ORIGINAL ARTICLE

Modelling cropland expansion and its drivers in Trans Nzoia County, Kenya

Harison Kiplagat Kipkulei[1,2,3] · Sonoko Dorothea Bellingrath-Kimura[1,2] · Marcos Lana[4] · Gohar Ghazaryan[1] · Mark Boitt[5] · Stefan Sieber[1,2]

Abstract
Population growth and increasing demand for agricultural production continue to drive global cropland expansions. These expansions lead to the overexploitation of fragile ecosystems, propagating land degradation, and the loss of natural diversity. This study aimed to identify the factors driving land use/land cover changes (LULCCs) and subsequent cropland expansion in Trans Nzoia County in Kenya. Landsat images were used to characterize the temporal LULCCs in 30 years and to derive cropland expansions using change detection. Logistic regression (LR), boosted regression trees (BRTs), and evidence belief functions (EBFs) were used to model the potential drivers of cropland expansion. The candidate variables included proximity and biophysical, climatic, and socioeconomic factors. The results showed that croplands replaced other natural land covers, expanding by 38% between 1990 and 2020. The expansion in croplands has been at the expense of forestland, wetland, and grassland losses, which declined in coverage by 33%, 71%, and 50%, respectively. All the models predicted elevation, proximity to rivers, and soil pH as the critical drivers of cropland expansion. Cropland expansions dominated areas bordering the Mt. Elgon forest and Cherangany hills ecosystems. The results further revealed that the logistic regression model achieved the highest accuracy, with an area under the curve (AUC) of 0.96. In contrast, EBF and the BRT models depicted AUC values of 0.86 and 0.77, respectively. The findings exemplify the relationships between different potential drivers of cropland expansion and contribute to developing appropriate strategies that balance food production and environmental conservation.

✉ Harison Kiplagat Kipkulei
harison.kipkulei@zalf.de

1 Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany

2 Institute of Agriculture and Horticulture, Faculty of Life Science, Humboldt University of Berlin, Berlin, Germany

3 Department of Geomatic Engineering and Geospatial Information Systems, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

4 Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

5 Institute of Geomatics, GIS & Remote Sensing (IGGReS), Dedan Kimathi University of Technology, Nyeri, Kenya

Introduction

Population growth and urbanization have pressured terrestrial landscapes, increasing land utilization to meet socioeconomic needs (Bowler et al. 2020; FAO 2017). As a result, agricultural production follows unsustainable practices that focus on enhancing the output per unit of land area. These practices may fail to achieve the intended purpose but drive the continuous impact on the environment as food production and ecosystem functions exhibit some form of interdependent relationship (Pellikka et al. 2013). The situation is even worse with the anticipation of 2.5 billion people being added to our planet by mid-century. Thus, the global demand for food will increase significantly, inducing anti-environmental effects (Tilman et al. 2011).

Globally, agricultural production demand is central to LULCCs on the Earth's surface. These changes involve transformations within and between various land uses.

The most widespread form of LULCCs relates to cropland expansion. This transformation is often accompanied by losses in forestlands, grasslands, wetlands, and other features of ecological importance (Lark et al. 2020; Zeng et al. 2018). Empirical evidence suggests that human actions are central to LULCCs (Mwaniki and Möller 2015). These changes vary across diverse spatial scales and magnitudes based on underlying biophysical and climatic conditions. Globally, cropland expansion resulting from LULCCs has been associated with the growing population, poorly formulated government action plans, environmental influences, and technological advancements (Hassan et al. 2016; Jellason et al. 2021; Kindu et al. 2015; Nakalembe et al. 2017; Pham and Smith 2014; Winkler et al. 2021).

In developing regions, cropland expansion portrays similar patterns and trends. Underlying this fact are the common challenges faced by smallholder farmers, who are the primary players in the food production chain in these regions. The challenges stem from an interplay of production factors such as land, income, market access, and prevailing climate conditions (Giller et al. 2021). Different from developed regions, unsustainable land-use practices such as charcoal burning, illegal encroachments, overgrazing, and relaxed enforcement of the law encompass the prevalent drivers of forestland, grassland, and wetland losses (Baldyga et al. 2008; Ewane 2021; Mwangi et al. 2020; Nakalembe et al. 2017). Consequently, these losses induce massive cropland conversions that have severe implications for ecosystem service provision (Song and Deng 2017), hydrological balances (Baldyga et al. 2008), and food production (Hoque et al. 2020). Therefore, understanding LULCCs and intrinsic drivers is a step towards developing tenable and coherent landscape practices that drive sustainable agricultural production (Kindu et al. 2015).

Regular and up-to-date information on land use dynamics and cropland expansion is required to formulate sound policies that foster sustainable human-environmental interactions. Moreover, information on the drivers of cropland expansion is paramount to offering precise and timely solutions to land-use decisions and regulatory measures. Remote sensing information combined with geospatial approaches provides the most feasible, cost-effective way of obtaining cropland expansion dynamics. The technology thus helps to address the issue of data limitation, especially in the data-sparse environments in developing regions. Kenya has experienced rapid conversions of natural ecosystems to croplands in the recent past (Bullock et al. 2021). The expansions challenge the ecosystems' provision capacities and expose the land to degradation, soil erosion, and biodiversity loss (Mulinge et al. 2016). The expansion in croplands is gradual in high potential agricultural production zones (Kogo et al. 2021). Consequently, it poses a threat to the sustainability of agricultural production, given that only 12% of Kenya's land mass falls under the high potential zones for production

(Kabubo and Karanja 2007). However, the effects of various drivers on cropland expansion in these zones remain uncertain, and comprehensive analysis has been lacking to date.

In recent years, various modelling approaches involving qualitative and quantitative data analysis have gained prominence in assessing drivers of cropland expansion. These approaches integrate remote sensing information and geospatial analysis that allows explicit assessments of LULCCs. Some studies, for instance, used linear and spatial regression to assess the drivers of deforestation and agricultural expansion. For example, de Espindola et al. (2021) combined satellite information and variables related to proximity, land management, technological resources, and environmental variables to assess drivers of LULCCs in the Amazon basin. Mwangi et al. (2020) combined boosted regression trees and geographically weighted regression to determine the significance and model the spatial influence of the drivers of LULCCs in Central Kenya. Nevertheless, in Kenya, Were et al. (2014) employed a logistic regression approach to uncover the drivers of LULCCs in Kenya-Afromontane forest environments. Other studies have utilized machine learning approaches such as random forest (RF) classification to determine and evaluate the importance of various drivers of LULCCs in the northeastern United States of America (Zhai et al. 2020). Other studies combined qualitative and quantitative data analysis, such as the study of Kindu et al. (2015), who evaluated drivers of LULCCs in Ethiopia. Moreover, Munthali et al. (2019) combined qualitative data analysis and geographic information systems (GIS)-based processing to assess the drivers of LULCCs in Malawi.

The reviewed studies modelled observed LULCCs changes derived through the analysis of remotely sensed imagery as a function of socioeconomic and biophysical attributes of the landscape. Subsequently, they linked the geographical distribution of land-use transitions to ancillary data to establish the significant drivers and uncover the underlying reasons for the observed patterns. Although their applications have been successful in LULCC studies, the use of evidence belief functions to assess drivers of LULCCs remains limited. Furthermore, multiapplication assessment synthesizes the inherent strengths of the individual approaches. Therefore, this study combined logistic regression (LR), boosted regression trees (BRTs), and evidence belief functions (EBFs) to assess the drivers of cropland expansion in Trans Nzoia County. Campbell et al. (2005) concluded that complexities in LULCC processes, especially their linkages with social, ecological, economic, and institutional contexts, require multiple approaches to disentangle the drivers of LULCCs.

The present study thus complements the literature in the following ways. First, three modelling techniques were applied to assess the accuracies of cropland expansions and the underlying processes. Second, the spatial prediction was conducted to depict varying probabilities of cropland

expansion across the study area. Finally, spatially modelled raster surfaces were used to enhance the definition of proximity variables by combining cost functions and linear network analysis in a geospatial environment. In this way, a more realistic measure of proximity is defined as opposed to the Euclidean and buffering approaches common in past studies (Sarkar and Chouhan 2020). Thus, this study aimed to achieve the following objectives:

1. To assess LULC changes in Trans Nzoia county and their contributions to cropland expansions.
2. To analyse the key drivers of cropland expansion using LR, BRTs, and the EBFs.
3. To assess the approaches for usability and the quality and accuracies in predicting cropland expansions at a county scale.

