Modelling and predicting Land Use Land Cover Change Using Machine Learning and Remote Sensing: A Case of Dzalanyama Forest, Central Malawi

Hamza Nkhumbwa[1 *], Chikondi Chisenga[1], Emmanuel Chinkaka[1,2], Harineck Tholo[1], Jabulani Nyengere[1].

[1]Ndata School of Climate and Earth Sciences, Malawi University of Science and Technology, P.O. Box 5196, Limbe, Malawi.

[2]Department of Geography and Spatial Sciences, University of Delaware, Newark, USA.

Corresponding author: Hamza Nkhumbwan[1], Ndata School of Climate and Earth Sciences, Malawi University of Science and Technology, P.O Box 5196, Limbe, Malawi. Cel: 265-0-997-386-793. Email: gis-021-20@must.ac.mw

Abstract

Dzalanyama Forest Reserve in Central Malawi is undergoing rapid land use land cover (LULC) changes driven by deforestation , charcoal production and agricultural expansion endangering vital ecological services. This research aims at modelling and predicting LULC dynamics from 2004 to 2034 using machine learning and remote sensing approach. Landsat imageries (2004,2014,2024) were classified using the Random Forest (RF) algorithms because of the ability to handle complex data. For future projections for 2034 , they were simulated using a hybrid model combining SVM and Cellular Automata-Markov (CA-Markov) in Google Earth Engine (GEE) and ArcGIS Pro. Results reveal a drastic decline in forest cover from 67.5% in 2004 to 17.03% in 2024, with further loss predicted to reach 9.07% by 2034. Concurrently , bareland and grassland are projected to expand reflecting increasing environmental damage by human activities. Accuracy assessment yielded overall accuracies above 85% with Kappa coefficients above 0.80 , validating the capability and capacity of the model. This findings align with broader Miombo Woodland degradation trends in Southern Africa. The project highlights the urgent need for targeted conservation interventions , including reforestation , Participatory-GIS (PGIS) and enforce anti-deforestation policies. With the aid of machine learning models and cloud-based platforms for LULC prediction , this research offers a scalable decision-support tool for forest conservation and land Management in Malawi and similar contexts.

Keywords: Dzalanyama Forest, LULC change, Machine learning, Remote sensing, CA-Markov, SVM, Malawi

Introduction

Forests at a global level play critical roles in maintaining ecological balance , mitigating and biodiversity conservation.(Grantham et al., 2020) Forests covering approximately 31% of the Earth's land surface ,they play roles as carbon, sinks, support millions of livelihoods and protect livelihoods(Pan et al., 2024). However , most forests globally face increasing threats from anthropogenic activities such as urbanization, industrial expansion's and agricultural expansions.(Alam et al., 2023). According to (Pendrill et al. (2022), global forest loss continue to strike more especially in the tropical regions experiencing the great negative impact. Sub-Saharan Africa in particular experiences significant forest degradation due to socio-economic pressures and weak regulatory mechanisms. The Miombo woodlands , which span large portions of Southern Africa are the most affected ecosystems due to shifting cultivation , charcoal production and population pressure(Kissanga et al., 2024)

In the African context, forest degradation contributed to declined ecosystem services together with loss of biodiversity, increased vulnerability to climate changes and reduced agricultural productivity. Nations like Ghana, Kenya and Zambia have recorded severe environmental and economic impacts from the sharp drop of forest health.(Ngoma et al., 2021) As mitigation measures, several nations have launched landscape restorations and reforestation programs. Even though these are the interventions, they are inadequately funded , or lack long term monitoring systems (Lindenmayer et al., 2020).A notable challenge across many African nations is the absence of predictive tools that can guide policy decisions and land management strategies based on future land use scenarios .

Malawi's forest reserves face similar challenges. Forests cover approximately 36% of the country, but deforestation rates remain among the highest in the Southern African Region with an estimated rate of 2.6% annually (Ngwira & Watanabe, 2019). The Dzalanyama Forest Reserve , located in Central Malawi , plays a critical role as it serves as a water catchment for the Lilongwe River and provides essential ecological services(Katumbi et al., 2017). Despite its positive impact, the reserve has experienced extensive degradation over for the past two decades(Skole et al., 2021). Between 2004 and 2024, over 50% in approximation of forest cover was lost due to illegal charcoal production, fuel harvesting, urban sprawl from Lilongwe city , and agricultural enchroachment (Nkwanda et al., 2021). This severe forest decline has led to reduced water quality and quantity, increased biodiversity decline, soil erosion together with heightened vulnerability to climate extremes.

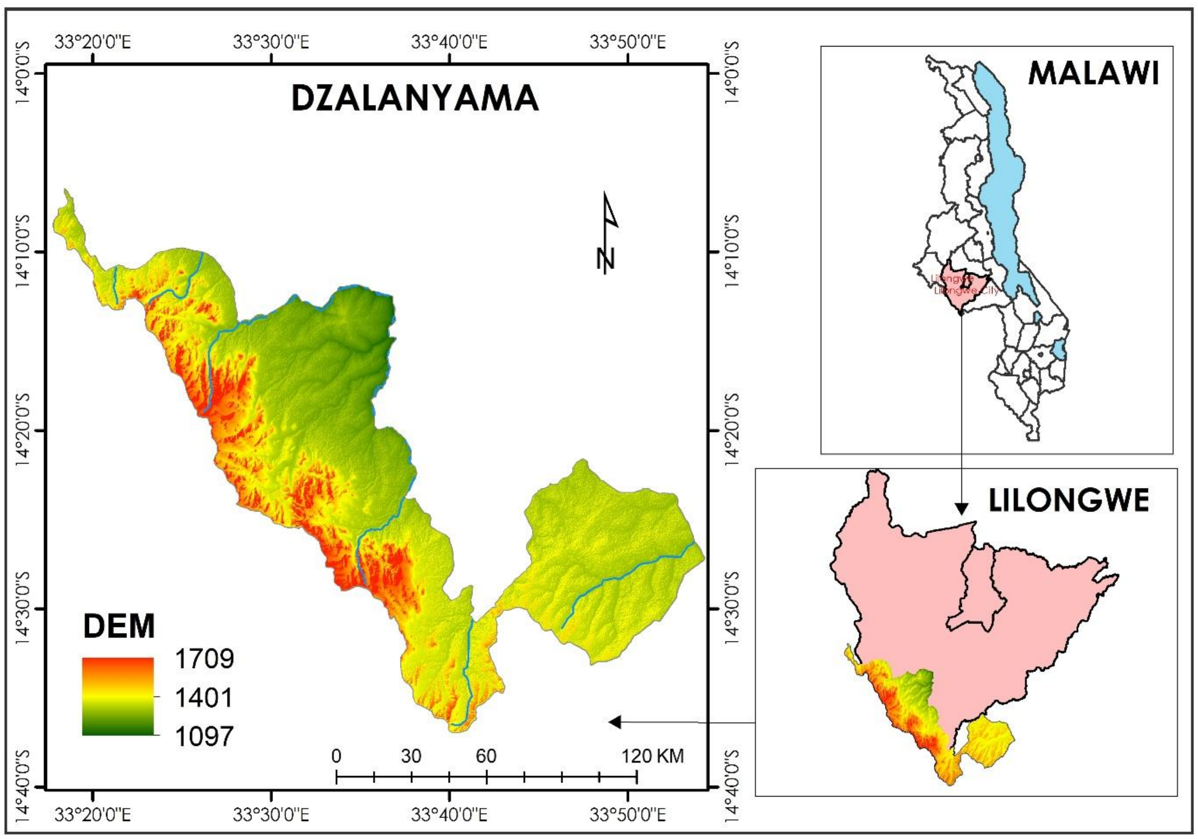
To address these issues, the Malawian government and stakeholders have initiated several programs, including the National Forest Landscape Restoration Strategy(2021-2030), afforestation campaigns and community forest management initiatives. While these efforts are commendable, they are largely reactive and struggle with enforcement, funding constraints and limited use of technology. The crucial part is lack of data driven models that can predict land use land cover (LULC) change and support proactive forest management. This gap hinders effective policy formulation and optimization of the intervention strategies.

