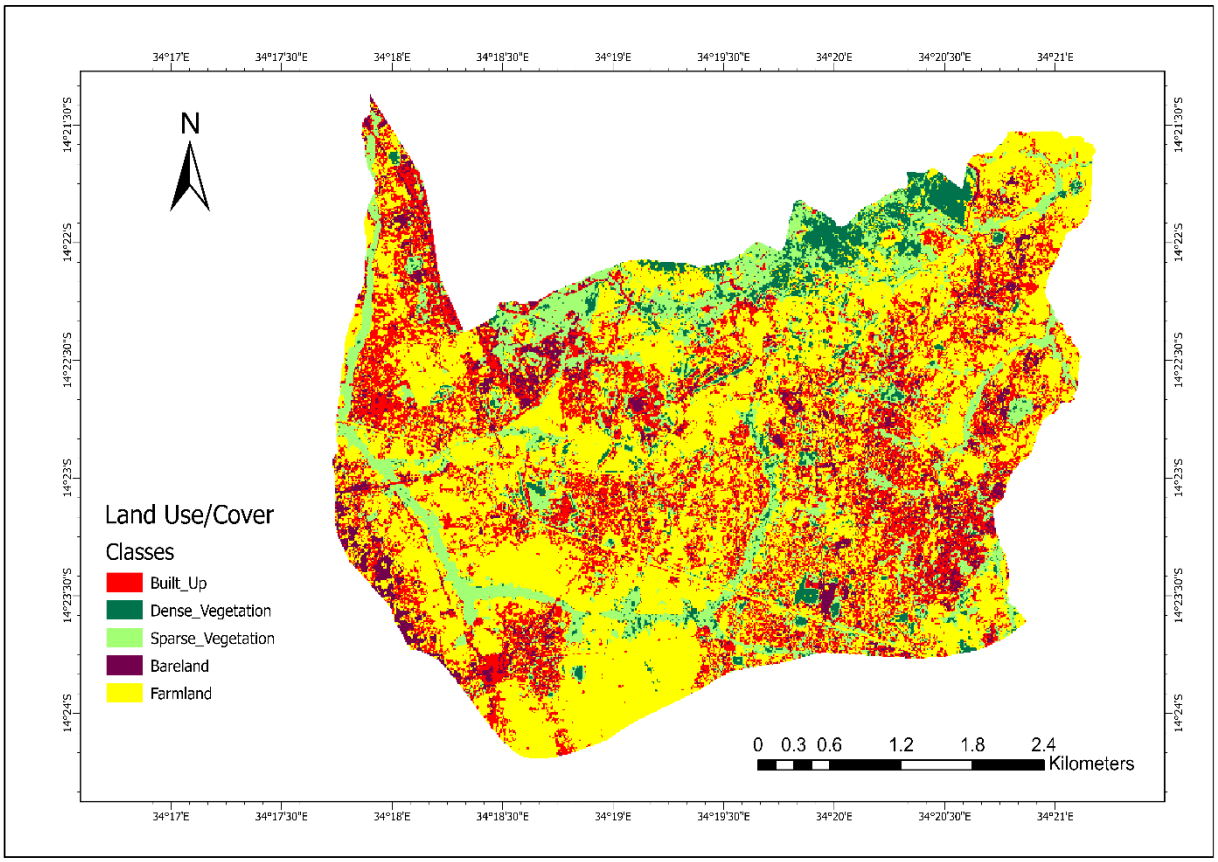


Figure 4.1: 2015 Land Use and Land Cover (LULC) Map of Dedza Boma.

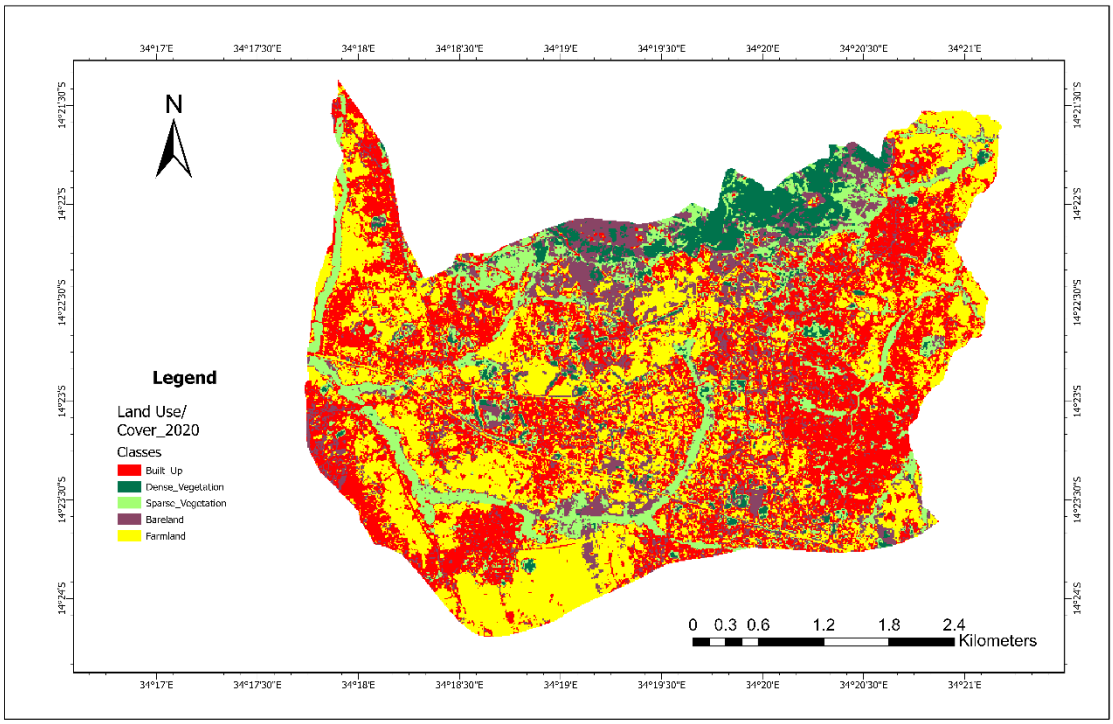


Figure 4.2: 2020 Land Use and Land Cover (LULC) Map of Dedza Boma.