Materials and methods

Study area

This study was conducted in Trans Nzoia County, situated in the western part of Kenya and bordering Uganda to the west (Fig. 1). Agriculture is the main economic activity characterized by both small- and large-scale farming. Small-scale farmers cultivate crops such as maize, beans, potatoes, and sorghum, while large-scale farmers focus on producing wheat, tea, and sugarcane (Mwaura and Kenduiywo 2021). Livestock keeping, poultry rearing, fishing, and apiculture are practised for subsistence and commercial purposes. Climatically, the county exhibits a bimodal rainfall pattern. The long and short rainy seasons occur between March and May and October and December, respectively. The average annual precipitation is approximately 1300 mm, while the mean annual minimum and maximum temperatures are 12 °C and 26 °C, respectively (Nyberg et al. 2020). The county hosts Mt. Elgon and Cherangany forest ecosystems, part of Kenya's prominent water towers (Langat 2018). These ecosystems are catchments for the Nzoia and Suam rivers, which drain their waters into Lake Victoria and Turkana. The county population is approximately 990,000 people, according to the 2019 Kenya population and housing census (KNBS 2019). Trans Nzoia County was selected for this study due to its leading role as the country's central food basket. In addition, recent substantial LULCCs in the region pose a serious challenge to food security and environmental sustainability.
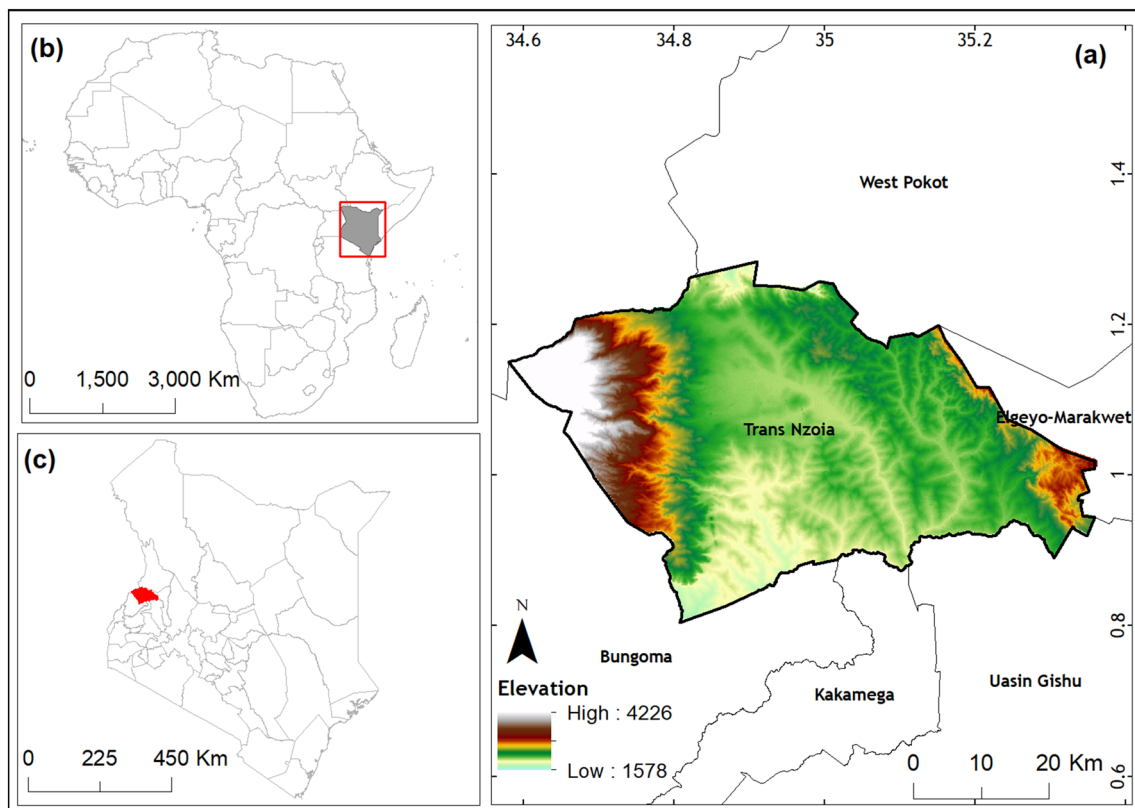

Fig. 1 Location of the study area and bordering counties (**a**), the context of Kenya in Africa (**b**), and the context of Trans Nzoia County in Kenya (**c**)

Data and data sources

This study used various datasets generated through primary and secondary data surveys, including archived remote sensing images, existing GIS databases, field observations, and discussions with land-use experts. The primary data collection was conducted between May and September 2021. The data collected during this period include ground observations used in the training and validation process of the RS image classification. In addition, land management experts from the Trans Nzoia County lands department provided information about land-based transformations and potential drivers.

The third set of data was sourced from secondary databases. The obtained variables include soil physical and chemical properties, precipitation, temperature, population density, accessibility to water sources, distance to major roads, and proximity to major trading centres. Soil information data was obtained from the International Soil Reference and Information Centre (ISRIC), https://www.isric.org/explore/isric-soil-data-hub; precipitation and temperature variables were sourced from the climatology laboratory of the University of California, https://www.climatologylab.org/terraclimate.html; The population density data were obtained from the Gridded Population of the World (GPW) Version 4 of the Socioeconomic Data and Applications Centre, https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/sets/browse; Road network data was obtained from the Humanitarian Data Exchange of the United Nations, https://data.humdata.org/dataset/kenya-roads; Rivers data was sourced from the World Resources Institute (WRI), https://datasets.wri.org/dataset/permanent-and-non-permanent-rivers-in-kenya; Market centres data was obtained from the Trans Nzoia County Department of Finance and Economic Planning, https://www.transnzoia.go.ke/. The proximity to roads, market centres, and rivers was modelled into raster surfaces using ArcGIS's cost distance functions. The detailed procedure for preparing proximity variables is outlined in the online resource of this article (ESM_1). In addition, the raster maps of all the variables used in this study are accessible from the online resource of this article (ESM_1).

The RS satellite images were acquired from different Landsat sensors, including thematic mapper (TM), enhanced thematic mapper (ETM+), and operational land imager/thermal infrared sensor (OLI/TIRS). The data was processed within the Google Earth Engine (GEE) platform, but the individual scenes are available from https://glovis.usgs.gov/. Three sets of Landsat images from 1990, 2005, and 2020 provided the data for mapping LULCCs. The spectral bands used in this study include the blue, red, NIR, and SWIR bands, with spatial resolutions of 30 m. The processed images were acquired during a relatively dry season between November and March of the succeeding year. The period allowed for the best comparison assessments across the time epochs, as the phenologies of the land features appear relatively similar. Table 1 outlines the sources, descriptions, and purposes of both the primary and secondary data used in modelling.

Image processing and LULC classification

The study used the GEE cloud computing environment to process the Landsat images and generate a time series of land cover maps for three epochs: 1990, 2005, and 2020. The platform permits large-scale data computing, thus minimising the tedious data downloading and storage requirements (Gorelick et al. 2017). Accordingly, surface reflectance data products were derived for the three epochs. The multitemporal products have already been preprocessed for radiometric and geometric corrections, and the products have also been corrected for absorbing and scattering gases and aerosol atmospheric effects. Therefore, the study used level 2 surface reflectance

Table 1 Overview of datasets used, sources, and their purpose in the study

Datasets	Where sourced	Purpose
Landsat 5 TM Landsat 7 ETM+ Landsat 8 OLI/TIRS	Processed from the Google Earth Engine	To derive LULC for the years 1990, 2005, and 2020
Climatic variables	Climatology laboratory, University of California	To derive climatic variables
Population density Rivers Road network Urban centres	SEDAC website WRI data portal Humanitarian data exchange Trans Nzoia County government	To provide the socioeconomic data
Soil	ISRIC data hub	To derive soil attributes
Reference land use and land cover data	Google Earth and In situ measurements	To provide training and validation samples for LULC classifications

WRI World Resources Institute, *SEDAC* Socioeconomic Data and Applications Centre, *ISRIC* International Soil Reference and Information Centre.

data corrected for radiometric and geometric defects. The cloud cover threshold was set at 20% to minimize the effect of clouds on the images. Any clouds present in the selected images were masked and replaced with pixels of images in the Landsat archive within 60 days of the acquisition date. The cloud score algorithm was used to mask pixels with high cloud cover based on a Landsat quality image file by computing a cloud-likelihood score from 0 (no clouds) to 100 (most cloudy). Finally, the normalised difference vegetation index, red, blue, NIR, and SWIR bands, were used as input features for the land use and land cover (LULC) classifications.

Supervised classification was performed using the random forest (RF) classifier (Breiman 2001) to map various LULC classes in the region. The algorithm is a nonparametric algorithm used for ensemble learning. It solves classification problems by estimating multiple decision trees from the training datasets and assigns probable class values to the pixels based on the maximum vote of the decision trees. The classifier is robust, achieves high accuracy, and effectively handles outliers and noisier datasets compared to other image classifiers (Belgiu and Drăguţ 2016). In this study, the number of decision trees was set at 500 to achieve a good balance between classification speed and accuracy (Belgiu and Drăguţ 2016). The default values were chosen for variablesPerSplit ($\sqrt{\text{n\_bands}}$), and the fraction of the input to bag per tree was set to 0.5.