This study aims to fill this gap by applying a hybrid modeling approach that combines machine learning and spatial simulation to forecast the future land cover changes in the Dzalanyama Forest Reserve. It integrates the Support Vector Machine (SVM) algorithm for supervised classification and the Cellular Automata-Markov (CA-Markov) model for spatio-temporal prediction(Rahaman et al., 2022).The two rely on satellite-driven data and environmental factors such as , NDVI, Slope, Proximity to roads and the Digital Elevation Model(DEM) together with Land Cover Transitions. Specifically, the objectives are to (1) assess LULC changes from 2004 to 2024, (2) predict LULC distribution in 2034, and (3) provide actionable insights for forest management and policy. Similar approaches have been successfully applied in other regions such as Kenya's Mau Forest Complex where Remote Sensing and Machine learning were utilized to simulate the LULC dynamics and aid in restoration together with preservation plans(Yuh et al., 2023). By modelling and predicting land cover changes for 2034 and beyond, this research contributes to evidence-based planning and forest conservation in Malawi. The insights generated can support decision makers ,development partners and environmental agencies in designing spatially explicit, forward looking policies. This study also align with MW VISION 2063 and SDG 13 (Climate Action) and 15(Life land) by promoting sustainable forest management through advanced geospatial technologies

Methodology and materials

Study Area

Dzalanyama Forest Reserve 93,500 hectares is situated in Lilongwe District, Central Malawi extending towards the Malawi-Mozambique border. The reserve lies between latitudes 14°13′30″S and 14°34′30″S and longitudes 33°30′0″E and 33°40′30″E. Dominated by Miombo woodlands (Brachystegia), the forest plays a critical role in regulating hydrological cycles, particularly for Kamuzu Dams supplying Lilongwe City.



Data Sources and Description

Three epochs of Landsat imagery were utilized: Landsat 5 (2004), Landsat 8 (2014), and Landsat 9 (2024),all accessed through Google Earth Engine (GEE) in the dates of 01/May to 31/October. Ancillary data included DEM (30m resolution) from SRTM, NDVI derived from

Landsat, road network from OpenStreetMap, population density from WorldPop, and slope calculated from DEM.

Table 1 Data Sources and description

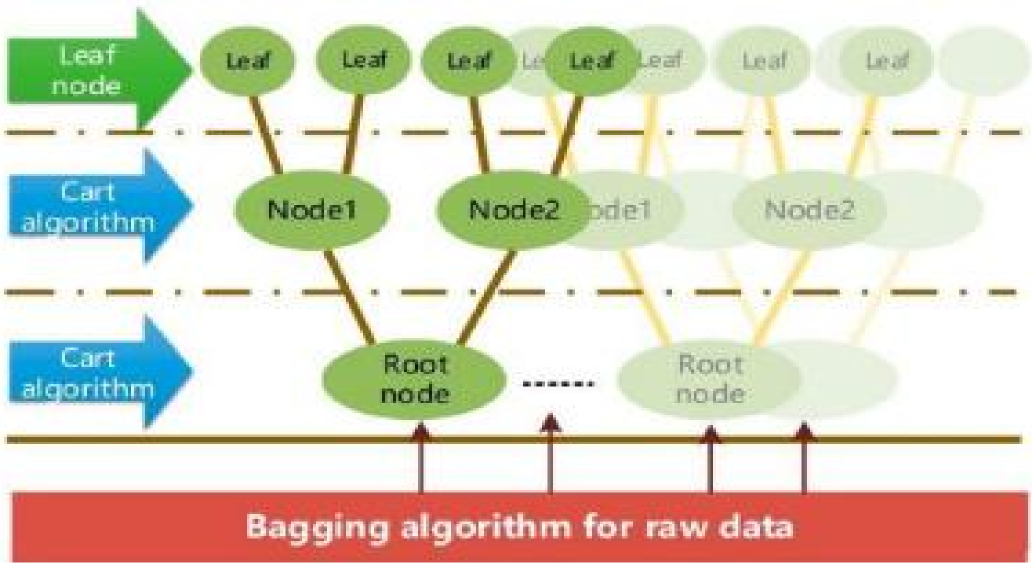
No table of figures entries found.	**Description**	**Source**
NDVI	It is essential for measuring the vegetation health/greenness (values range from -1 to +1). High-resolution, open, accurate, comparable, and timely land use maps are critical for decision-makers in many industry sectors(*Ndvi-Vegetation-Performance-Evaluation-Using-Rs-and-Gis*, 2021)	Landsat imagery ,source GEE (LANDSAT/LT05/C02/T1_TOA, LANDSAT/LT08/C02/T1_TOA, LANDSAT/LT09/C02/T1_TOA)
Land Cover	Classifying the land into categories like forest, grassland, water and bare land for LULC. High-resolution, open, accurate, comparable, and timely land use maps are critical for decision-makers in many industry sectors and developing nations(*Essd-2021-251*, n.d.)	Landsat imagery, source GEE (LANDSAT/LT05/C02/T1_TOA, LANDSAT/LT08/C02/T1_TOA, LANDSAT/LT09/C02/T1_TOA)
DEM	The Elevation data essential for slope / terrain analysis and one of the determining factors in prediction. Key specifications include: Geographic Horizontal Datum WGS84, Vertical Datum EGM96 (Earth Gravitational	Source in GEE ,SRTM(30m) USGS/SRTMGL_1003

	Model 1996), Vertical Units in meters, Spatial Resolution of 1 arc-second (approx. 30 meters) for global coverage, Raster Size of 1-degree tiles, and C-band Wavelength of 5.6 cm.(*SRTM_User_Guide_V3*, n.d.)	
Roads	Vector data for road networks	OpenStreetMap(OSM),Imported to GEE
Population density	Human Population per unit Area	NSO,WorldPop and WorldBank
Slope	Terrain Steepness(derived from DEM)	Source in GEE
Weather Data	Precipitation and temperature	Source in GEE ,WORLDCLIM/V1/BIO

Preprocessing

Images acquired by Landsat sensors are subject to distortion due to atmospheric ,solar, sensor and topographic effects which has effect the final result expected (Bilousov, 2022).Satellite imagery preprocessing is a critical step in ensuring accurate land cover classification as it attempts to minimize these effects to the extent of having desired results for a particular application .. In this study, Landsat images were carefully selected and processed to reduce atmospheric, sensor, and seasonal distortions. Using cloud masking with the QA_PIXEL band, median compositing, and spectral band selection sensitive to vegetation, water, and bare land, the workflow delivered consistent, high-quality inputs. Cloud masking refers to the process of identifying and removing pixels in a satellite image that are contaminated by clouds and in the context of Google Earth Engine Bitmask was applied to remove pixels affected by clouds. (Gonzalo Mateo-Garcia, 2018) This standardized approach enhanced the reliability of classification and enabled meaningful change detection across the 20-year period which was used in studies like Sentinel-2 Classification Using Deep Learning Method (Ekram Mokhtar Rewehel, 2022)and An Artificial Neural Network Classifier Using Spectral & Textural Context (Ramakrishna M. V. Malladi, 2018)**LULC Classification**

Land cover classification for study was performed using Supervised Classification in Google Earth Engine (GEE) , to impose the power of the Random Forest algorithm. Random Forest is a powerful ensemble machine learning algorithm that combines the predictions of multiple decision trees to make a more accurate and stable classification. (Vrushali Y Kulkarni, 2013) It is crucial to have a substantial amount of high quality data for effective data collection. This technique utilizes the concept of decision trees, constructing a collection of decision trees and aggregating their outcomes to generate the ultimate prediction. (Hasan Ahmed Salman, 2024) Every decision tree inside a random forest is constructed using random subsets of data, and each individual tree is trained on a portion of the whole dataset. Subsequently, the outcomes of all decision trees are amalgamated to derive the ultimate forecast.Random Forest (RF) mitigates the correlation between decision trees by employing a random selection of samples and features. Initially, an equivalent quantity of data is randomly chosen from the training sample in the original training data. Furthermore, a subset of the features is chosen at random to construct the decision tree. (X. Solé, 2014)



1.Expected Loss Function

$$E_{XY}[L(Y, f(X)] \qquad (1)$$

2. Loss Functions

Squared Error Loss (Regression):

$$L(Y, f(X) - (Y - f(X))^2 \qquad (2)$$

Zero-One Loss (Classification):

$$L(Y, f(X)) - I(Y \neq f(X)) - \begin{cases} 0 & if\ Y - f(X) \\ 1 & otherwise \end{cases} \qquad (3)$$