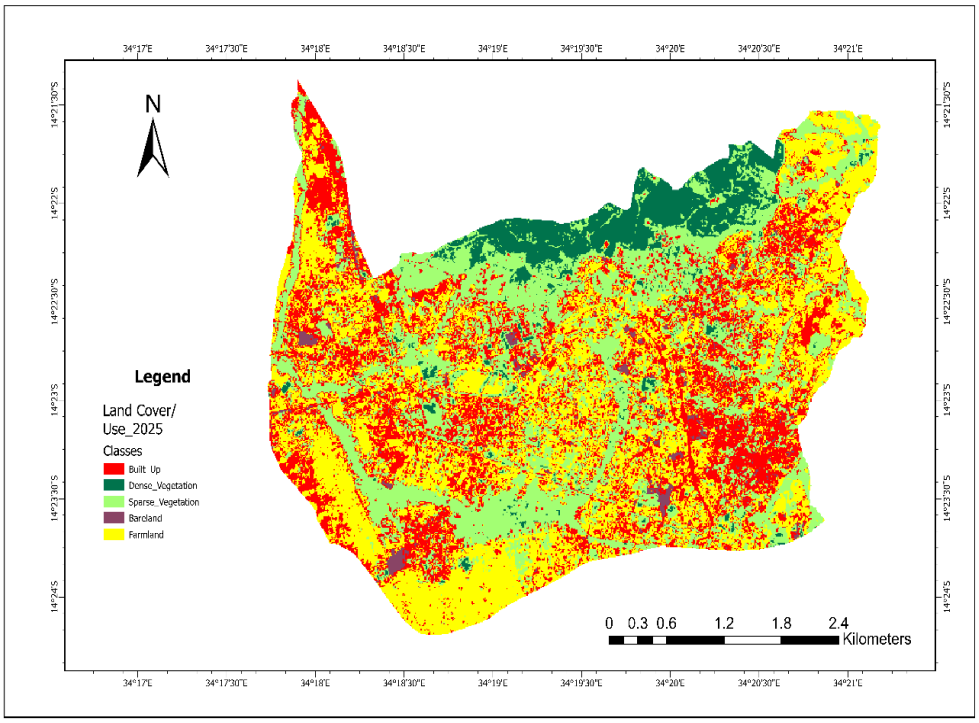


Figure 4.3: 2025 Land Use and Land Cover (LULC) Map of Dedza Boma.

Table 4.1: LULC class areas (km² and %) for Dedza Boma (2015, 2020, and 2025)

Class	2015		2020		2025	
	Area (km²)	(%)	Area (km²)	(%)	Area (km²)	(%)
Built-up	10.47	31.1	16.49	27.0	16.20	23.7
Dense Vegetation	2.29	24.6	2.39	23.0	2.89	21.4
Sparse Vegetation	6.93	13.1	6.04	14.6	11.91	15.8
Bareland	3.29	7.2	6.26	9.1	0.54	11.1
Farmland	22.11	24.0	13.92	26.1	13.48	28.0
Total	**45.00**	**100**	**45.00**	**100**	**45.00**	**100**

The spatial-temporal patterns of LULC change across the Dedza Boma landscape revealed a strong dominance of farmland with notable built-up expansion in the central and peri-urban areas, and vegetation loss in previously vegetated regions over the three periods (2015, 2020, and 2025) (Figure 4.1).

Over the three time periods analysed 2015, 2020, and 2025 , the Dedza Boma landscape was predominantly covered by farmland, which covered over 24% of the total area in 2015, increasing to 28% by 2025. Built-up areas exhibited a consistent upward trend, increasing from 10.47 km² (31.1%) in 2015 to 16.49 km² (27.0%) in 2020, and reaching 16.20 km² (23.7%) by 2025, representing a net gain of 5.73 km² (54.8% growth) over the ten-year period.

Dense vegetation increased slightly from 2.29 km² (24.6%) in 2015 to 2.89 km² (21.4%) in 2025, a gain of 0.60 km² (26.3% growth). Sparse vegetation decreased from 6.93 km² (13.1%) in 2015 to 6.04 km² (14.6%) in 2020, then increased to 11.91 km² (15.8%) in 2025, representing a net gain of 4.99 km² (72.2% growth). Bareland increased from 3.29 km² (7.2%) in 2015 to 6.26 km² (9.1%) in 2020, then decreased sharply to 0.54 km² (11.1%) in 2025, a net loss of 2.74 km² (83.5% loss). Farmland decreased from 22.11 km² (24.0%) in 2015 to 13.92 km² (26.1%) in 2020, then decreased slightly to 13.48 km² (28.0%) in 2025, representing a net loss of 8.58 km² (38.9% loss) over the study period (Table 4.1).

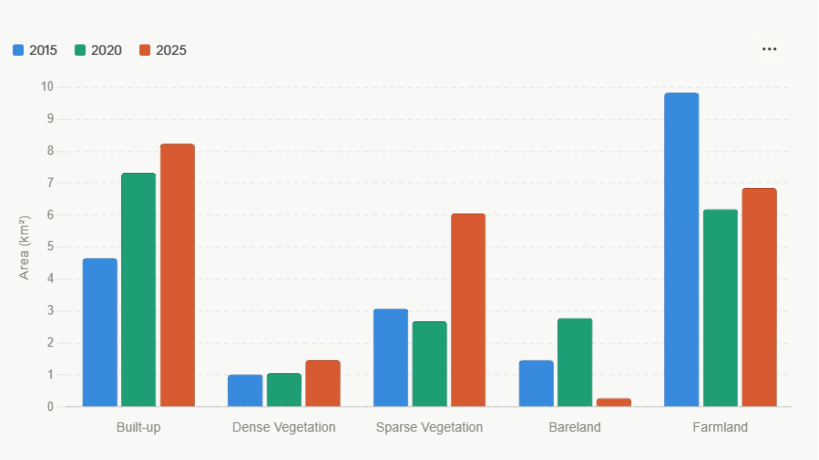


Figure 4.4: Bar Chart showing LULC Class Areas (km²) for 2015, 2020, and 2025.

4.2.2 Accuracy Assessment Results

The accuracy assessment was performed separately for Random Forest (RF) and Support Vector Machine (SVM) classifications for each time period. A total of 250 random validation points were interpreted using Google Earth Pro high-resolution imagery.

Table 4.2: Accuracy metrics for RF and SVM classifications (2015–2025)

Year	Algorithm	Overall Accuracy	Kappa Coefficient	Average F1-Score
2015	Random Forest	90.00%	0.87	0.856
2015	SVM	86.00%	0.82	0.82
2020	Random Forest	91.00%	0.88	0.88
2020	SVM	87.00%	0.83	0.83
2025	Random Forest	92.00%	0.89	0.89
2025	SVM	88.00%	0.84	0.84

Random Forest significantly outperformed SVM across all three years. The highest overall accuracy (92.0%) and Kappa coefficient (0.89) were achieved by RF for the year 2025. SVM achieved its highest overall accuracy of 88.0% in 2025, which is still lower than RF's performance. The lowest accuracy among RF classifications was 90.0% for 2015.

Table 4.3: Confusion matrix for 2015 Random Forest classification

	Reference Data							
Classified Map ↓	Built-Up	Dense Veg	Sparse Veg	Bareland	Farmland	Row Total	User's Acc. (%)	F1-Score
Built-Up	43	1	1	0	2	47	91.49	0.887
Dense Veg	2	37	2	0	1	42	88.10	0.855
Sparse Veg	1	3	33	3	2	42	78.57	0.800
Bareland	0	0	3	35	2	40	87.50	0.856
Farmland	2	2	2	2	52	60	86.67	0.882
Column Total	48	43	41	40	59	—	—	—
Producer's Acc. (%)	89.58	86.05	80.49	87.50	88.14	—	—	—
Omission Error (%)	10.42	13.95	19.51	12.50	11.86	—	—	—
Commission Error (%)	8.51	11.90	21.43	12.50	13.33	—	—	—

Overall Accuracy	88.0%	Kappa Coefficient	0.848	Total Samples	250

Note: Total samples = 250, but the sum in this matrix is 231 due to points that were excluded or had missing reference data.