The classification was conducted based on existing LULC classes in the study area, which include croplands, forestlands, wetlands, grasslands, built-up areas, and other lands. The other land category comprises barren lands, unclassified areas, and other exposed surfaces that do not fall in the former LULC categories. The training and validation datasets were collected from field surveys, existing topographical maps, documented historical land use plans, the local knowledge of the two authors, and visual interpretation of high-resolution imagery derived from Google Earth. The period of study determined the training and validation data used. For instance, historical information was used for the 1990 and 2005 image classifications, whereas ground survey data and updated county spatial plans were used for the 2020 image classification. The number of sample points in 1990, 2005, and 2020 was 975, 1105, and 1544, respectively. The samples were split into training samples for training the RF classifier (70%) and verification samples for accuracy verification (30%).

Accuracy assessment

Accuracy assessment is an integral part of digital image processing, as it reveals the quality and reliability of the classified images. The accuracy of the LULC maps was assessed based on a confusion matrix using an independent validation set of ground-based data. Accuracy assessment metrics, such as producer accuracy, user accuracy, and overall accuracy, were used to evaluate the overall classification process (Congalton

1991). Producer accuracy indicates how the ground features are correctly shown on the classified map. In contrast, user accuracy reveals how often a class on the classified map is depicted on the ground surface. The overall accuracy provides the percentage of correctly classified pixels for all class types. The recommendations of Olofsson et al. (2014) were adopted for accuracy assessment. The method outlines good practices for area estimation and accuracy assessment of RS image classification. In addition, it provides guidelines for proper reference sample selection and precise allocation of different class strata to achieve the desired samples.

Change analysis of land cover maps

Change analysis is a post-classification procedure that detects and quantifies changes in independently produced LULC classifications for different dates. The method provides transitions between land covers, quantifies the land cover changes, and presents information on the distribution of changes in the landscape. In this study, the analysis of changes and their distributions was used to derive binary maps of areas that were converted to croplands in the two time periods. Subsequently, they were used to assess the potential drivers of cropland expansion.

Modelling cropland expansion

Logistic regression

LR is a machine learning regression technique that assesses the relationships between dependent variables (binary or continuous) and a set of independent variables (Peng et al. 2002). LR involves logit transformation of the dependent variable. The model has the following form:

$$\text{Logit } Y = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \qquad (1)$$

$$\pi(x) = Y|X = x_1 \ldots x_n = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n}} \qquad (2)$$

where $\pi(x)$ is the probability of the outcome of interest, α is the y-intercept, β represents the regression coefficients, ε corresponds to the model error term, and x represents a set of explanatory variables. The antilog of Eq. 1 yields Eq. 2, which predicts the probability of the occurrence of the outcome of interest. The parameters α and β are estimated using the maximum likelihood (ML) method. LR is ideal for handling dichotomous outcomes and can be applied in instances of nonnormality of the dependent variable.

Cropland conversion maps were created and used to define the binary outcome. Converted areas were coded as 1, whereas nonconverted zones were coded as 0. The potential

drivers of cropland conversion included proximity to rivers, population density, soil type, precipitation, soil organic carbon, time to the nearest urban centres, time to the nearest major road, elevation, soil pH, and slope (Fig. 2). The spatial structure of the drivers was assessed using a semivariogram and interpolation conducted in ArcGIS version 10.8.1. 5,000 random points were generated in the ArcGIS environment and used as representative samples to model the relationship between cropland expansion and the potential drivers. Spatial dependency effects were minimized by maintaining a minimum distance of 200 m between each sample pair.

The selected points formed the basis for extracting the explanatory variables from the corresponding interpolated surfaces. The data were converted into ASCII format for ready import in R software. Subsequently, LR was implemented using the generalised linear model function (R Core Team 2020). 80% of the samples were used for training the model, and 20% were used for model validation. First, a full model was fitted, followed by a multicollinearity assessment among the predictors using the variance inflation factor (VIF) function in the companion to applied regression (CAR) package in R (Fox et al. 2019). Stepwise regression using the backwards selection procedure was then used to select the statistically significant variables at a 5% significance level. The regression technique fits a full model and then iteratively drops predictors with less contribution to the outcome variable. The model with the lowest Akaike information criterion (AIC) was selected as the parsimonious model for generating probability surfaces of cropland expansion.

Boosted regression tree modelling

Boosted regression tree (BRT) modelling is an ensemble model that combines regression trees and boosting algorithms to generate nonparametric statistical models (Schapire 2003). In contrast to conventional statistical models, it fits multiple statistical modes to improve prediction accuracy. The rationale is that fitting multiple trees from several approximate rules and averaging them is easier than obtaining a single highly predictive model. The strengths of the BRT model include the potential for handling missing data, accommodation of different predictor variables, and robust modelling of nonlinear interactions between variables (Elith et al. 2008). The BRT model was implemented using the generalised boosted models package in R statistical software (Ridgeway 2005). The parameters specified for the model include a bagging fraction of 0.5, as recommended by (Elith et al. 2008), a tree complexity of 5, and a learning rate of 0.005. The bagging fraction determines the split between the training and validation data, tree complexity controls whether interactions are fitted, and

the learning rate determines the contribution of each tree to the growing model. The sample points used for training and evaluating the LR were also used in the BRT modelling.

Evidential belief function

The evidence belief function (EBF) model is founded on the Dempster–Shafer theory of belief (Dempster 1968). It is a data-driven approach that computes mass functions of belief, disbelief, plausibility, and uncertainty using spatial occurrences of phenomena on the Earth's surface (Park 2011). The concept behind estimating these functions is that the locations of geographical phenomena caused by diverse earth processes can be utilized to determine the probabilities of confounding variables. Accordingly, the confounding factors are categorised into several class groups, which are then used to document the various EBF functions. The EBF model is ideally suited for assessing spatial integration processes such as LULC changes (Arasteh et al. 2019). Accordingly, an evidential map layer of the geographical phenomenon is required to compute the various functions.

Based on the functions, high belief values indicate a high likelihood of a factor contributing to an event within a class category, whereas high disbelief values indicate a lower chance. Therefore, computations of the belief and disbelief functions integrate the total number of unit cells or pixels within a class category, the number of unit cells of the evidential map layer within the class category, and the total number of unit cells in the exploration area. Equations 3–6 were used to compute the belief and disbelief values, where F_{ij} represents i confounding factors (drivers) with j class categories. $N(F_{ij})$ represents the total number of unit cells in class j, whereas $N(F_{ij} \cap A)$ is the number of unit cells in class j that were converted to cropland. $N(A)$ and $N(T)$ indicate the total number of unit cells converted to cropland and the total number of unit cells in the exploration area, respectively.

$$\text{Bel}_{F_{ij}} = \frac{W_{F_{ij(\text{Converted pixels})}}}{\sum_{j=1}^{n} W_{F_{ij(\text{Converted pixels})}}} \tag{3}$$

$$W_{F_{ij(\text{Changed pixels})}} = \frac{N(F_{ij} \cap A)/N(F_{ij})}{\left[N(A) - N(F_{ij} \cap A)/[N(T) - N(F_{ij})]\right]} \tag{4}$$

The numerator and the denominator in Eq. 4 correspond to the proportion of unit cells converted to croplands in each class factor and the ratio of unit cells converted to other land uses, respectively.

$$\text{DIS}_{F_{ij}} = \frac{W_{F_{ij(\text{Non-converted pixels})}}}{\sum_{j=1}^{n} W_{F_{ij(\text{Non-converted pixels})}}} \tag{5}$$