3.Optimal Predictor

Regression :

$$f(x) - E(Y|X - x) \tag{4}$$

Classification

$$f(x) - arg\overset{max}{_{y \in Y}}P(Y - y|X - x) \tag{5}$$

Deleted[hamza]: <math>

4.Ensemble Predictor Equations

$$f(x) - \frac{1}{j}\sum_{j=1}^{j} h_j\,(x) \tag{6}$$

Classification (Majority vote):

$$f(x) - arg\overset{max}{_{y \in Y}}\sum_{j=1}^{j}(y - h_j(x)) \tag{7}$$

5.Node Splitting Criteria

Regression (Mean Squared Residual):

$$Q = \frac{1}{n}\sum_{i=1}^{n}(y_i - y)^2 \tag{8}$$

Where $$y = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{9}$$

Classification (Gini Index):

$$Q = \sum_{k \neq k'} P_k P_{K'} \tag{10}$$

where $$P_{k=\frac{1}{n}\sum_{i=1}^{n}\, I(y_i - k)} \tag{11}$$

Split Criterion (Weighted impurity)

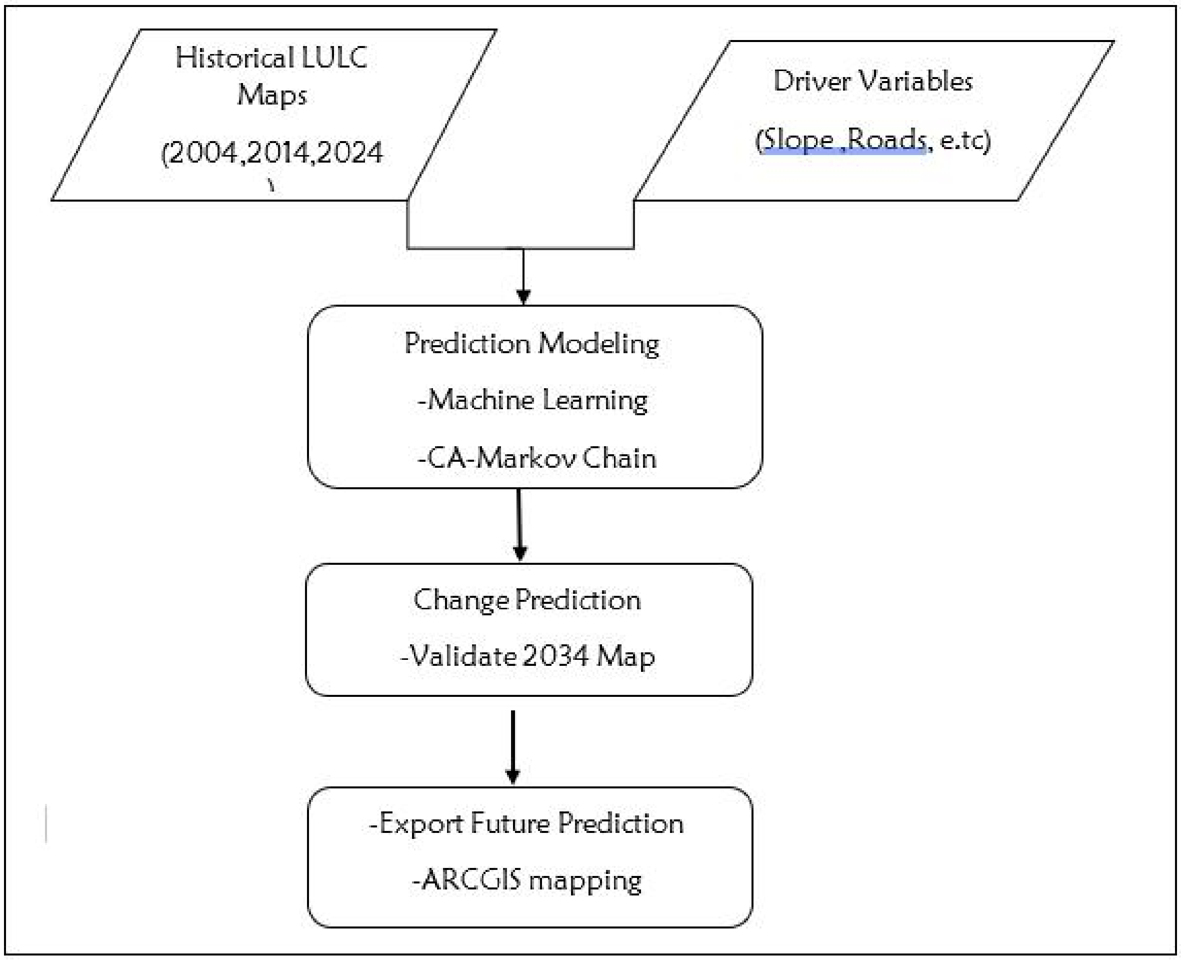
$$Q_{split} - n_L Q_L + n_R Q_R \tag{12}$$

These equations together define how Random Forest construct evalutes and select decision tree

High-quality training samples of the Water, Forest, Bare-Land, were collected using skills and knowledge of spectral band combinations and the interactive tools within GEE. These samples captured unique spectral signatures, which formed the basis for model training. The Random Forest classifier with its great robustness, speed, and resistance to overfitting was trained using Landsat Surface Reflectance bands according to (Xu et al., 2019) . Once trained, the model was applied to preprocessed imagery from 2004, 2014 and 2024 to generate accurate land cover maps. This process ensured precise classification across years and provided a foundation for change detection and future land cover prediction.

Prediction Modelling

Deleted[hamza]:



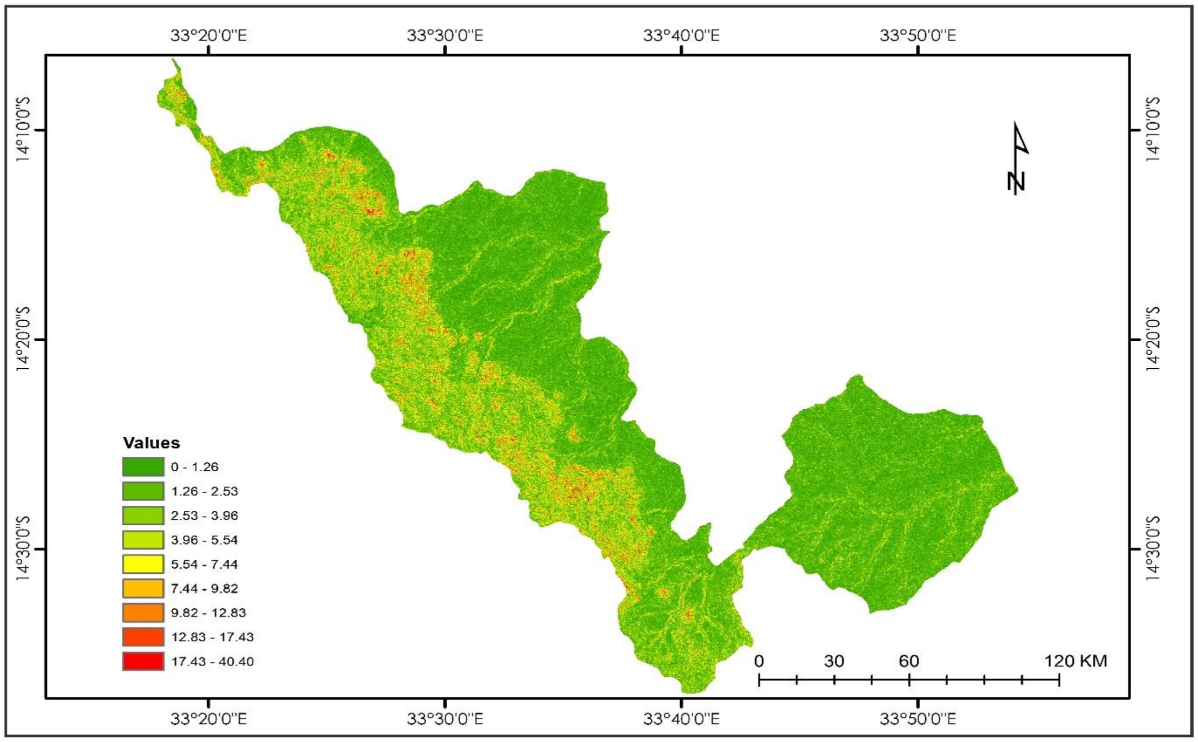


Figure showing slope

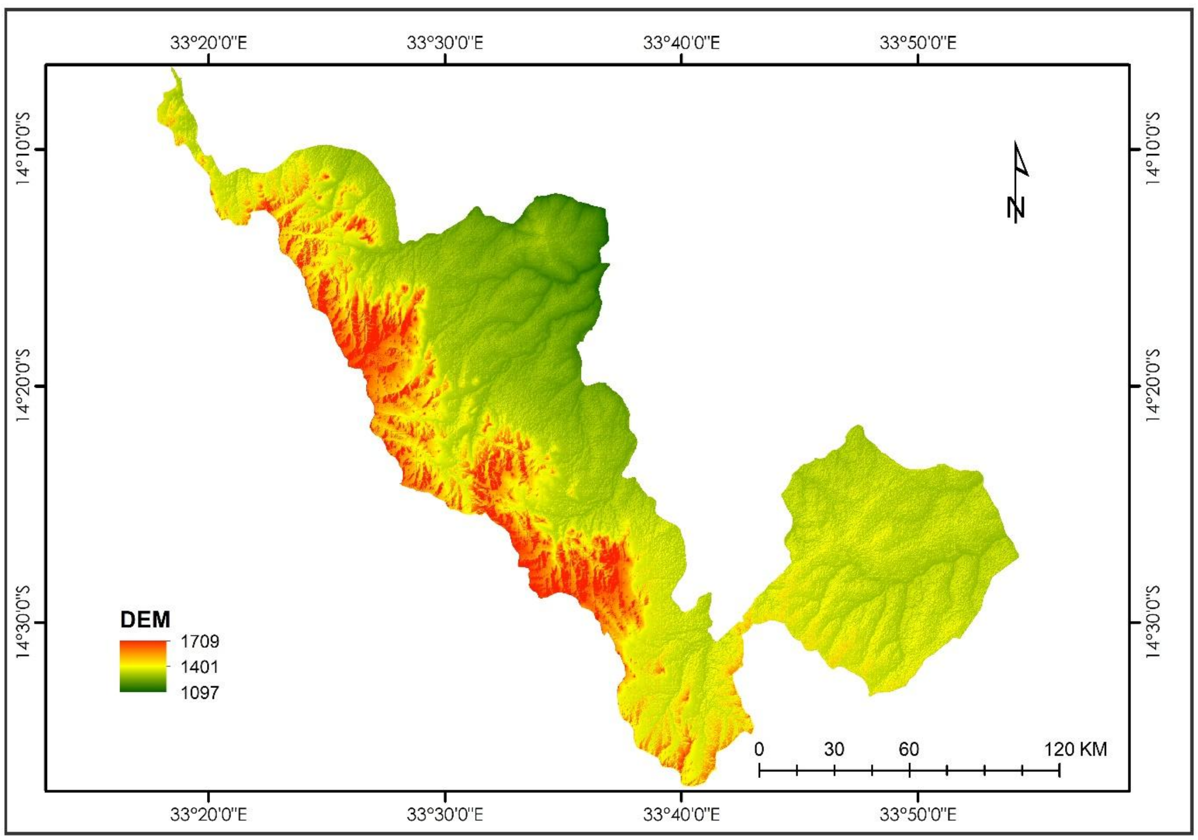


Figure showing DEM

Future LULC distribution for 2034 was projected using a hybrid model comprising of SVM classification and CA-Markov chain simulation. This was employed to enhance the accuracy and spatial realism of land use / land cover (LULC) change prediction. The SVM algorithm has robustness in handling high dimensional and non-linear data and was used for supervised classification of multi-temporal satellite imagery to ensure precise delineation of existing land cover classes. (Islam Atef, 2024) The Markov Chain Model was applied to analyze historical LULC transitions and compute the corresponding transition probability matrices , capturing the nature of temporal land dynamics. To embed the spatial dependency and effects of neighborhood into the prediction process , Cellular Automata were incorporated for spatial allocation of future land states based on local change rules and influence of surrounding cells (Salman A. H. Selmy 1, 2023).

CA–Markov Chain Model Formulas

Markov Chain Transition Equation

$$S(t+1) = P_{ij}.S(t) \qquad (13)$$

S(t): State vector (land use proportions at time t)

S(t+1): Predicted state vector at time t+1

P_{ij} *: Transition probability matrix*

$$P_{ij} = \begin{bmatrix} p_{11} & \cdots & p_{13} \\ \vdots & \ddots & \vdots \\ p_{31} & \cdots & p_{33} \end{bmatrix} \qquad (14)$$

Where :

P_{ij} *= probability of land use class I changing to class j*

$$0 \leq p_{ij} \leq 1, and\ \sum_{j=1}^{n} p_{ij} = 1 \qquad (15)$$

Cellular Automata Transition Rule

$$S_{i,j}^{t+1} = f(S_{i,j}^{t+1}, \Omega_{i,j}^{t}, V) \qquad (16)$$

$S_{i,j}^{t+1}$ *:State of cell (i, j) at time (t+1)*

$S_{i,j}^{t}$ *: current state of cell (i,j) at time t*

$\Omega_{i,j}^{t}$ *: Neighborhood state of the cell*

V: suitability factors (e.g slope,distance to road, population density)

f: Transition function (based on suitability + neighborhood influence)

Combined CA-Markov Simulation

Final predicted state results from combining the temporal transition probabilities with the spatial allocation :

Markov gives: how much change occurs (quantity, temporarily)

CA: allocates where change happens (location , spatially 0

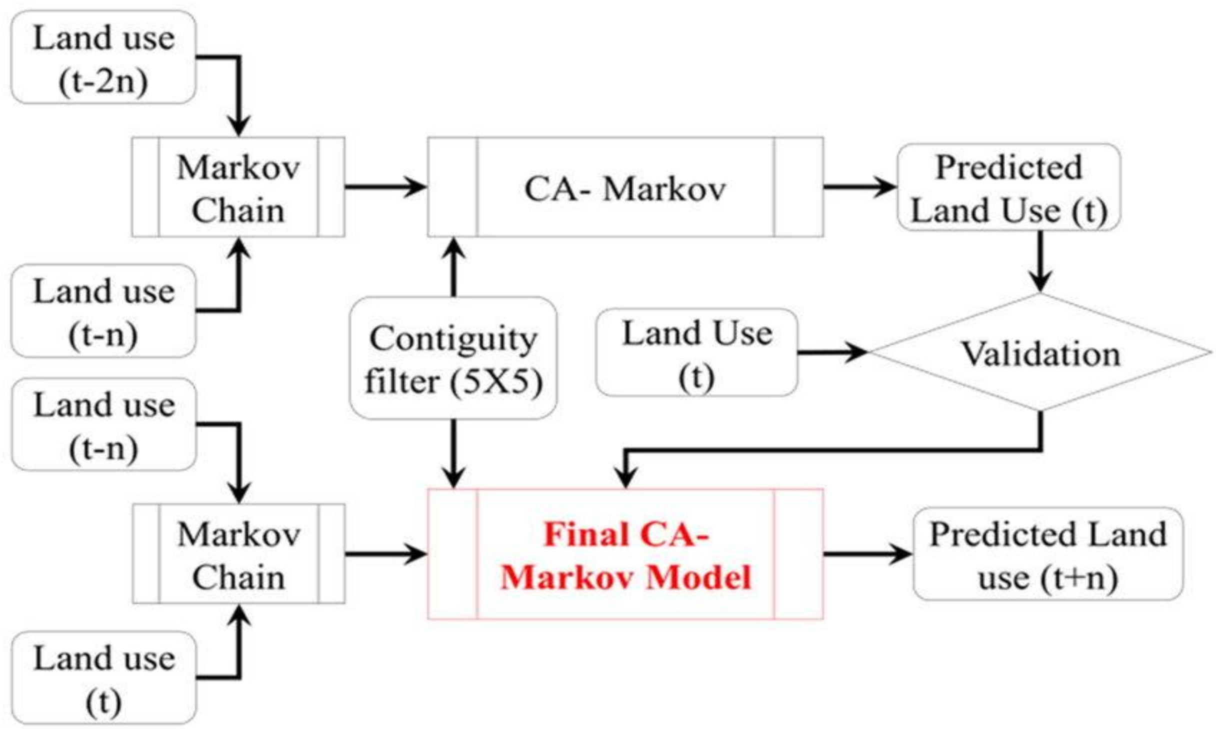


Figure showing how CA-Markov Model

Support Vector machine is a powerful supervised machine learning algorithm used for LULC prediction due to strong classification capabilities and robustness in handling high-dimensional ,non-linear datasets. (Lemenkova, 2024) In prediction , SVM constructs optimal hyperplane that separates data points of different land cover classes with maximum possible margin ,enhancing the model's generalizability to unseen data. (Noamene Baccari, 2024) For cases where the classes are not linearly separable in the original feature space, SVM use's kernel function to project the data into higher dimensional space, to capture complex spatial patterns. Once trained on historical LULC data ,the SVM model can accurately classify future land states. (Bwalya Mutale, 2024)