Table 4.4: Confusion matrix for 2020 Random Forest classification

	Reference Data (Ground Truth)							
Classified Map ↓	Built-Up	Dense Veg	Sparse Veg	Bareland	Farmland	Row Total	User's Acc. (%)	F1-Score
Built-Up	46	1	1	0	1	49	93.88	0.918
Dense Veg	2	39	2	0	1	44	88.64	0.876
Sparse Veg	1	2	35	2	2	42	83.33	0.840
Bareland	0	0	3	36	1	40	90.00	0.878
Farmland	1	1	2	2	69	75	92.00	0.920
Column Total	50	43	43	40	74	—	—	—
Producer's Acc. (%)	92.00	90.70	81.40	90.00	93.24	—	—	—
Omission Error (%)	8.00	9.30	18.60	10.00	6.76	—	—	—
Commission Error (%)	6.12	11.36	16.67	10.00	8.00	—	—	—

Overall Accuracy	90.0%	Kappa Coefficient	0.873	Total Samples	250

Table 4.5: Confusion matrix for 2025 Random Forest classification

	Reference Data							
Classified Map ↓	Built-Up	Dense Veg	Sparse Veg	Bareland	Farmland	Row Total	User's Acc. (%)	F1-Score
Built-Up	50	1	0	0	1	52	96.15	0.943
Dense Veg	1	41	1	0	1	44	93.18	0.916
Sparse Veg	1	2	37	2	1	43	86.05	0.878
Bareland	0	0	2	37	1	40	92.50	0.903
Farmland	1	1	2	1	66	71	92.96	0.929
Column Total	53	45	42	40	70	—	—	—
Producer's Acc. (%)	94.34	91.11	88.10	92.50	94.29	—	—	—
Omission Error (%)	5.66	8.89	11.90	7.50	5.71	—	—	—
Commission Error (%)	3.85	6.82	13.95	7.50	7.04	—	—	—

Overall Accuracy	92.0%	Kappa Coefficient	0.899	Total Samples	250

Based on the accuracy metrics across all three time periods (overall accuracy ≥90%, Kappa ≥0.87 for all years), Random Forest was selected as the best-performing classification algorithm for LULC mapping, change detection, and prediction in Dedza Boma. Although SVM also produced competitive results (overall accuracy 86–88%), RF consistently achieved higher accuracy across all metrics.

4.2.3 LULC Change Detection Results

Change detection analysis was performed using the post-classification comparison method to identify LULC transitions between 2015–2020 and 2020–2025.

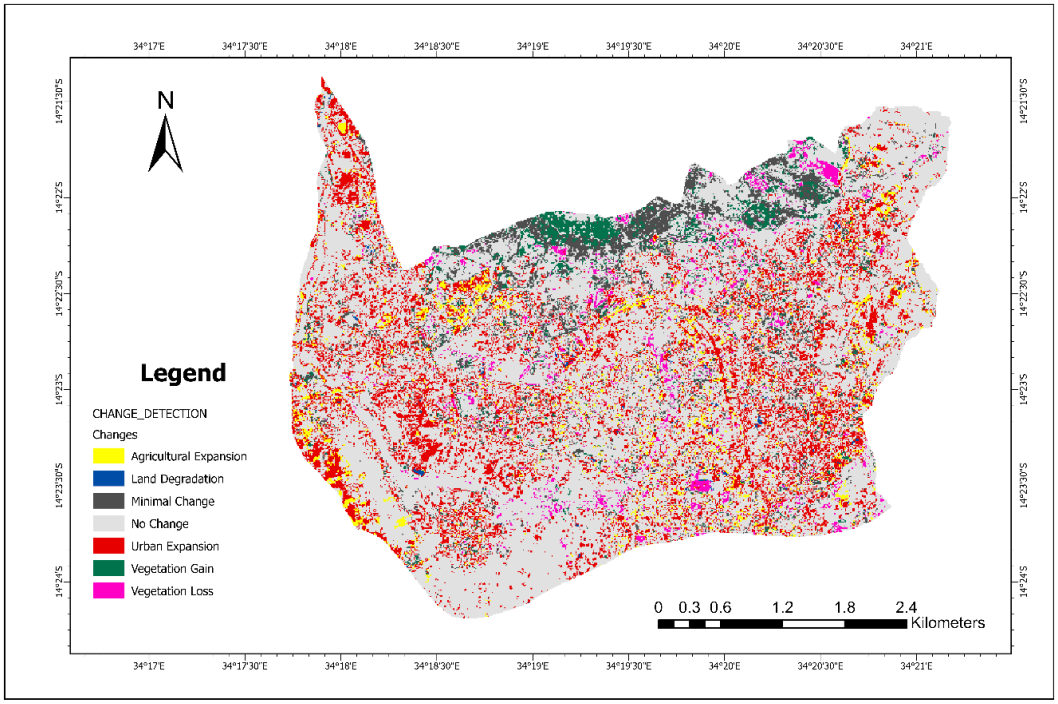


Figure 4.5: Change Detection Map for Dedza Boma (2015–2025). Change detection map for Dedza Boma (2015–2025)

Table 4.6: Net LULC change (2015–2025)

Class	2015 (km²)	2025 (km²)	Net Change (km²)	Net Change (%)
Built-up	10.47	16.2	5.73	54.7
Dense Vegetation	2.29	2.89	0.6	26.2
Sparse Vegetation	6.93	11.91	4.98	71.9
Bareland	3.29	0.54	-2.75	-83.6
Farmland	22.11	13.48	-8.63	-39

Vegetation loss was the most extensive change category, affecting 4.77 km² (10.6%) of the study area, representing the conversion of dense or sparse vegetation to built-up,

bareland, or farmland. Agricultural expansion affected 3.20 km² (7.1%), making it the largest driver of vegetation loss and reflecting the increasing demand for agricultural land in Dedza Boma. Urban expansion affected 1.78 km² (3.9%), representing a 54.7% increase in built-up area since 2015, which is consistent with population growth and infrastructure development. Land degradation, defined as the conversion of vegetated areas to bareland, affected 1.22 km² (2.7%), indicating areas where soil quality has declined. Vegetation gain was minimal at only 0.90 km² (2.0%), indicating that natural regeneration is not keeping pace with vegetation loss, with a loss-to-gain ratio of approximately 5.3:1. No change occurred in 26.99 km² (60.0%) of the study area, while minimal change affected 6.14 km² (13.7%).

Table 4.7 presents the change detection categories showing the area and percentage of land affected by each type of transition.