Fig. 2 Summary of the workflow integrating land-use changes and potential drivers of cropland expansion based on machine learning and evidence belief functions
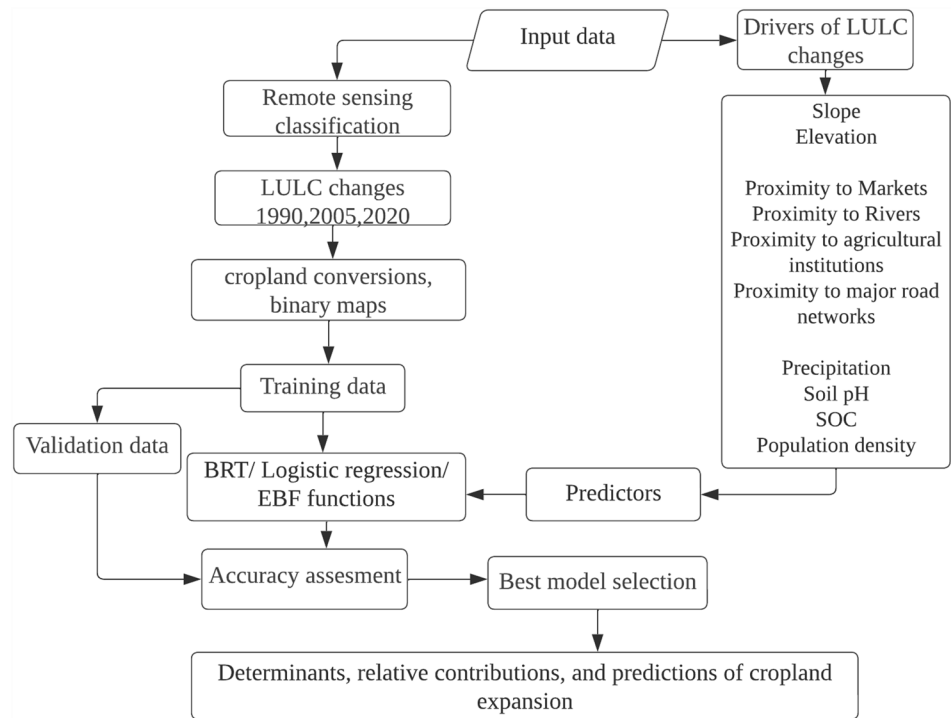
$$W_{F_{ij}(\text{Unchanged pixels})} = \frac{N(F_{ij}) - N(F_{ij} \cap A)/N(F_{ij})}{\left[N(T) - N(A) - N(F_{ij}) + N(F_{ij} \cap A)/[N(A) - N(F_{ij})] \right]} \tag{6}$$

Evaluation of the BRT, LR, and EBF models

The models were evaluated using the area under the curve (AUC) of the receiver-operating characteristic (ROC) at separate classification thresholds. The ROC curve is a plot of sensitivity against specificity (Mas et al. 2013). Sensitivity gives the proportion of the positive class that was correctly classified, while specificity indicates the ratio of the negative class that was correctly classified. The AUC ranges from 0 to 1. A perfect model yields an ROC value of 1, which indicates an exact agreement between the predicted values and the observations. ROC was implemented using the PROC package in R (Robin et al. 2011).

Results

Land use and land cover changes in Trans Nzoia County

The land cover maps for Trans Nzoia County reveal changing land use and land cover dynamics. Across the study area, the dominant land cover class in the studied epochs was cropland. The initial coverage of cropland in 1990 was 33%

of the study area, and the coverage further increased to 66% and 72% in 2005 and 2020, respectively. The changes demonstrate a rapid expansion of croplands between 1990 and 2005, followed by a slow, albeit increasing, expansion until 2020. Over the period, the area under croplands grew at the cost of forestlands, wetlands, and grasslands. Spatially, the land cover distribution indicates that croplands occupied the central areas of the county, whereas forestland and grassland classes dominated the western and northeastern parts, respectively (Fig. 3).

Built-up areas recorded positive growth over the study period. The coverage in 1990 was 2.4 km^2, which translates to approximately 0.1% of the total land area. During the period, few pockets of built-up zones were evident in the main town, located at the centre of the county. However, the coverage increased to approximately 10 km^2 by 2005, with most expansions occurring in the surrounding areas of the main town. Additionally, some sections of the western and northern parts of the county experienced a notable increase in built-up coverage. The highest coverage of built-up areas was recorded in 2020, with an approximate area of 36 km^2. In this recent period, built-up areas expanded exponentially and extended along the main transport corridors through the county. Although the built-up area class accounted for the smallest

proportion of the total land area, the findings of this study show that it experienced the highest growth in the study period.

Forestland recorded a decline in the 30 years. In 1990, the area under forest cover was approximately 500 km^2. However, the coverage was reduced to 447 km^2 and 335 km^2 in 2005 and 2020, respectively. The decline in forest cover was higher (− 25%) in the 2005–2020 period than in the 1990–2005 period (− 11%). The reduction was more pronounced in the forested areas of Mt. Elgon and Cherangany hills. The land cover maps further indicate that tree cover vegetation along river courses and creeks was drastically removed. Similarly, planted forest patches were gradually cleared. The effect is evident from the 2020 LULC map (Fig. 3c). Likewise, wetlands exhibited a continuous decline in the 30 years. The study area's wetlands comprise permanent streams, open water, seasonal and permanent marshes, riverine vegetation, scrub, forested wetlands, and seasonal flood plains. In 1990, wetlands occupied 270 km^2. However, the coverage declined to 230 km^2 and 73 km^2 in 2005 and 2020, respectively. The decline was higher in the 2005–2020 period (− 68.2%) than in the 1990–2005 period (− 15%).

Grasslands covered 19% of the total area in 1990 and declined by 16% in 2005. In 2020, there was a slight increase of 1% in grassland cover. Grasslands dominated the western region of the county and the periphery of the Cherangany hills forest in 1990. The LULC dynamics show that extensive grassland areas were rapidly transformed into croplands in the 1990–2005 period. Approximately 72% of grassland cover was converted to croplands during this period (Fig. 3a,b).

Accuracy assessment

The accuracy assessment statistics of the LULC classifications are presented in Table 2. The overall accuracies for 1990, 2005, and 2020 are 82%, 93%, and 93%, respectively. The producer accuracies were above 70%, except for the built-up area in 2020, which was 53%. The low accuracy in the built-up area classification can be attributed to the mixed-pixel problem common in urban areas, as they rarely transition to other land covers.
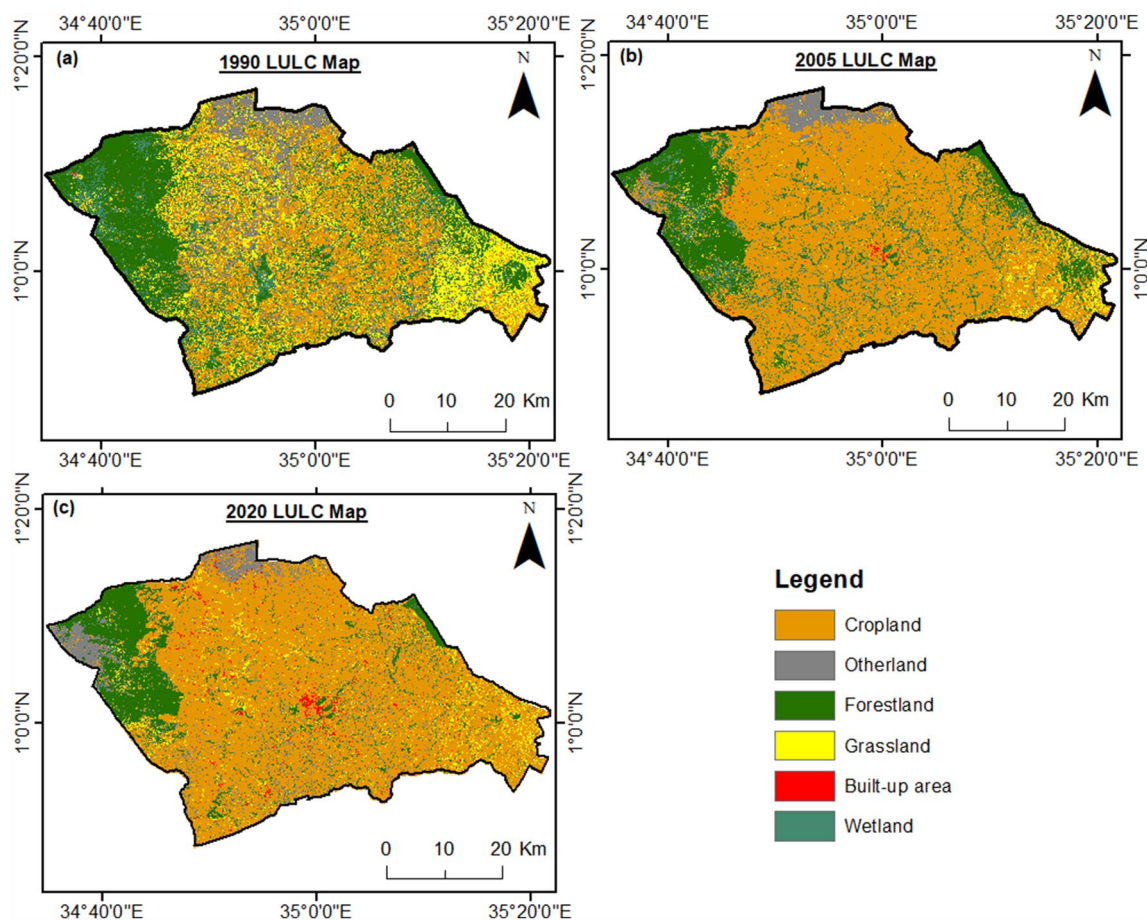

Fig. 3 LULC maps for **a** 1990, **b** 2005, and **c** 2020

Change analysis of land cover maps

The Sankey plot (Fig. 4) characterizes the distribution of the major LULC transitions in the 30 years. A noticeable trend from the plot is that most of the LULC transitions were directed towards croplands, visible from the width of the links connecting each land cover node. From the plot, 44% of wetland conversion between 1990 and 2005 occurred at the expense of cropland expansion, and the expansion increased to 62% in the 2005–2020 period.