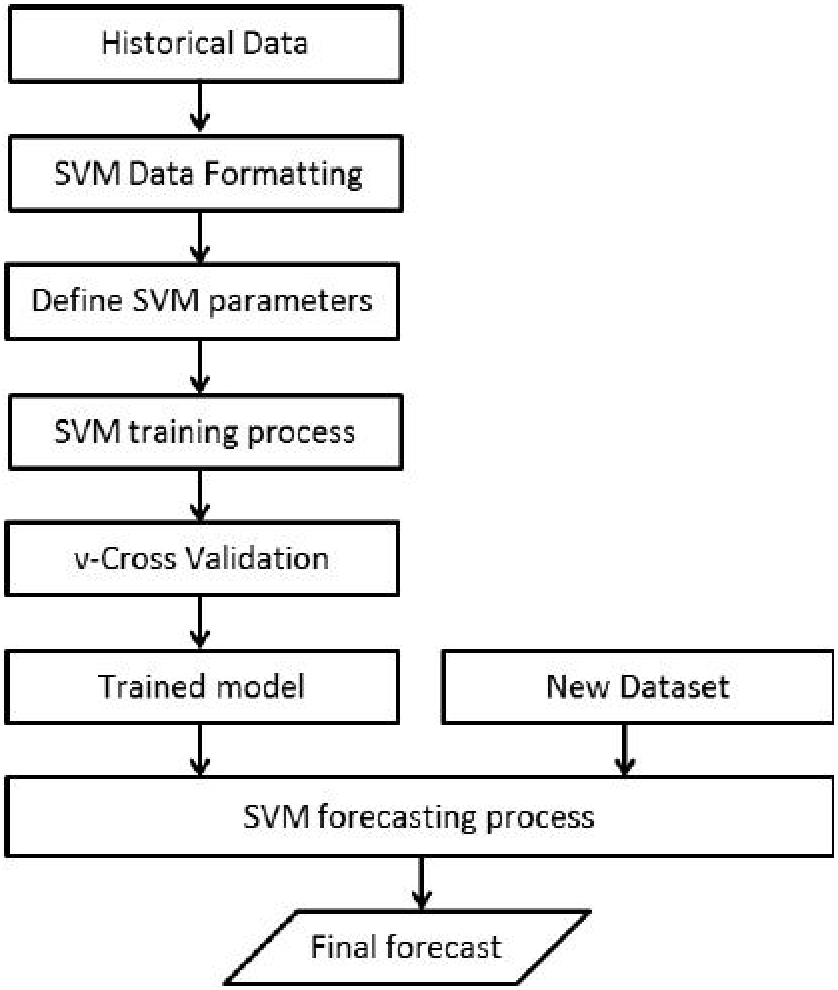


Figure showing SVM flow chart

Support Vector Machine (SVM) Prediction Formulas

Linear SVM Decision Function :

$$f(x) = W^T X + b \quad (17)$$

X: Feature vector (e.g , spectral bands, NDVI)

W: Weight vector (normal to the hyperlane)

b: Bias (offset from the origin)

decision rule:

$$y = sign(f(x) = \begin{cases} +1, & if\ f(x) \geq 0 \\ -1, & otherwise \end{cases} \quad (18)$$

Optimization Problem (Hard Margin SVM)

$$\min_{\cdot w,b} \frac{1}{2}||w||^2 \ subject\ to\ \ y_i(W^T X_i + b) \geq 1, \forall_i \quad (19)$$

Soft Margin SVM (for non-separable data)

$$\min_{w,b,\varepsilon} \frac{1}{2}||w||^2 + \ C \sum_{i=1}^{n} \varepsilon_i \ subject\ to\ \ y_i\,(W^T X_i + \ b) \geq 1, \forall_i \quad (20)$$

Kernel Trick (Nonlinear SVM)

Linear:

$$K(x, x^{'}) = x^T x^{'} \qquad (21)$$

Polynomial:

$$K(x, x^{'}) = (x^T x^{'} + 1)^d \qquad (22)$$

Radial Basis Function (RBF):

$$K(x, x') = \exp\,(-\gamma||x - x^{'}||^2) \qquad (23)$$

Final prediction function (Dual Form with Kernels):

$$f(x) = \sum_{i=1}^{n} \sigma_i y_i K(K(x, x') + b \qquad (24)$$

σ_i *: lagrange multipliers from optimization*

K : kernel function

The integration of SVM's classification strength with the probabilistic forecasting capability of Markov Chains together with the spatial diffusion mechanism of Cellular Automata yields a comprehensive and reliable tool for simulating complex spatio-temporal land cover changes. (by Auwalu Faisal Koko 1ORCID, 2022) This CA-Markov-SVM hybrid model is particularly valuable in regions experiencing rapid urbanization and environmental transformation , as it offers both predictive precision and spatial coherence , essential for sustainable land use planning and environmental management. (Zhenyu Zhang 1, 2023)

Predictor variables included DEM, slope, NDVI, distance to roads, temperature and population density. A total of 5000 stratified sample points were used with 70;30 split into training and testing subsets. SVM is effective for high dimensional spaces and performs well making it suitable for complex classification tasks (NG, 2020). SVM works very well in a high dimension features space. It is effective even when feature dimension is greater than the number of samples in the training data. (Anirban Dutta Choudhury, 2023) Furthermore CA-Markov model was used along side the SVM due to the ability to detect dynamic degree of estimating spatial and temporal changes of the various land uses in Dzalanyama Forest hence forming a Hybrid model.(Abdelsamie et al., 2024.) To predict the LULC pattern for 2034 , Dzalanyama Forest, hybrid modeling approach was implemented through integrating spatial simulation and machine learning techniques. Using GEE , supervised classification was done using Support Vector Machine (SVM) algorithms where the 2004, 2014 and 2024 LULC maps served as the baseline data. A set of environmental predictors was developed and stacked to help contribute to the prediction. The predictors were Digital Elevation Model(DEM), Slope, Normalized Vegetation Index (NDVI), Worldclimate Temperature (BI001), Distance to Roads, and population density (Anantha M. Prasad, 2022) . 5000 training samples points were stratified from the 2024 LULC and enriched with predictors. The samples were then split into a ratio of 7:3 of training and testing subsets.(Phinzi et al., 2022). The SVM model was then trained and applied to classify the future 2034 in parallel with the Cellular Automata-Markov (CA-MARKOV) chain model which was used to simulate the land change dynamics and capturing temporal transitions and all spatial dependencies.(Nyamekye et al., 2021) This was possible because of the 2004,2014, and 2024 maps of which the CA-Markov was perfect in projecting spatial land change patterns using all the historical transitions probabilities and neighborhood influence.(Abdelsamie et al., 2024) The hybrid model was ideal due to robust classification with dynamic land change modeling and the integration is widely recognized for LULC prediction because of its ability to reflect the spatial-temporal dynamics and the statistical learning. This methodology ensured a reliable , data-driven prediction framework suit for land management and conservation in Dzalanyama.**Accuracy Assesment**

The accuracy metrics comprised of the Overall Accuracy , Producer's Accuracy , User's Accuracy and Kappa Coefficient. Accuracy assessment is an important step for data validation, and all the land cover classification results underwent this step. If focused on comparing the classified maps with some reference data to determine the reliability of the classification. The accuracy assessment generally focused on Overall Accuracy(OA) quantifying pixels correctly classified and the Producers Accuracy reflects on the probability on how well a map maker's classification identifies true instance's of a specific class. (*Russell G. Congalton* , 2014). Furthermore the User's Accuracy its used for denoting the likelihood that the pixel classified is really in that category. The User Accuracy is the probability that a value predicted to be in a certain class really is that class. The probability is based on the fraction of correctly predicted values to the total number of values predicted to be in a class and the Kappa Coefficient measures the agreement between the reference data and the classified map Over 400 reference points were used to validate all the results obtained.