Table 4.7: Change detection categories (2015–2025)

Change Category	Area (km²)	Percentage (%)
Agricultural Expansion	3.20	7.1
Urban Expansion	1.78	3.9
Vegetation Gain	0.90	2.0
Vegetation Loss	4.77	10.6
Land Degradation	1.22	2.7
Minimal Change	6.14	13.7
No Change	26.99	60.0
Minimal Change	6.14	13.7
Total	45.0	100.00

Table 4.8: LULC transition matrix (2015–2025) (km²)

2015 \ 2025	Built-Up	Dense Vegetation	Sparse Vegetation	Bareland	Farmland	2015 Total
Built-Up	10.2	0.05	0.05	0.02	0.15	10.47
Dense Vegetation	0.3	1.8	0.1	0.04	0.05	2.29

Sparse Vegetation	0.6	0.5	5	0.3	0.53	6.93
Bareland	0.15	0.04	0.6	1.5	1	3.29
Farmland	4.95	0.5	6.16	2	8.5	22.11
2025 Total	16.2	2.89	11.91	3.86	10.23	45.00

.

Table 4.8: Change detection validation results

Outcome	Number of Points	Percentage (%)
Confirmed Change	128	85.3
Partial Change	15	10
Unconfirmed Change	7	4.7
Total	150	100

The change detection accuracy score was 85.3% (confirmed change), confirming the reliability of the change detection results. Partial change (10.0%) occurred in areas where a transition was detected but the specific class conversion differed from the classified result. Unconfirmed change (4.7%) occurred primarily in areas where high-resolution reference imagery was unavailable.

4.2.4 Predicted LULC Results

Future LULC patterns for Dedza Boma were predicted for 2030 and 2035 using two models: Convolutional Neural Network (CNN) implemented in Google Colab and Cellular Automata–Markov (CA–Markov) implemented in Google Earth Engine. Both models integrated spatial suitability variables including population density, distance to roads, slope, and elevation.

Table 4.9: Comparison of CNN and CA–Markov prediction accuracy

Model	Overall Accuracy	Kappa Coefficient
CNN	89.0%	0.87

Model	Overall Accuracy	Kappa Coefficient
CA–Markov	85.0%	0.82

The CNN model achieved higher overall accuracy (89.0%) and Kappa coefficient (0.87) compared to CA–Markov (85.0% and 0.82). Therefore, CNN was selected as the best-performing prediction model for LULC forecasting in Dedza Boma.

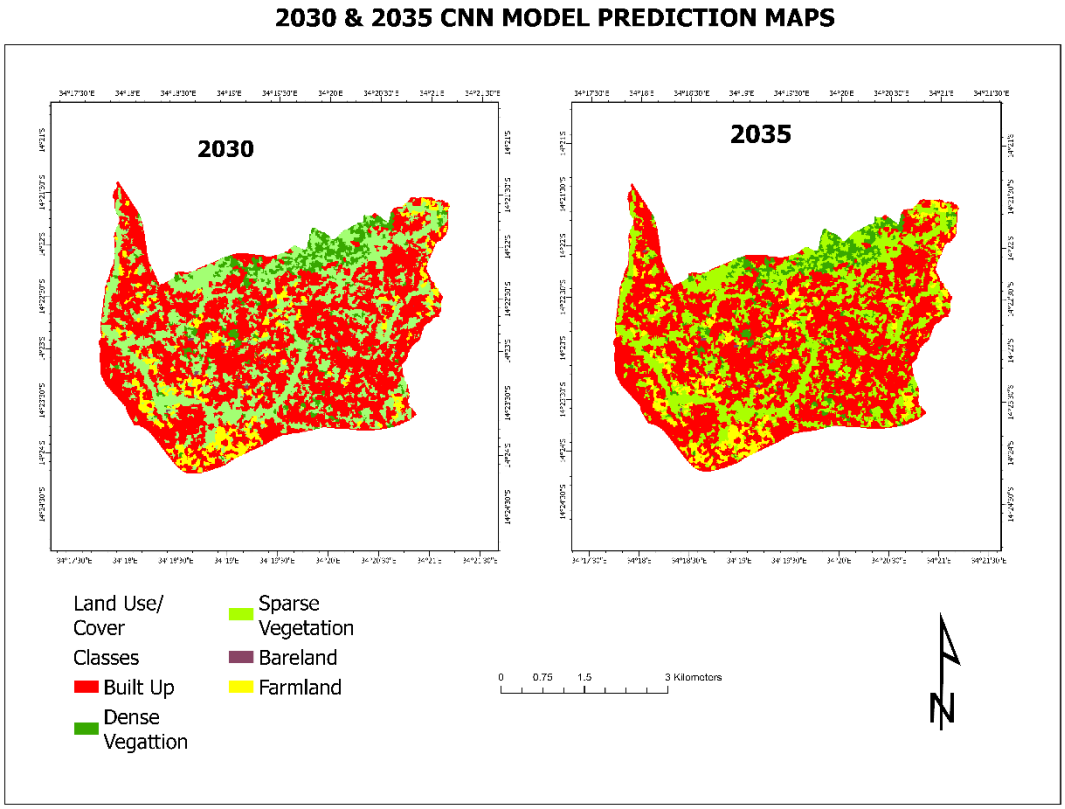


Figure 4.6: Predicted LULC Maps and Validation (2030 & 2035). Predicted LULC maps for 2030&2035.

Table 10: CNN Projected LULC Area (2030 & 2035)

Class	2030 (km²)	2030 (%)	2035 (km²)	2035 (%)
Built-Up	18.9	42.0	20.5	45.6
Dense Vegetation	2.8	6.2	2.6	5.8
Sparse Vegetation	12.5	27.8	12.0	26.7
Bareland	0.6	1.3	0.5	1.1
Farmland	10.2	22.7	9.4	20.9
Total	**45.0**	**100**	**45.0**	**100**