The dynamics of the grassland cover show that 72% of grassland coverage converted to croplands between 1990 and 2005. However, the conversion declined to 62% between 2005 and 2020. On the same note, the forestland class lost 26% of its coverage to croplands between 1990 and 2005. otherlands category, comprising barren land, artificial surfaces, pockets of water features, and other unclassified features, lost 67.9% and 32.4% to cropland in the 1990–005 and 2005–2020 periods, respectively. Overall, croplands gained significantly from the major land covers in the study area.

Drivers of cropland expansion

Logistic regression

The logistic regression results provide the relative influences of the individual drivers on the cropland expansion in the study area. Multicollinearity assessment revealed low Pearson correlation values among the drivers except for access to major roads and market centres (0.98). Accessibility to major roads, therefore, was dropped from any further analysis. The retained drivers had low VIF values (< 2), thus confirming the absence of multicollinearity. The drivers were further analysed using stepwise regression to quantify their effect on cropland expansion. The aim was to obtain a parsimonious model with statistically significant drivers that explain cropland expansion (Table 3). The drivers' estimates were transformed into percentage odds to assess the relative contribution of each driver to cropland expansion.

The findings show that the contribution of elevation was positive and revealed a increase of 2.1 percent odds in cropland expansion for every 1 unit increase in elevation while controlling for other variables (Table 3). Similarly, the

Table 2 Overall, producer and user accuracies of the LULC classifications in 1990, 2005, and 2020

LULC	1990		2005		2020	
	PA	UA	PA	UA	PA	UA
Cropland	0.94	0.95	0.95	0.95	0.98	0.94
Forestland	0.91	0.97	0.94	0.91	0.92	0.99
Grassland	0.79	0.63	0.80	0.73	0.76	0.89
Built-up	0.86	0.97	0.86	0.79	0.53	0.92
Wetland	0.87	0.81	0.73	0.8	0.89	0.77
Other lands	0.91	0.92	0.92	0.97	0.66	0.96
Overall Accuracy	0.82		0.93		0.93	

PA producer accuracy, *UA* user accuracy

Fig. 4 Sankey plot showing land cover transitions in the 1990–2005 and 2005–2020 epochs
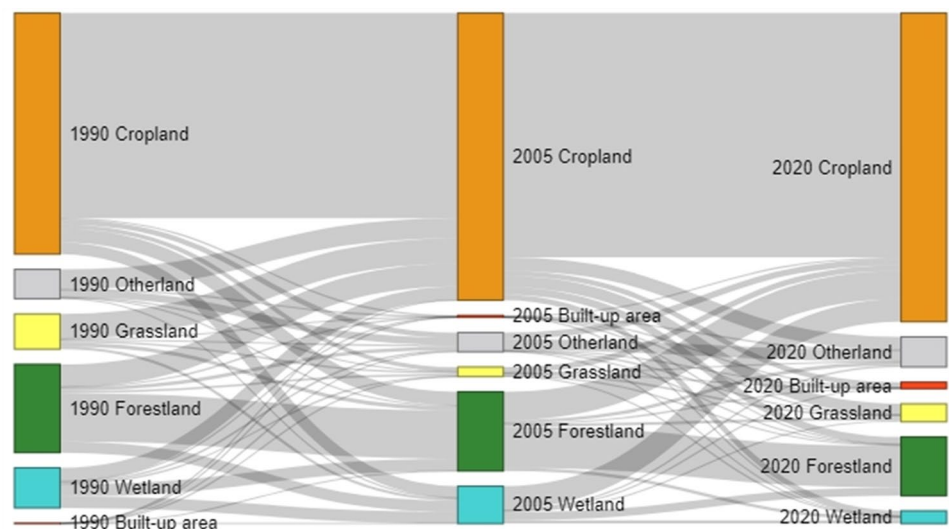
contribution of accessibility to market centres was also positive, and it showed that a unit increase from market centres increases the odds of cropland expansion by 0.7%. The result confirmed that croplands expand farther from market centres into rural setups, which is typical in Kenyan landscape environments. Regarding soil drivers, soil pH was the only variable that showed statistical significance and a large marginal effect among the drivers. Accordingly, a unit increase in soil pH increased the odds of cropland expansion by 25%. Accessibility to water sources depicted a negative trend where a unit increase in proximity to water sources decreased the odds of cropland expansion by 1.27%.

Boosted regression trees

Partial dependence plots (PDPs) summarised the results obtained from the BRT model (Fig. 5). The plots model each driver's relationship to cropland expansion while controlling for other factors. The probability of cropland expansion is indicated on the *y*-axis, whereas the data distribution is plotted on the *x*-axis. The assessed drivers demonstrated varied influence on cropland expansion based on the data range on the *x*-axis. For instance, areas with low slopes (< 10%) depicted a high likelihood of cropland expansion, with declined probability as the gradient increased. PDP for elevation revealed regions that range from 1750 to 2000 m to have a high likelihood of cropland expansion.

Soil properties indicated that regions characterised by low soil organic carbon (SOC) had high tendencies to be converted to croplands. However, areas with high SOC values were less likely to be converted to croplands. The possible explanation is that marginal areas are increasingly experiencing exploitation for agricultural production. Regarding soil pH, regions with values that range between 5.5 and 6.0 showed a high likelihood of expansion, which declined in the neutral and less acidic zones. Although the relative contributions of population density and proximity variables were low, the results showed a significant trend in their factor ranges. Notably, areas within 1-hour access of rivers indicated a high probability of cropland expansion. The plot showed a high likelihood of cropland expansion for population density in low-populated regions (0–900 people per square kilometre).

Evidence belief functions

The EBF functions of cropland expansions are presented in Table 4. The drivers showed high belief values, thus supporting strong evidence of cropland expansion. They comprise proximity to rivers, proximity to market centres, SOC, elevation, population density, soil pH, and precipitation.

The EBF model indicated that regions within 1 h of proximity to water sources have a high likelihood of cropland expansion, as revealed by a high belief value (0.342). Similarly, areas with a population density between 240 and 440 people per square kilometre showed high probabilities of cropland expansion (Bel = 0.31). Low SOC (Bel = 0.423) and high soil pH zones (Bel = 0.42) revealed a high likelihood of conversion. For access to market centres, areas within 2 hours of proximity showed a high probability of experiencing cropland expansion (Bel = 0.281). The results further indicated that low precipitation zones were more likely to be converted to croplands (Bel = 0.368). The high belief values in low SOC and low precipitation zones imply that cropland expansions also target marginal zones. Cropland expansions were also prevalent in elevation ranges between 2000 and 2800 m and between 2400 and 2800 m, as indicated by high belief values and frequency ratio scores.

Performance of the models

The model performances were evaluated based on the receiver operating characteristic (ROC) curves (Fig. 6). The AUC values ranged from 0.77 to 0.96, with the LR model showing excellent performance and the BRT model achieving the lowest accuracy. The AUC value obtained using the EBF model was moderate (0.86). Nonetheless, the obtained values showed good to excellent performances of cropland expansion assessment in Trans Nzoia County.