Overall Accuracy Formula

The proportion of correctly classified samples:

$$Overall\ Accuracy = \frac{\sum_{i=1}^{k} n_{ii}}{N} \qquad (25)$$

Producer's Accuracy (Sensitivity)

Represents how well reference pixels of a given class are correctly classified:

$$Producer's\ Accuracy = \frac{n_{ii}}{\sum_{j=1}^{k} n_{ij}} \qquad (26)$$

User's Accuracy (Precision)

Indicates the reliability of the map for each class (how many of the predicted class are correct :

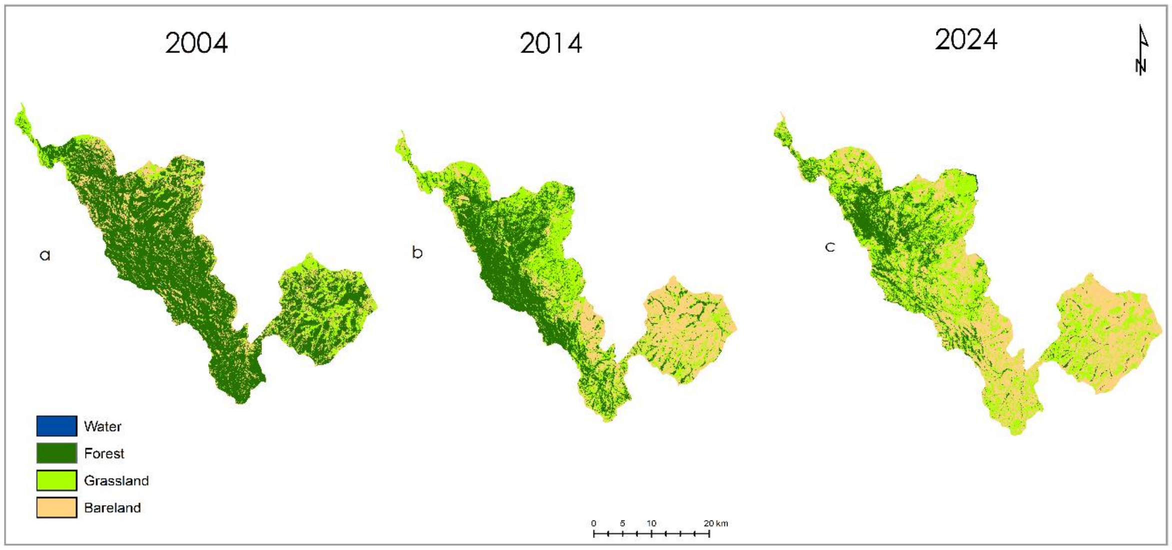
$$User's\ Accuracy = \frac{n_{ii}}{\sum_{j=1}^{k} n_{ji}} \qquad (27)$$

Kappa Coefficient

Accounts for agreement occurring by chance:

$$Kappa\ Coefficient(k) = \frac{N \sum n_{ii} - \sum (n_{i+}.n_{+i})}{N^2 - \sum (n_{i+}.n_{+i})} \qquad (28)$$

Results LULC Changes (2004-2024)



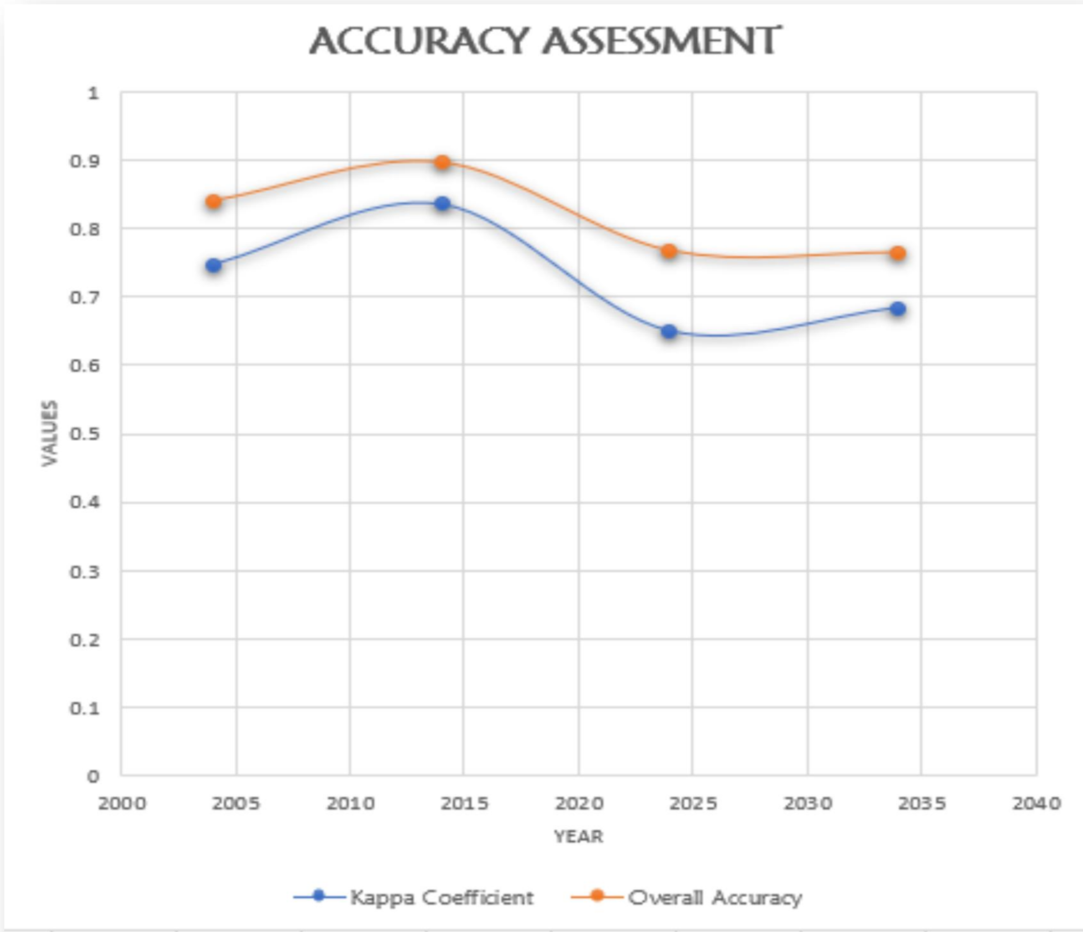
Forest cover decreased from 66,189 ha (67.5%) in 2004 to 17,009 ha (17.03%) in 2024. Bareland expanded from 19,595 ha (20%) to 41,777 ha (42%), while grassland increased from 12,630 ha (12%) to 39,608 ha (40%). Water bodies remained minimal with changes.