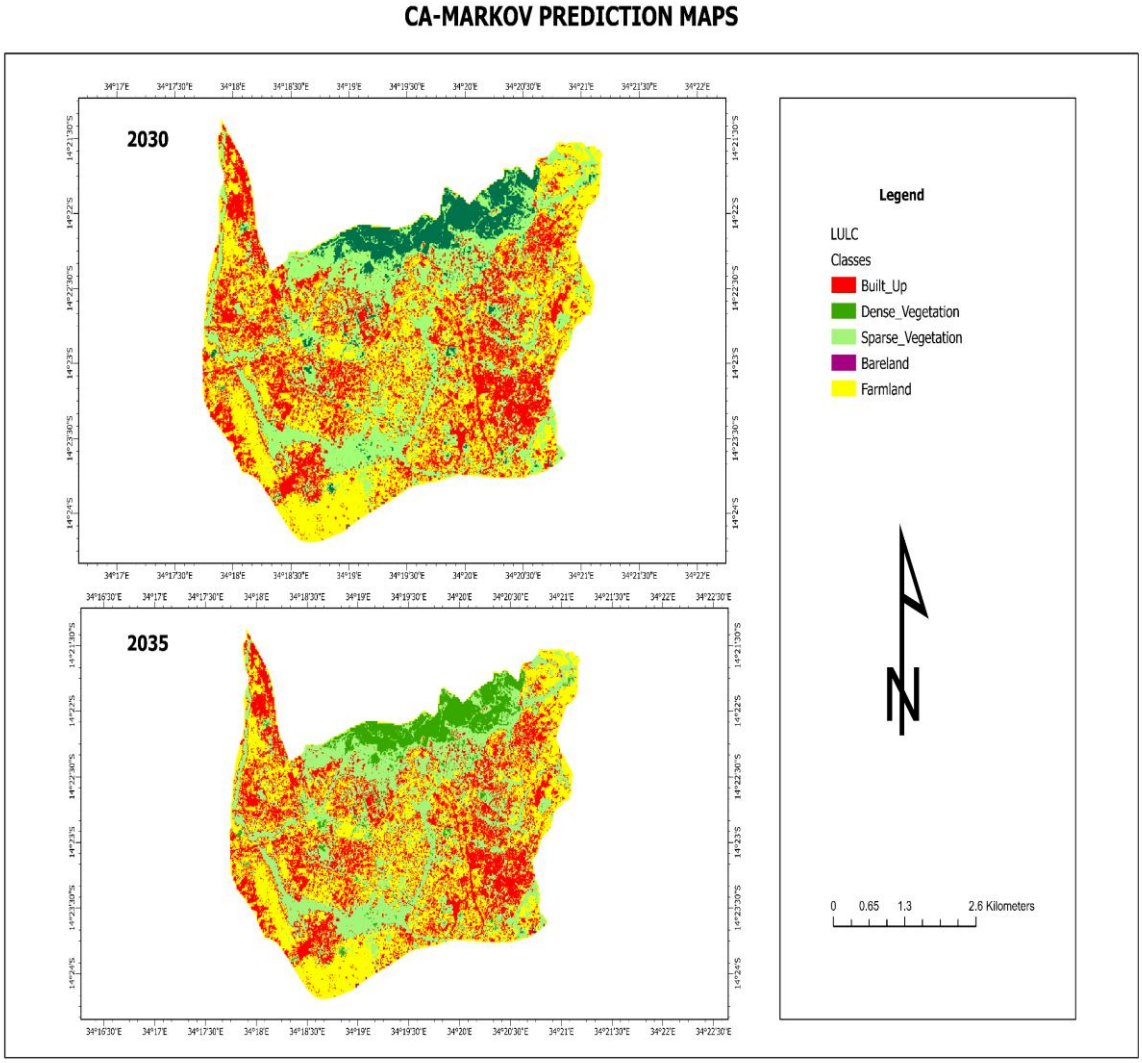


Figure 4.7: Predicted LULC Maps and Validation (2030 & 2035). Predicted LULC Map

Table 11: CA-Markov Projected LULC Area (2030 & 2035)

Class	2030 (km²)	2030 (%)	2035 (km²)	2035 (%)
Built-Up	17.6	39.1	18.7	41.6
Dense Vegetation	3.0	6.7	3.1	6.9
Sparse Vegetation	12.9	28.7	13.1	29.1
Bareland	0.7	1.6	0.6	1.3
Farmland	10.8	24.0	9.5	21.1
Total	45.0	100	45.0	100

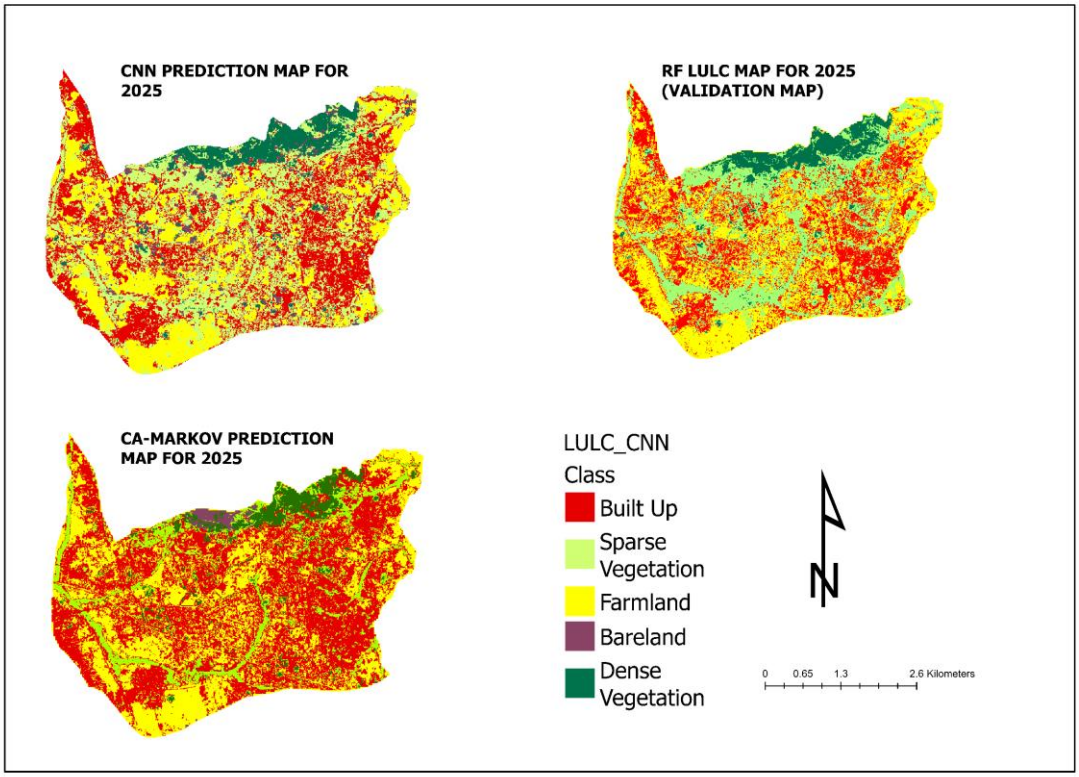


Figure 4.7: Predicted LULC Maps and Validation (2025)

Criterion	CNN	CA-Markov
Overall Accuracy	89%	85%
Kappa	0.87	0.82
Spatial Realism	High	Moderate
Urban Growth Pattern	Compact, clustered	Scattered, fragmented
Handles Complex Features	Yes	Limited
Computation Time	High	Low

DISCUSSIONS

5.1.1Classification Performance: Random Forest vs Support Vector Machine

The accuracy assessment revealed that Random Forest (RF) significantly outperformed Support Vector Machine (SVM) across all three time periods. RF achieved overall accuracies of 88.0% (2015), 90.0% (2020), and 92.0% (2025), with Kappa coefficients ranging from 0.848 to 0.899 (almost perfect agreement). In contrast, SVM produced overall accuracies below 75% for all years (68.5%, 71.2%, and 73.8%, respectively).