Probability of cropland expansion

Cropland expansion probability maps (Fig. 7) were generated to visualize the spatial patterns and to establish the drivers' contributions across the study area. The models agreed well in the characterization of the cropland expansion patterns. The predicted surfaces revealed that the western parts

Table 3 Logistic regression model estimates and percentage odds of cropland expansion drivers

	Estimate	Standard error	*z* value	Percentage odds
(Intercept)	− 5.429***	0.965	− 5.627	− 99.56
Rivers	− 0.013***	0.004	− 3.499	− 1.27
Soil pH	0.228**	0.112	2.032	25.55
Access to Markets	0.007***	0.002	2.902	0.70
Elevation	0.002***	0.001	5.194	0.21

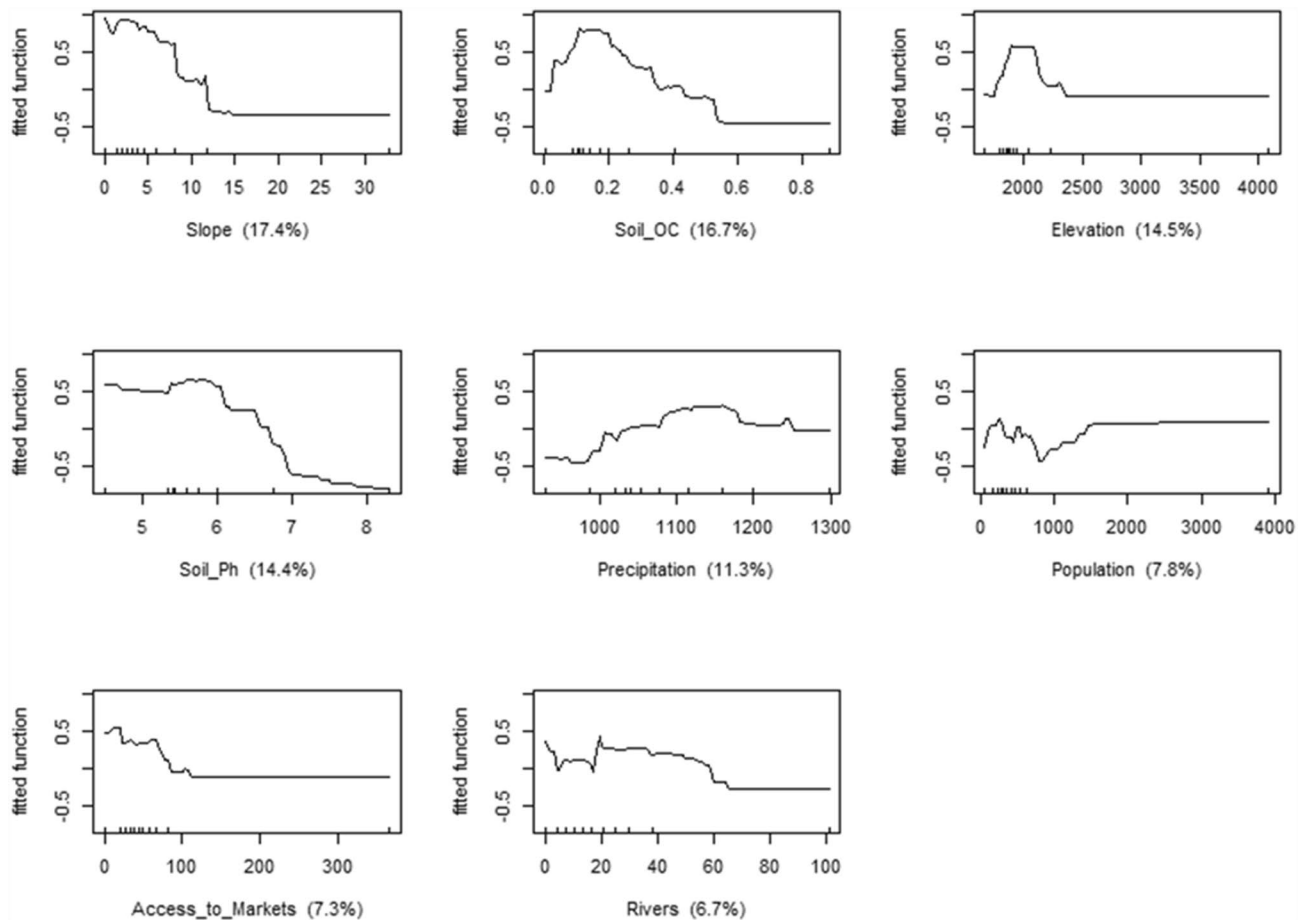***, **, *Denote significance at the 0.01, 0.05, and 0.1 levels of probability, respectively

Fig. 5 Partial dependence plots indicating marginal effects of the drivers on cropland expansion. The *x*-axis shows the data distribution, and the *y*-axis indicates the probability of cropland expansion

bordering Mt. Elgon forest and the Cherangany hills ecosystems are at a high risk of cropland expansion. The BRT and EBF models revealed more similar expansion patterns than the LR model. Nonetheless, all models showed a high probability of cropland expansion in the western parts, which declined towards the central regions and increased in the western region.

The LR model showed a higher likelihood of cropland expansion in the western and eastern parts than in the southern region. Additionally, few areas in the north revealed a high probability of cropland expansion. The predictions showed a close fit between the different approaches, thus demonstrating the robustness of the assessed drivers in characterizing cropland expansion in the region. Blending different techniques provides an array of statistical measures that help to understand the magnitude, nature, and direction of cropland expansions. For instance, LR and BRT provided relative contributions of various factors, whereas BRT and EBF provided the factors' range of influence on cropland expansion.

Discussion

LULC classification and accuracies

In this study, a detailed evaluation of long-term LULCCs (1990–2020) was conducted in Trans Nzoia County. Six dominant land covers were mapped to assess their spatial coverage in the 30 years. The overall classification accuracies were above 82%, with a high classification accuracy of 93% in the 2005 and 2020 classifications. Similarly, the user and producer accuracies strongly agreed between the mapped classes and the reference data. One exception was in the classification of built-up areas, where the producer accuracy was 53%. A possible explanation for this observation is that the classifier may have misinterpreted urban areas because of mixed pixels. The spatial coverage of a single Landsat pixel used in the classification process was 30 m. However, buildings and surfaces in the region have less coverage, resulting in mixed land use classes. The problem that always results in spectral confusion is dominant in characterising urban footprints (Forget et al. 2018).

Table 4 Belief, disbelief, and uncertainty values for different drivers of cropland expansion

Spatial layers	Class	Number of class pixels	Pixels converted croplands	Frequency ratio	EBF Functions		
					Bel	Dis	Unc
Proximity to rivers	0–30	2,088,315	243,269	1.020	0.342	0.155	0.502
	30–60	561,271	68,146	1.063	0.341	0.155	0.504
	60–90	68,726	4949	0.630	0.197	0.165	0.638
	90–120	26,881	634	0.206	0.065	0.173	0.762
	120–150	18,697	50	0.023	0.007	0.177	0.816
	>150	13,221	230	0.152	0.048	0.174	0.778
Population density	58–240	737,552	73,316	0.870	0.180	0.202	0.618
	241–440	1,047,579	147,348	1.231	0.310	0.188	0.502
	441–910	901,684	88,889	0.863	0.175	0.203	0.622
	911–2100	69,103	5717	0.724	0.156	0.205	0.640
	2101–3921	21,193	2008	0.829	0.179	0.202	0.619
Precipitation	920–1000	280,371	49,948	1.559	0.368	0.182	0.450
	1001–1100	1,450,326	157,473	0.950	0.199	0.200	0.601
	1101–1200	579,328	68,445	1.034	0.231	0.196	0.573
	1201–1300	448,189	41,110	0.803	0.171	0.203	0.626
	>1300	18,897	302	0.140	0.031	0.219	0.750
Soil organic carbon	0–0.87	226,380	50,303	1.945	0.423	0.223	0.354
	0.87–1.5	2,002,234	208,702	0.912	0.148	0.268	0.584
	1.5–3.96	460,655	45,226	0.859	0.167	0.263	0.571
	>3.96	87,842	13,047	1.300	0.262	0.247	0.492
Proximity to market centres	0–30	1,354,817	141,233	0.912	0.121	0.173	0.706
	30–60	881,317	113,542	1.128	0.172	0.166	0.663
	60–90	196,842	37,519	1.668	0.252	0.154	0.594
	90–120	63,845	14,001	1.919	0.281	0.149	0.570
	120–150	50,280	5955	1.037	0.149	0.169	0.682
	>150	230,010	5028	0.191	0.026	0.189	0.785
Soil PH	0–5.4	485,422	46,715	0.842	0.162	0.264	0.575
	5.4–5.7	1,919,182	198,996	0.908	0.149	0.267	0.583
	5.7–7.1	159,356	24,171	1.328	0.269	0.246	0.485
	7.1–8.3	213,151	47,396	1.946	0.420	0.223	0.357
Elevation	1578–2000	2,064,586	225,351	0.971	0.241	0.196	0.566
	2000–2400	382,067	61,448	1.430	0.412	0.181	0.406
	2400–2800	181,400	24,763	1.214	0.331	0.188	0.480
	2800–3300	101,752	716	0.063	0.016	0.217	0.767
	3300–4226	47,306	0	0.000	0	0.218	1

LULC changes and cropland expansion

The LULCCs show that the study area experienced losses in forestland, grassland and wetland land covers. Conversely, built-up land and croplands increased in the same period, with gains from wetlands, forestlands, and grasslands. Similar cropland expansion trends and intra-class transitions have been observed in other studies conducted in western Kenya (Becker et al. 2016; Masayi et al. 2021; Rotich and Ojwang 2021). In the present study, the area under crops increased from 83,132 ha in 1990 to 166,420 ha in 2005 and then further increased to 180,222 ha in 2020. The observed

expansion may be attributed to market forces, government extension initiatives, and population growth in the study region. The demographic assessment shows that in the 1990s, Kenya witnessed a paradigm shift in the agricultural sector, brought about by market liberalization and increased access to credit services (De Groote et al. 2006). These initiatives fostered rapid agricultural development, especially in high-potential regions such as Trans Nzoia County.