Net-change

LULC	2004 Area in ha	2014 Area in ha	2024 Area in ha
WATER	45.39	50.68	65.31
FOREST	66189.91	36557.65	17009.21
GRASSLAND	12629.58	32945.01	39608.37
BARELAND	19594.96	28955.47	41776.93

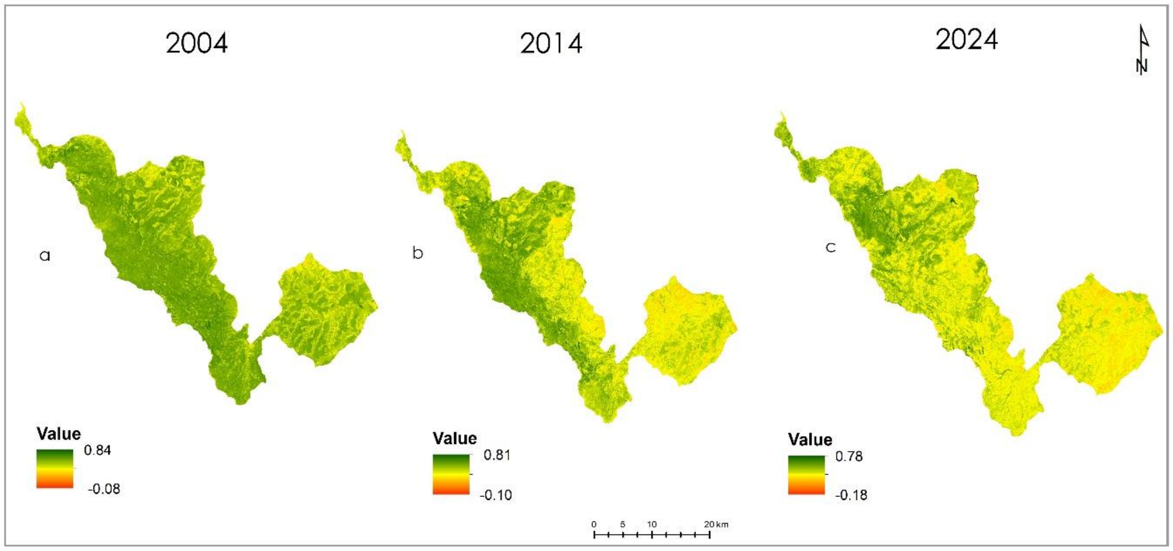
Accuracy Assessment

Classification achieved Overall Accuracies above 85% with Kappa coefficients exceeding 0.80, indicating strong model reliability.



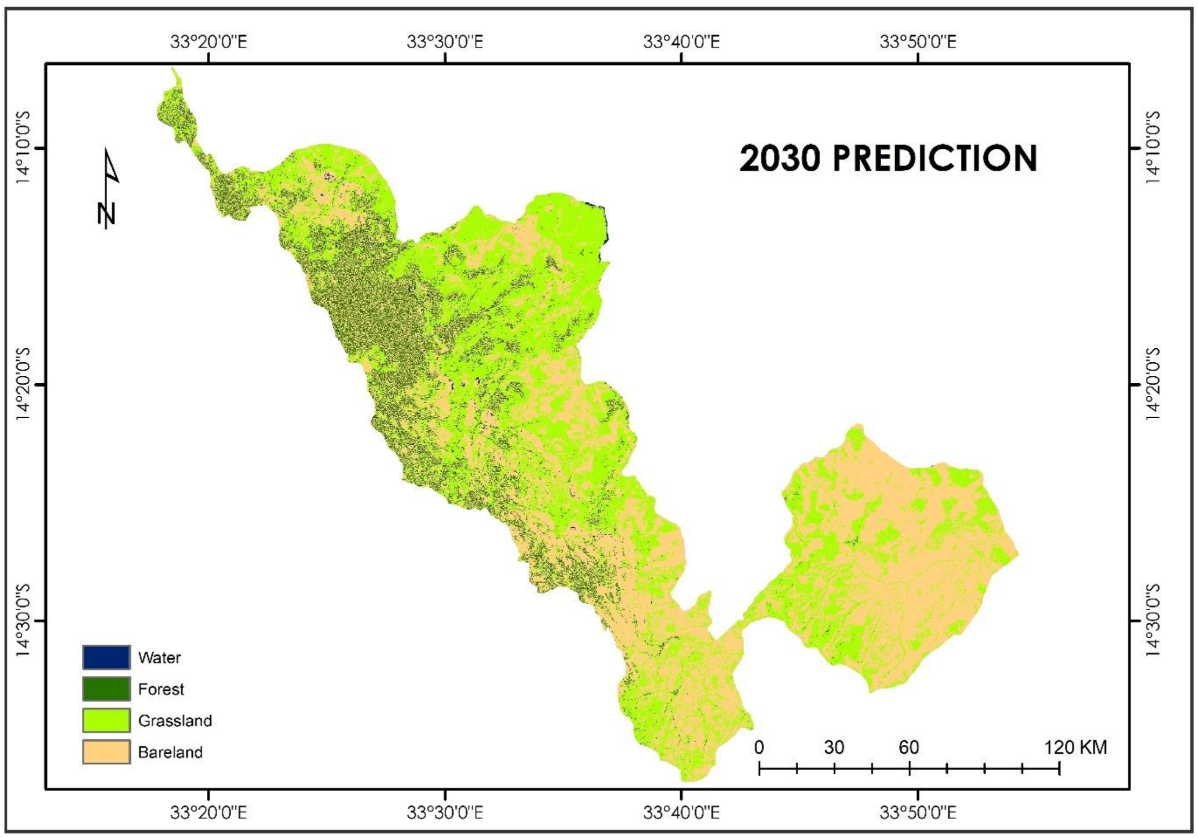
NDVI Trends

NDVI values declined over time , reflecting forest degradation. In 2004, NDVI ranged from -0.08 to +0.84 and the year 2024 ranged from -0.18 to 0.78 , which shows a decrease or decline in vegetation productivity.



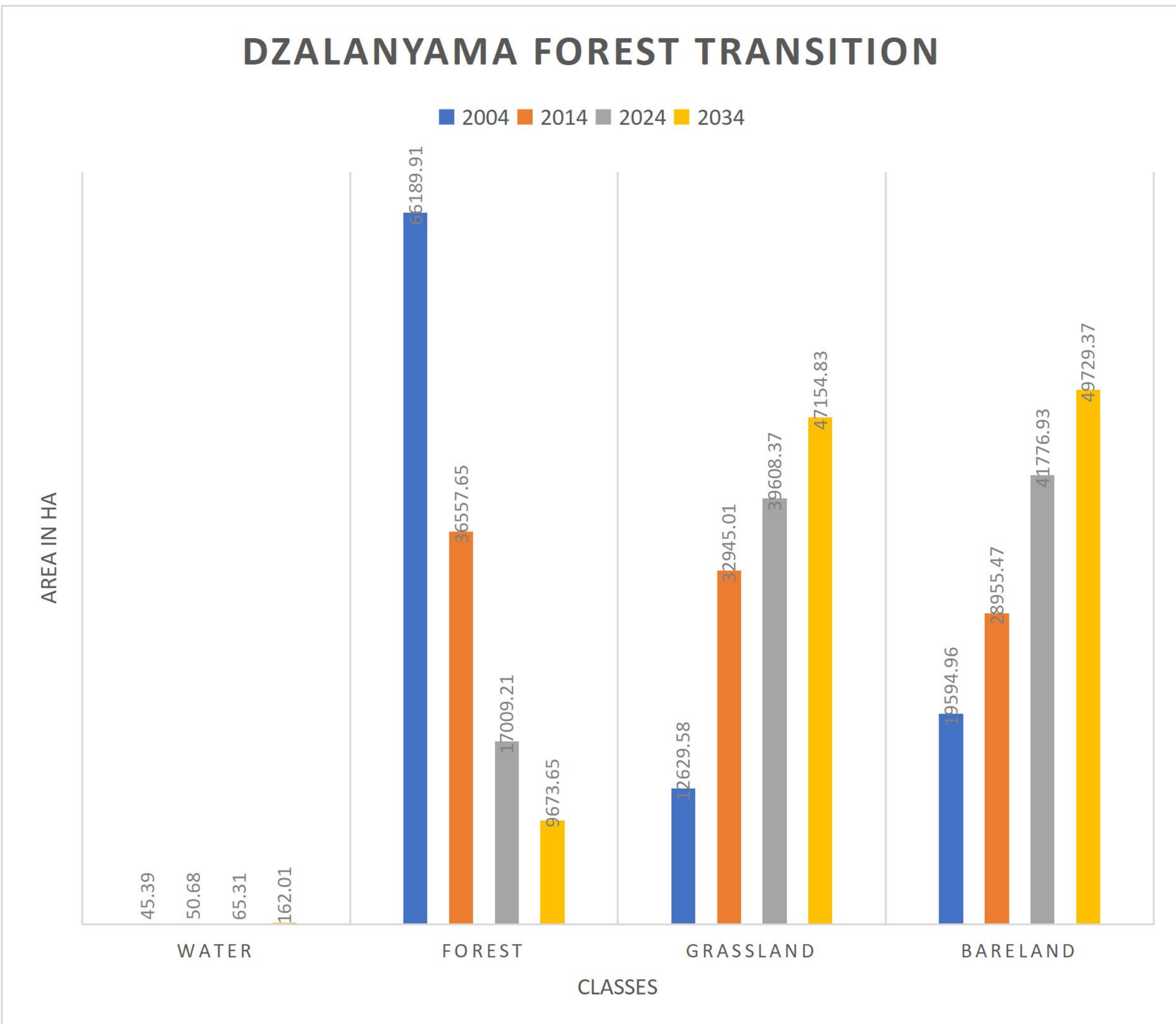
LULC Prediction for 2034

The model predicts further forest decline to 9,674 ha (9.07%). Bareland is projected to increase to 49,729 ha (46.6%), and grassland to 47,155 ha (44.18%). Water bodies show a marginal increase. These results highlight accelerating land transformation and forest fragmentation.