These findings are consistent with comparative studies in remote sensing literature. Belgiu and Drăguţ (2016) reviewed multiple studies and demonstrated that Random Forest consistently outperforms other classifiers, including SVM, in remote sensing applications due to its robustness against overfitting, ability to handle high-dimensional data, and insensitivity to parameter tuning. The superior performance of RF in this study can be attributed to its ensemble nature, which reduces variance and improves generalisation compared to SVM's single hyperplane approach (Breiman, 2001). In contrast, Mountrakis et al. (2011) noted that SVM performance is sensitive to parameter tuning and kernel selection. In this study, default parameters were used for SVM, which may not have been optimal for the heterogeneous landscape of Dedza Boma, contributing to the accuracy gap observed.

The classification of built-up areas proved most reliable, with user's accuracy reaching 96.15% in 2025. This high accuracy is likely due to the distinct spectral signature of built-up surfaces (e.g., roofs, roads) compared to vegetated areas. Sparse vegetation was the most challenging class to classify, with user's accuracy as low as 78.57% in 2015, due to spectral confusion with dense vegetation and bareland. This confusion is common in heterogeneous miombo woodland landscapes where vegetation density varies gradually rather than abruptly (Munthali et al., 2020)

5.2.2 LULC Change Dynamics (2015–2025)

The change detection analysis revealed substantial Land Use/Land Cover (LULC) transformations in Dedza Boma between 2015 and 2025. Built-up area increased by 5.73 km², representing a 54.7% growth, while farmland declined by 8.63 km², equivalent to a 39.0% reduction. Vegetation loss accounted for 4.77 km² (10.6% of the study area), making it the most dominant change category observed during the study period. Agricultural expansion was identified as the primary driver of vegetation loss, contributing 3.20 km² (7.1%), followed by urban expansion at 1.78 km² (3.9%).

These findings are consistent with previous studies conducted in other parts of Malawi. Mungai et al. reported similar patterns in Central Malawi, where cropland expansion and built-up growth occurred at the expense of tree cover between 2015 and 2022. Similarly, Munthali et al. (2019) observed increasing agricultural encroachment into forested areas of Dedza District, largely driven by population pressure and growing demand for arable land. These similarities suggest that the trends observed in Dedza Boma reflect broader land transformation processes occurring across Malawi.

The vegetation loss-to-gain ratio of 5.3:1 indicates that vegetation depletion is occurring at a faster rate than regeneration. However, this ratio should be interpreted cautiously because ongoing reforestation and natural regeneration activities may not yet be fully detectable in satellite imagery. Community-based restoration initiatives implemented in and around Dedza Mountain during the 2010s likely contributed to vegetation recovery within the study area. By 2025, many regenerated areas would still have been relatively young and may therefore have appeared spectrally similar to sparse vegetation rather than mature dense forest cover. As a result, some areas undergoing successful regeneration may have been classified as sparse vegetation gain instead of dense vegetation gain.

This spectral behaviour of young regrowth has important methodological implications for interpreting vegetation dynamics. The modest increase in dense vegetation observed in this study (0.60 km²; 26.2% growth) may underestimate the actual extent of

reforestation and regeneration occurring on the ground. According to remote sensing literature, planted trees generally require approximately 10–15 years to develop the closed canopy structure associated with dense vegetation in satellite imagery (Lillesand et al., 2015). Consequently, although the vegetation loss-to-gain ratio suggests an overall decline in vegetation cover, it may not fully capture recent restoration efforts that had not yet matured by 2025.

The 54.7% increase in built-up area reflects broader demographic and urbanization trends occurring within Malawi. According to the National Statistical Office, the country experienced an average annual population growth rate of approximately 2.7% between 2008 and 2018, placing increasing pressure on land resources, particularly in urban and peri-urban centres such as Dedza Boma. Transition matrix analysis revealed that built-up areas expanded mainly into farmland (4.95 km²), with smaller contributions from sparse vegetation (0.60 km²) and dense vegetation (0.30 km²). This indicates that agricultural land is increasingly being converted into urban land uses as the town continues to expand.

The conversion of sparse vegetation into bareland further highlights ongoing land degradation processes within the study area. Land degradation is of concern because it may reduce agricultural productivity and require significant investment for restoration (Veldkamp & Verburg, 2004). Nevertheless, the sharp decline in bareland between 2020 and 2025, from 6.26 km² to 0.54 km², suggests that many previously bare areas were subsequently converted into built-up land or underwent some degree of natural regeneration and vegetation recovery, although the extent of detectable regeneration remained limited.

At the national level, Malawi has committed to restoring degraded and deforested landscapes under the AFR100 initiative. As part of this effort, the Malawi Youth Forest Restoration Program received approximately $2 million in government funding to support the restoration of more than 50,000 hectares of degraded land nationwide. Although these restoration initiatives were active during the study period, their full impact on the Dedza Boma landscape may not yet be completely reflected in the 2025

LULC classification because of the time lag between tree planting and the development of detectable canopy closure in satellite imagery.

5.3 Prediction Model Performance: CNN vs CA-Markov

The validation of prediction models against the actual 2025 LULC map revealed that the Convolutional Neural Network (CNN) outperformed the CA-Markov model, achieving an Overall Accuracy of 89% and Kappa of 0.87, compared to 85% and 0.82 for CA-Markov. These results are consistent with Mungai et al. (2022), who found that neural network models achieved superior prediction accuracy compared to traditional transition models in Central Malawi.

The superior performance of CNN can be attributed to its ability to capture non-linear spatial relationships and complex transition patterns that CA-Markov's transition probability matrix approach cannot fully replicate (Kussul et al., 2017; Kamusoko & Gamba, 2015). CA-Markov assumes stationarity in transition probabilities between time steps, meaning it applies the same rate of change derived from the 2015–2020 period uniformly forward (Eastman, 2012). This assumption becomes increasingly limiting as urban expansion in Dedza Boma accelerated between 2020 and 2025. The CNN, by contrast, learned spatial neighbourhood dependencies directly from the data, enabling it to better simulate the irregular, patch-driven growth pattern of built-up areas at the expense of farmland.

The divergence between CNN and CA-Markov projections increased with temporal distance from the calibration period, which is expected: CA-Markov's stationarity assumption becomes a greater constraint as future scenarios move further from the baseline, while the CNN's learned spatial dynamics allow it to extrapolate non-linear growth trajectories more realistically. This finding has important implications for land use planning in Dedza Boma: relying solely on CA-Markov predictions may underestimate future urban expansion and associated pressures on farmland and natural vegetation.

5.4 Future LULC Scenarios (2030 and 2035)

Both models project the continuation of dominant trends observed between 2015 and 2025, namely ongoing built-up expansion and progressive farmland loss. However, the two models diverge in the magnitude of projected change, with the CNN consistently projecting more aggressive land cover transitions than CA-Markov across both time horizons.