Population growth also contributed substantially to cropland expansion in the county. According to Kenya's population and housing census statistics, the population of Trans Nzoia County in 1989 was 393,682 people (GOK 1994).
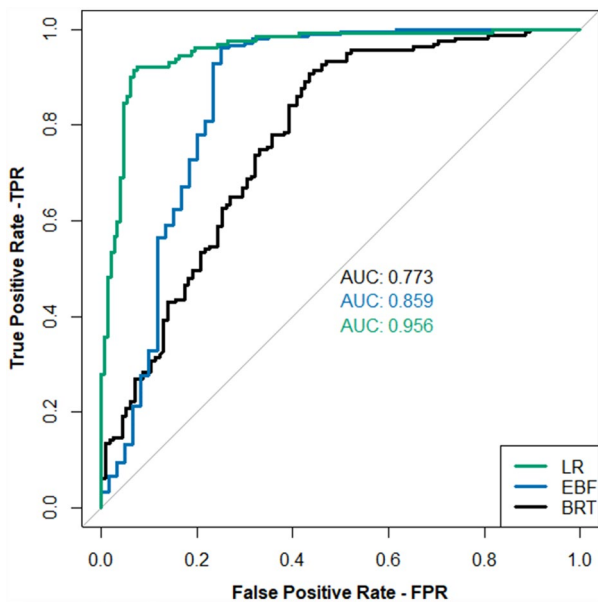
Fig. 6 ROC curves and AUC values showing the accuracies of the LR, BRT, and EBF models

The population rose by 44% between 1989 and 1999 and increased to 818,000 by 2009 (GOK 2010). The observed growth may have induced a direct effect on cropland expansion due to the rising demand for land for settlement and food production purposes. KEFRI (2017) reported that human activities have increased in the area due to the rising population in Mt. Elgon and Cherangany forest hills and their borders. Our results also highlight massive losses of forested areas, wetlands, and grasslands bordering these towers. Population growth, therefore, played a primary role in the loss of forestland, wetlands, and grassland cover. Jayne and Muyanga (2012) also noted that Kenya's western region is densely populated.

The demographic dynamics in the region affected existing farming systems and management of land, encouraging unsustainable environmental practices such as deforestation, poor soil management, and encroachment of natural ecosystems (Allaway and Cox 1989). Ongugo et al. (2014) observed that politically motivated excisions in the Mt. Elgon forest ecosystem led to massive destruction of forests for settlement purposes. As a result, the beneficiaries took
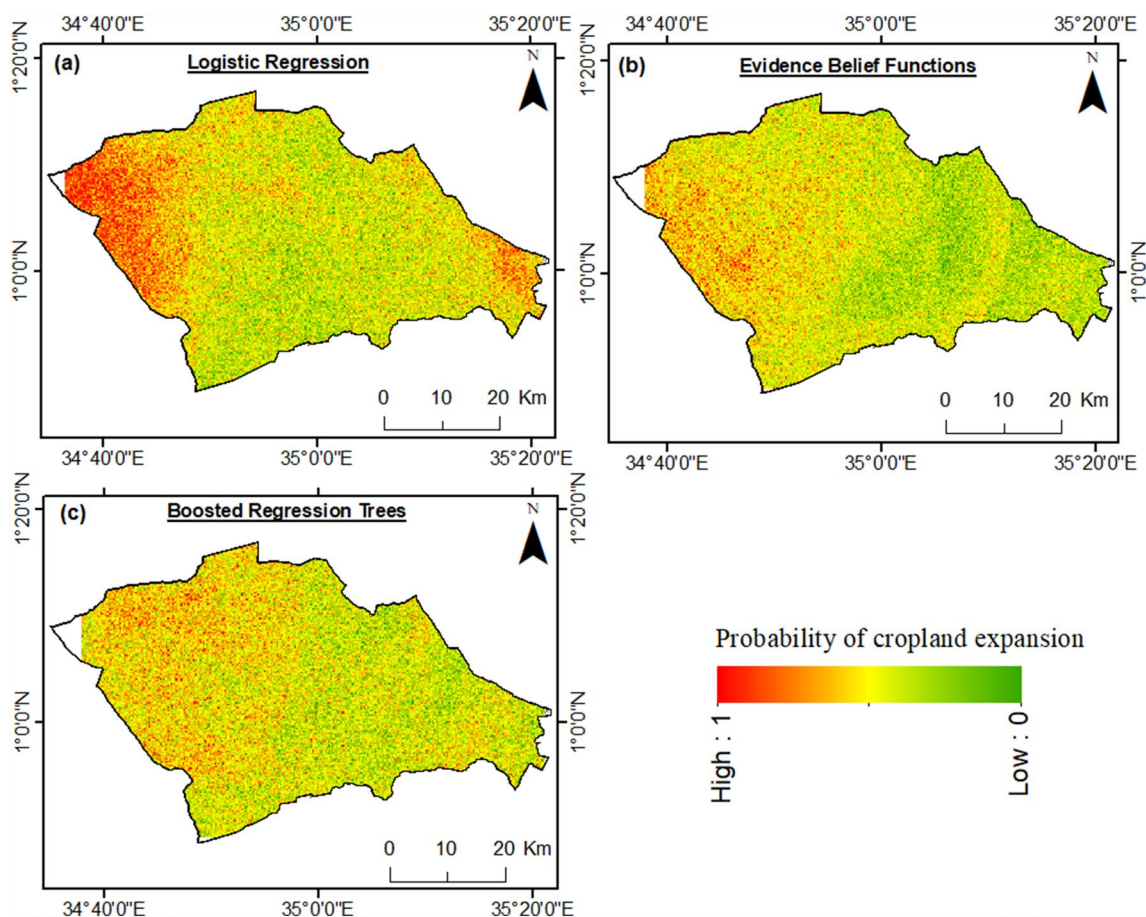
Fig. 7 Surfaces showing the probability of cropland expansion based on **a** logistic regression, **b** evidence belief functions, and **c** boosted regression trees

advantage of the scheme to clear extra land for agricultural use. Masayi et al. (2021) also noted intensive mixed farming in the Mt. Elgon forest ecosystem resulted in massive forest and biodiversity losses. The LULCCs dynamics and demographic influences corroborate previous studies conducted in western Kenya (Kogo et al. 2021; Mutoko et al. 2014).

Spatially, buffer zones and foothills of Mt. Elgon and Cherangany Hills forest ecosystems experienced rapid cropland expansion over the study period. Further expansions were prominent along the major river channels of the Nzoia and Sabwani Rivers. The finding agrees with the national environmental management authority (NEMA) district environmental action plan of Trans Nzoia county, which documented encroachments on protected areas by the locals to boost their household food production (GOK 2019). Similarly, Maua et al. (2022) and Ondiek et al. (2020) observed that the rich wetland resources of the Nzoia and Lake Victoria drainage systems experienced rampant anthropogenic exploitation within the period under our study. Studies in Kenya have also attributed cropland expansion to agricultural extensification practices (Eckert et al. 2017; Mwangi et al. 2018). At the SSA scale, cropland expansion in the late twentieth century resulted from rising economic activities, improved technology, declining soil fertility, and climate change variability (Jellason et al. 2021).

Cropland expansion and drivers

Cropland expansion in Trans Nzoia County was analysed using three modelling techniques, BRT, LR and EBF. The results revealed some consistencies in drivers' influences and cropland expansion prediction. The three models revealed factors such as proximity to rivers, elevation, soil pH, and market accessibility as crucial factors of cropland expansion in the region. Furthermore, the BRT and EBF functions found that slope, SOC, and population density significantly affected cropland expansion ($p < 0.05$). On the other hand, LR associated elevation and proximity to rivers with cropland expansion. Regular physical and chemical analyses of water quality parameters in the Nzoia Basin confirmed high levels of phosphates and nitrates in water sources caused by intensive agricultural activities (KEFRI 2018; Twesigye et al. 2011). Likewise, Enanga et al. (2011) found human activities to be the fundamental causes of increased pollution in riparian buffer strips of hydrological watersheds in Kenya. The rising contamination can be attributed to intensive agricultural activities close to key water features.

The BRT model revealed a high likelihood of cropland expansions in areas of high soil acidity. According to Hijbeek et al. (2021), soils in western Kenya are predominantly acidic, which might have resulted in the more acidic samples used in the BRT model training. The EBF model uses class categories; thus, the observations might have

captured the dynamics in the alkaline zones. In other studies in Kenya, for example, Were et al. (2014) found that soil pH was a key driver of LULCCs in the Eastern Mau forest reserve. Regarding population density, a high likelihood of expansion was reflected by the EBF model, whereas the BRT model revealed a medium influence. Accordingly, areas with a population density ranging from 240 to 440 people per square kilometre demonstrated a high likelihood of expansion based on the EBF model. One possible explanation is that areas with low population densities have more space to be utilized for agricultural use than highly populated areas.