LULC	2034 Area in Ha
WATER	162.01
FOREST	9673.65
GRASSLAND	47154.82
BARELAND	49729.37

Temporal Changes from 2004 to 2034



Discussion

The observed and predicted LULC trends in Dzalanyama Forest mirror broader patterns of Miombo woodland degradation in Southern Africa. Drivers include charcoal production, subsistence agriculture, and population growth, particularly around road networks and settlements. The model's predictions are consistent with similar studies in Ghana and Kenya, which reported forest loss linked to human encroachment and infrastructure expansion.

The integration of environmental predictors, particularly slope, NDVI, and road proximity, enhanced the model's spatial realism. Areas near roads exhibited higher deforestation rates, aligning with previous findings in Borneo and Sub-Saharan Africa. NDVI served as a sensitive indicator of vegetation health, capturing gradual degradation not visible in categorical maps.

Policy implications are significant. Without intervention, continued forest loss will jeopardize Lilongwe's water security, biodiversity conservation, and Malawi's Vision 2063 environmental targets. Machine learning-based prediction models, such as the one employed here, offer valuable decision-support tools for land managers and policymakers. Community-based conservation, enforcement of forestry laws, and alternative livelihoods (e.g., agroforestry, renewable energy) are urgently needed.

Conclusion

This study demonstrates the successful application of machine learning and remote sensing to model and predict LULC change in Dzalanyama Forest Reserve. Forest cover has declined from 67.5% in 2004 to 17.03% in 2024 and is projected to drop further to 9.07% by 2034. Bareland and grassland expansion reflect growing anthropogenic pressures.

The hybrid SVM and CA-Markov model, validated with high classification accuracy, provides a robust framework for forecasting land dynamics. By integrating environmental predictors, the model offers spatially explicit insights for targeted conservation.

Urgent interventions are needed to reverse degradation trends, including stricter policy enforcement, technology-driven monitoring using platforms like GEE, and community-led reforestation. The methodology is scalable and can inform land management strategies in other degraded Miombo ecosystems across Africa.

REFERENCES

Abdelsamie, E. A., Mustafa, A. A., El-Sorogy, A. S., Maswada, H. F., Almadani, S. A., Shokr, M. S., & El-Desoky, A. I. (n.d.). *Current and Potential Land Use/Land Cover (LULC) Scenarios in Dry Lands Using a CA-Markov Simulation Model and the Classification and Regression Tree (CART) Method: A Cloud-Based Google Earth Engine (GEE) Approach*.

Accuracy_assessment_and_validation_of_remotely_sen (1). (2020).

Alam, N., Bohnett, E., Zafar, M., Sher, H., Ahmad, B., Ramadan, M., Ahmad, M., Ullah, Z., Ali, A., Khan, S. M., & Syed, K. (2023). Impact of Anthropogenic Threats on Species Diversity: A Case Study of the Sub-Himalayan Tropical Dry Deciduous Forests of Pakistan. *Sustainability*. https://doi.org/10.3390/su15032829

Grantham, H., Duncan, A., Evans, T., Jones, K., Beyer, H., Schuster, R., Walston, J., Ray, J., Robinson, J., Callow, M., Clements, T., Costa, H., DeGemmis, A., Elsen, P., Ervin, J., Franco, P., Goldman, E., Goetz, S., Hansen, A., … Watson, J. (2020). Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-19493-3

Katumbi, N., Nyengere, J., & Mkandawire, E. (2017). *Drivers of Deforestation and Forest Degradation in Dzalanyama Forest Reserve in Malawi*. https://consensus.app/papers/drivers-of-deforestation-and-forest-degradation-in-katumbi-nyengere/e7a9f122138c50839848e0e00ce0c9a6/

Kissanga, R., Catarino, L., Máguas, C., & Cabral, A. (2024). Dynamics of land-cover change and characterization of charcoal production and trade in southwestern Angola. *Remote Sensing Applications: Society and Environment*. https://doi.org/10.1016/j.rsase.2024.101162

Lindenmayer, D., Woinarski, J., Legge, S., Southwell, D., Lavery, T., Robinson, N., Scheele, B., & Wintle, B. (2020). A checklist of attributes for effective monitoring of threatened species and threatened ecosystems. *Journal of Environmental Management*, *262*, 110312. https://doi.org/10.1016/j.jenvman.2020.110312

Ndvi-vegetation-performance-evaluation-using-rs-and-gis. (2021).

Ngoma, H., Pelletier, J., Mulenga, B., & Subakanya, M. (2021). Climate-smart agriculture, cropland expansion and deforestation in Zambia: Linkages, processes and drivers. *Land Use Policy*, *107*, 105482. https://doi.org/10.1016/J.LANDUSEPOL.2021.105482

Ngwira, S., & Watanabe, T. (2019). An Analysis of the Causes of Deforestation in Malawi: A Case of Mwazisi. *Land*. https://doi.org/10.3390/LAND8030048

Nkwanda, I., Feyisa, G., Zewge, F., & Makwinja, R. (2021). Impact of land-use/land-cover dynamics on water quality in the Upper Lilongwe River basin, Malawi. *International Journal of Energy and Water Resources*, *5*, 193–204. https://doi.org/10.1007/s42108-021-00125-5

Nyamekye, C., Kwofie, S., Agyapong, E., Ofosu, S. A., Arthur, R., & Appiah, L. B. (2021). Integrating support vector machine and cellular automata for modelling land cover change in the tropical rainforest under equatorial climate in Ghana. *Current Research in Environmental Sustainability*, *3*, 100052. https://doi.org/10.1016/j.crsust.2021.100052

Pan, Y., Birdsey, R., Phillips, O., Houghton, R., Fang, J., Kauppi, P., Keith, H., Kurz, W., Ito, A., Lewis, S., Nabuurs, G., Shvidenko, A., Hashimoto, S., Lerink, B., Schepaschenko, D., Castanho, A., & Murdiyarso, D. (2024). The enduring world forest carbon sink. *Nature*, *631 8021*, 563–569. https://doi.org/10.1038/s41586-024-07602-x

Phinzi, K., Ngetar, N. S., Pham, Q., Chakilu, G. G., & Szabó, S. (n.d.). Understanding the role of training sample size in the uncertainty of high-resolution LULC mapping using random forest. *Earth Science Informatics*.

Rahaman, Z., Kafy, A., Faisal, A.-A.-, Rakib, A. A., Jahir, D. Md. A., Fattah, M., Kalaivani, S., Rathi, R., Mallik, S., & Rahman, M. (2022). Predicting Microscale Land Use/Land Cover Changes Using Cellular Automata Algorithm on the Northwest Coast of Peninsular Malaysia. *Earth Systems and Environment*, *6*, 817–835. https://doi.org/10.1007/s41748-022-00318-w

Skole, D., Samek, J., Mbow, C., Chirwa, M., Ndalowa, D., Tumeo, T., Kachamba, D., Kamoto, J., Chioza, A., & Kamangadazi, F. (2021). Direct Measurement of Forest Degradation Rates in Malawi: Toward a National Forest Monitoring System to Support REDD+. *Forests*, *12*, 426. https://doi.org/10.3390/F12040426

SRTM_User_Guide_V3. (2017).

Xu, S., Zhao, Q., Yin, K., Zhang, F., Liu, D., & Yang, G. (2019). Combining random forest and support vector machines for object-based rural-land-cover classification using high spatial resolution imagery. *Journal of Applied Remote Sensing*, *13*(01), 1. https://doi.org/10.1117/1.JRS.13.014521

Yuh, Y., Tracz, W., Matthews, H., & Turner, S. (2023). Application of machine learning approaches for land cover monitoring in northern Cameroon. *Ecol. Informatics*, *74*, 101955. https://doi.org/10.1016/j.ecoinf.2022.101955