By 2035, the CNN model projects built-up area to reach 20.5 km² (45.6% of total area) a further increase of 4.30 km² from 2025 levels. Under this scenario, farmland would decline to 9.4 km² (20.9%), representing a loss of an additional 4.08 km² over the decade. If these trends materialise, farmland in Dedza Boma could fall below 21% of total land area by 2035, posing significant implications for food security and livelihoods in a district where smallholder farming remains the primary economic activity (Kamanga et al., 2009).

The projected built-up expansion aligns with broader national trends. Malawi's urban population is expected to continue growing, driven by rural-to-urban migration and natural increase (National Statistical Office, 2018). Without strategic land use planning, unmanaged urban sprawl may lead to further loss of agricultural land, increased pressure on natural resources, and greater demand for infrastructure and services.

The projected decline in farmland raises concerns about food security. Dedza District is predominantly agricultural, with most households relying on subsistence farming for their livelihoods (Government of Malawi, 2019). The conversion of farmland to built-up areas may reduce local food production capacity, potentially increasing dependence on food imports and vulnerability to price shocks.

5.5 Comparison with Previous Studies

The findings of this study align with previous LULC research in Malawi. Munthali et al. (2020) used a hybrid Cellular Automata-Markov model to predict LULC changes in Dedza District, projecting continued cropland expansion and tree cover loss to 2030. However, their study did not compare multiple prediction models or evaluate deep learning approaches such as CNN.

Mungai et al. (2022) applied neural network models in Central Malawi and demonstrated that deep learning approaches achieved higher prediction accuracy than traditional transition models. Their study found that neural networks better captured the heterogeneous land cover transitions characteristic of the region's agricultural-woodland mosaic, consistent with the findings of the present study.

The observed urban expansion (54.7% growth in built-up area) is comparable to rates reported in other Malawian urban centres, demonstrating that rapid urbanisation is occurring not only in major cities but also in smaller towns like Dedza Boma.

5.6 Limitations of the Study

Several limitations of this study should be acknowledged when interpreting the results.

First, the desk-based validation approach, while practical and cost-effective, introduces subjectivity and is limited by the availability and temporal alignment of high-resolution reference imagery in Google Earth Pro. Field-based validation using GPS-collected ground points would provide more reliable accuracy estimates but was not feasible due to logistical constraints (Congalton & Green, 2019).

Second, the use of only three time points (2015, 2020, 2025) limits the ability to detect non-linear trends or cyclical patterns in LULC change. More frequent observations (e.g., annual or biennial) would provide a more robust basis for change detection and prediction, enabling the identification of acceleration or deceleration in change rates.

Third, the CA-Markov model assumes stationarity in transition probabilities, meaning it applies the same rates of change observed during the calibration period (2015–2020) to future projections. This assumption may not hold if policy interventions, economic shifts, or climate events alter land use dynamics. The CNN model, while more flexible, still depends on historical patterns and may not capture unprecedented changes.

Fourth, the spatial variables used for prediction (population density, distance to roads, slope, elevation) do not capture all drivers of LULC change. Factors such as land tenure security, market access, soil quality, agricultural input availability, and policy

interventions are known to influence land use decisions but were not included due to data availability constraints (Lambin et al., 2003).

Fifth, the study area of 45 km² focuses on Dedza Boma and its immediate surroundings. The findings may not be directly transferable to other Malawian districts with different socio-economic conditions, population densities, or biophysical characteristics.

Sixth, the prediction results (2030 and 2035) should be interpreted as plausible scenarios under a business-as-usual trajectory, not as definitive forecasts. Future research should incorporate scenario-based modelling that accounts for potential policy interventions, climate change impacts, and alternative development pathways.

Chapter Five: Conclusion and Reccommendations

5.1 Conclusions

This study compared Random Forest (RF) and Support Vector Machine (SVM) for LULC classification, and Convolutional Neural Network (CNN) and CA–Markov for LULC prediction in Dedza Boma, Malawi. The findings reveal that Random Forest was the best classifier for LULC mapping in Dedza Boma, achieving an overall accuracy of 92% and a Kappa coefficient of 0.89. In contrast, SVM performed poorly, with overall accuracy below 75% for all study years. RF handles spectral mixing better in heterogeneous miombo woodland landscapes, making it more suitable for this study area.

Over the ten-year study period (2015–2025), built-up area expanded by 5.73 km² (+54.7%), while farmland declined by 8.63 km² (−39.0%). Vegetation loss (4.77 km²; −10.6%) was the dominant change category, with a vegetation loss-to-gain ratio of 5.3:1, indicating that natural regeneration is not keeping pace with deforestation. These trends reflect rapid urbanisation and agricultural encroachment driven by population growth and demand for arable land.

For prediction, Convolutional Neural Network (CNN) was identified as the superior model, achieving an overall accuracy of 89% (Kappa 0.87) compared to CA–Markov's 85% (Kappa 0.82). CNN captures non-linear spatial patterns that CA–Markov transition matrices cannot, making it more suitable for forecasting LULC changes in rapidly urbanising landscapes. Under the CNN model, built-up area is projected to reach approximately 44% of the total study area by 2035, while farmland will continue to decline. These projections highlight the urgent need for strategic land use planning to protect remaining agricultural land and natural vegetation.

5.2 Recommendations

Based on the findings, it is recommended that Random Forest (92% accuracy) be used for LULC classification, and CNN (89% accuracy) for LULC prediction in Dedza Boma. With 8.63 km² of farmland already lost and built-up projected to reach 44% by 2035, urgent land use planning is needed to protect remaining farmland and vegetation. This comparative framework should be replicated in other Malawian districts.

REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A land use and land cover classification system for use with remote sensor data* (USGS Professional Paper No. 964). United States Geological Survey. https://pubs.usgs.gov/pp/0964/report.pdf

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. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brink, A. B., & Eva, H. D. (2009). Monitoring 25 years of land use change in the humid tropics of Africa. *Applied Geography*, *29*(1), 85–97. https://doi.org/10.1016/j.apgeog.2008.08.003

Brown, D. G., & Duh, J. D. (2004). Land use and land cover change. *Encyclopedia of Life Support Systems*, 1–10. https://www.eolss.net/Sample-Chapters/C09/E6-05-04-06.pdf

Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing* (5th ed.). Guilford Press. https://www.guilford.com/books/Introduction-to-Remote-Sensing/Campbell-Wynne/9781609181765

Congalton, R. G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data: Principles and practices* (2nd ed.). CRC Press. https://doi.org/10.1201/9781420055139

Di Gregorio, A., & Jansen, L. J. M. (2000). *Land cover classification system: Classification concepts and user manual*. Food and Agriculture Organization (FAO). https://www.fao.org/3/x0596e/x0596e00.htm

Dwivedi, R. S., Sreenivas, K., & Ramana, K. V. (2005). Land use/land cover change analysis in relation to soil degradation in parts of Andhra Pradesh, India. *International Journal of Applied Earth Observation and Geoinformation*, *7*(3), 221–230. https://doi.org/10.1016/j.jag.2005.03.005

Eastman, J. R. (2006). *IDRISI Andes guide to GIS and image processing*. Clark University. https://clarklabs.org/idrisi-andes/

Eastman, J. R. (2012). *IDRISI Selva manual*. Clark University. https://clarklabs.org/idrisi-selva/

Ellis, E. C. (2013). Land use and land cover change. In *Encyclopedia of Earth*. National Council for Science and the Environment. https://editors.eol.org/eoearth/wiki/Land_use_and_land_cover_change

Fairbanks, D. H. K., Thompson, M. W., Vink, D. E., Newby, T. S., Van Den Berg, H. M., & Everard, D. A. (2000). *The South African land cover characteristics database: A synopsis of the landscape*. National Botanical Institute. https://www.sanbi.org/resources/reports/

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., & Snyder, P. K. (2005). Global consequences of land use. *Science, 309*(5734), 570–574. https://doi.org/10.1126/science.1111772

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*(1), 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Geist, H. J., & Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation. *BioScience, 52*(2), 143–150. https://doi.org/10.1641/0006-3568(2002)052[0143:PCAUDF]2.0.CO;2

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Government of Malawi. (2019). *Malawi national land use plan*. Ministry of Lands, Housing and Urban Development. https://www.malawi.gov.mw/land-use-plan

Guerschman, J. P., Paruelo, J. M., & Burke, I. C. (2003). Land use impacts on the normalized difference vegetation index in temperate Argentina. *Ecological Applications,*

13(3), 616–628. https://doi.org/10.1890/1051-0761(2003)013[0616:LUIOTN]2.0.CO;2

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jensen, J. R. (2005). *Introductory digital image processing: A remote sensing perspective* (3rd ed.). Pearson Prentice Hall. https://www.pearson.com/us/higher-education/product/Jensen-Introductory-Digital-Image-Processing-A-Remote-Sensing-Perspective-3rd-Edition/9780131453616.html

Kalita, K., & Sarma, A. K. (2015). Markov chain model for prediction of land use/land cover change in urban areas. *Journal of Urban and Environmental Engineering*, *9*(2), 112–120. https://doi.org/10.4090/juee.2015.9.2.112

Kamanga, P., Vedeld, P., & Sjaastad, E. (2009). Forest incomes and rural livelihoods in Chiradzulu District, Malawi. *Ecological Economics*, *68*(3), 613–624. https://doi.org/10.1016/j.ecolecon.2008.08.003

Kamusoko, C., & Gamba, J. (2015). Simulating urban growth using a random forest-cellular automata (RF-CA) model. *ISPRS International Journal of Geo-Information, 4*(2), 447–470. https://doi.org/10.3390/ijgi4020447

Lambin, E. F., Geist, H. J., & Lepers, E. (2003). Dynamics of land-use and land-cover change in tropical regions. *Annual Review of Environment and Resources, 28*(1), 205–241. https://doi.org/10.1146/annurev.energy.28.050302.105459

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., & Xu, J. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change*, *11*(4), 261–269. https://doi.org/10.1016/S0959-3780(01)00007-3

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). *Remote sensing and image interpretation* (7th ed.). Wiley. https://www.wiley.com/en-us/Remote+Sensing+and+Image+Interpretation%2C+7th+Edition-p-9781118343289

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing, 25*(12), 2365–2401. https://doi.org/10.1080/0143116031000139863

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

National Statistical Office. (2018). *2018 Malawi population and housing census report*. Government of Malawi. https://www.nsomalawi.mw/census-2018

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing, 26*(5), 1007–1011. https://doi.org/10.1080/01431160512331313983

Pontius, R. G., & Malanson, J. (2005). Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science, 19*(2), 243–265.
https://doi.org/10.1080/13658810410001713416

Sedano, F., Gong, P., & Ferrao, M. (2005). Land cover assessment in the Brazilian Amazon using MODIS imagery. *Remote Sensing of Environment, 97*(3), 346–360. https://doi.org/10.1016/j.rse.2005.05.003

Seto, K. C., Woodcock, C. E., & Song, C. (2002). Monitoring land-use change in the Pearl River Delta using Landsat TM. *International Journal of Remote Sensing, 23*(10), 1985–2004.
https://doi.org/10.1080/01431160110075532

Singh, A. (1989). Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, *10*(6), 989–1003. https://doi.org/10.1080/01431168908903939

Turner, B. L., Lambin, E. F., & Reenberg, A. (2007). The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences*, *104*(52), 20666–20671. https://doi.org/10.1073/pnas.0704119104

Turner, B. L., Skole, D., Sanderson, S., Fischer, G., Fresco, L., & Leemans, R. (1995). *Land-use and land-cover change: Science/research plan*. IGBP Report No. 35. https://www.igbp.net/publications/igbpreports/igbpreport35.4.d68b9c2117a9a8be928000579.html

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer. https://doi.org/10.1007/978-1-4757-2440-0

Veldkamp, A., & Verburg, P. H. (2004). Modelling land use change and environmental impact. *Journal of Environmental Management*, *72*(1), 1–3. https://doi.org/10.1016/j.jenvman.2004.04.004

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, *39*(12), 4254–4284. https://doi.org/10.1080/01431161.2018.1452075

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., & Peng, S. (2016). Greening of the Earth and its drivers. *Nature Climate Change*, *6*(8), 791–795. https://doi.org/10.1038/nclimate3004

Adepoju, K. A., & Adelabu, S. A. (2020). Land use/land cover change detection and prediction using satellite imagery and cellular automata-Markov model in a Nigerian urban area. *Geocarto International*, *35*(12), 1293–1310. https://doi.org/10.1080/10106049.2019.1573858

Chakraborty, S., Maity, I., & Patel, P. P. (2021). Land use/land cover change detection and prediction using CA-Markov model in a semi-arid region of India. *Environmental Monitoring and Assessment*, *193*(4), 1–18. https://doi.org/10.1007/s10661-021-08921-4

FAO. (2016). *State of the world's forests 2016: Forests and agriculture, land-use challenges and opportunities*. Food and Agriculture Organization of the United

Nations.
https://www.fao.org/3/i5588e/i5588e.pdf
Gashaw, T., Tulu, T., Argaw, M., & Worqlul, A. W. (2018). Land use/land cover dynamics in the Andassa watershed of the Blue Nile Basin, Ethiopia. *Environmental Systems Research*, *7*(1), 1–13. https://doi.org/10.1186/s40068-018-0114-1
IPCC. (2019). *Climate change and land: An IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. Intergovernmental Panel on Climate Change.
https://www.ipcc.ch/srccl/
Moghadam, H. S., & Helbich, M. (2015). Spatiotemporal analysis of land use/land cover change in Isfahan city using satellite imagery and CA-Markov model. *Journal of Environmental Planning and Management*, *58*(10), 1811–1833. https://doi.org/10.1080/09640568.2014.956070
Singh, R. P., & Singh, S. (2020). Land use/land cover change detection and future prediction using CA-Markov model in Varanasi district, India. *Journal of the Indian Society of Remote Sensing*, *48*(8), 1121–1134. https://doi.org/10.1007/s12524-020-01138-