Other studies within the Kenyan context pointed out some of the drivers identified in this study as significant to cropland expansion. For instance, Serneels and Lambin (2001) noted accessibility to markets and agroclimatic factors as critical drivers in the Narok district. In addition, Mwangi et al. (2020) reported population, proximity to rivers, and proximity to roads as crucial drivers of cropland conversions in Central Kenya. Moreover, Were et al. (2014) found soil pH, population density, precipitation, distance to towns, and rivers to be the significant drivers of LULCCs in the Eastern Mau forest reserve. A study in the Eastern Arc Mountains of Taita in coastal Kenya found strong associations between proximity variables and woodland-cropland conversions (Maeda 2011). Another study in the agro-pastoral regions of the Kajiado district found that changing preferences from herding to crop production and population density were the prominent drivers of cropland expansion in the Kajiado district (Campbell et al. 2005).

Comparable trends in cropland expansions are also noted in studies conducted in East Africa and the larger SSA region. Kindu et al. (2015) and Betru et al. (2019) found drivers of LULCCs in Ethiopia to fall within the broader categories of social, economic, environmental, policy, demographic, and technological forces. According to the researchers, some fundamental drivers included population density, livestock ranching, climate change, accessibility to markets, and accessibility to major road networks. In addition, Msoffe et al. (2011) found population and subsistence-driven agriculture to be the main drivers of cropland expansion in Northern Tanzania. A review of studies conducted in Uganda by Kilama Luwa et al. (2021) similarly noted population density as the primary driver of the observed LULCCs in the region. In Malawi, Li et al. (2021) and Munthali et al. (2019) found elevation, proximity to water sources, population, and human activities to be significant drivers of LULCCs.

Similar patterns of LULCCs and the associated drivers have also been found in studies conducted outside of Africa. For instance, Duraisamy et al. (2018) found access to water sources and improved road networks to be India's main drivers of LULCCs. In another study by Zaveri et al. (2020), dry rainfall anomalies contributed to cropland expansion

in other developing regions outside the African context. Similar trends in cropland expansion have also been witnessed in developed countries. For example, a study in the United States of America shows that cropland expansions marginally targeted agricultural zones (Lark et al. 2020). In the European Union, however, a study by Kuemmerle et al. (2016) showed mixed findings with declines and hotspots of cropland expansions based on the region. There is a clear divide in the factors driving cropland dynamics between developing and developed areas. While the developed regions appear to have well-regulated policies that guide land use and management, less developed regions still face challenges in enforcing land management policies. Poor land tenure practices and systems for controlling land management are considered barriers to effective land management in these regions. Similarly, inequalities in land ownership promote illegal land-use activities that target land expansion (Keijiro and Frank 2014; Schürmann et al. 2020).

Overall, our study confirms that, similar to other global regions, drivers of cropland expansions in Trans Nzoia County, Kenya, are multifaceted, cutting across socio-economic, climatic, and proximity factors. The modelling approaches used for assessment demonstrated an agreement on the fundamental aspects of cropland expansions and their spatial prediction. At the same time, it provides an understanding of the drivers of changes in a spatially explicit manner.

Conclusion

The present study investigated cropland expansion and its drivers, driven by LULCCs in Trans Nzoia County, Kenya. The study found that the county experienced rapid land use transformation in the past three decades, resulting in the loss of forests, wetlands, and grasslands. Consequently, cropland areas increased at the expense of these land cover types. The study noted the expansion to have resulted from an interplay of socioeconomic, climatic, and soil drivers. LR, BRTs, and EBFs models assessed the relative contribution of the drivers and generated spatial prediction surfaces for cropland expansion.

Based on the modelling results, we conclude that elevation, soil pH, and proximity to rivers are critical drivers of cropland expansion in the region. In particular, soil pH has a significant influence based on the relative contribution assessment from the LR and BRT models and the belief values from the EBF model. We also conclude that climatic, soil, and biophysical drivers influence cropland expansion across landscapes. Logistic regression showed a better performance in characterising the drivers of cropland expansion in the region. The application of the model and incorporating soil, proximity, and topographical variables is recommended for credible cropland expansion modelling in the future.

Therefore, the study provides insights into target areas for sustainable land management and conservation of the natural environment. Counties and national governments should integrate the drivers of LULCCs into their routine resource planning to foster harmony between food production and environmental protection. Such approaches are currently scarce, and the findings of this study create a solution pathway, especially in light of the rapidly growing population, urbanization, and increased human pressure on natural resources.

Author contributions HK contributed to the formal analysis, methodology and writing of the first draft. SS, BK, ML, and MB supervised, reviewed and edited the manuscript. GG reviewed and edited the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors declare that they have no competing interests. Any suggestions, results, discussions and conclusions or recommendations presented in this study are those of the authors and do not express the views of the funding agencies.

References

Allaway J, Cox PMJ (1989) Forests and competing land uses in Kenya. Environ Manag 13:171–187

Arasteh R, Ali Abbaspour R, Salmanmahiny A (2019) Non-path dependent urban growth potential mapping using a data-driven evidential belief function. Environ Plan B Urban Anal City Sci 48:555–573. https://doi.org/10.1177/2399808319880219

Baldyga TJ, Miller SN, Driese KL, Gichaba CM (2008) Assessing land cover change in Kenya's Mau Forest region using remotely sensed data. Afr J Ecol 46:46–54. https://doi.org/10.1111/j.1365-2028.2007.00806.x

Becker M, Alvarez M, Heller G, Leparmarai P, Maina D, Malombe I, Bollig M, Vehrs H (2016) Land-use changes and the invasion dynamics of shrubs in Baringo. J East Afr Stud 10:111–129. https://doi.org/10.1080/17531055.2016.1138664

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: a review of applications and future directions. ISPRS J Photogramm Remote Sens 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Betru T, Tolera M, Sahle K, Kassa H (2019) Trends and drivers of land use/land cover change in Western Ethiopia. Appl Geogr 104:83–93. https://doi.org/10.1016/j.apgeog.2019.02.007

Bowler DE, Bjorkman AD, Dornelas M, Myers-Smith IH, Navarro LM, Niamir A, Supp SR, Waldock C, Winter M, Vellend M, Blowes SA, Böhning-Gaese K, Bruelheide H, Elahi R, Antão LH, Hines J, Isbell F, Jones HP, Magurran AE, Cabral JS, Bates AE (2020) Mapping human pressures on biodiversity across the planet uncovers anthropogenic threat complexes. People Nat 2:380–394. https://doi.org/10.1002/pan3.10071

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Bullock EL, Healey SP, Yang Z, Oduor P, Gorelick N, Omondi S, Ouko E, Cohen WB (2021) Three decades of land cover change in East Africa. Land. https://doi.org/10.3390/land10020150

Campbell DJ, Lusch DP, Smucker TA, Wangui EE (2005) Multiple methods in the study of driving forces of land use and land cover change: a case study of SE Kajiado District, Kenya. Hum Ecol 33:763–794. https://doi.org/10.1007/s10745-005-8210-y

Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37:35–46. https://doi.org/10.1016/0034-4257(91)90048-B

de Espindola GM, de Silva FE, Picanço Júnior P, dos Reis Filho AA (2021) Cropland expansion as a driver of land-use change: the case of Cerrado-Caatinga transition zone in Brazil. Environ Dev Sustain 23:17146–17160. https://doi.org/10.1007/s10668-021-01387-z

De Groote H, Kimenju S, Owuor G, Wanyama J (2006) Market Liberalization and Agricultural Intensification in Kenya (1992–2002). In: International Association of Agricultural Economists, 2006 Annual Meeting, August 12–18, 2006, Queensland

Dempster AP (1968) A generalization of Bayesian inference. J R Stat Soc Ser B (methodological) 30:205–247

Duraisamy V, Bendapudi R, Jadhav A (2018) Identifying hotspots in land use land cover change and the drivers in a semi-arid region of India. Environ Monit Assess 190:535. https://doi.org/10.1007/s10661-018-6919-5

Eckert S, Kiteme B, Njuguna E, Zaehringer JG (2017) Agricultural expansion and intensification in the foothills of Mount Kenya: a landscape perspective. Remote Sens. https://doi.org/10.3390/rs9080784

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Enanga E, Shivoga W, Gichaba M, Creed I (2011) Observing changes in riparian buffer strip soil properties related to land use activities in the River Njoro Watershed, Kenya. Water Air Soil Pollut 218:587–601. https://doi.org/10.1007/s11270-010-0670-z

Ewane EB (2021) Land use land cover change and the resilience of social-ecological systems in a sub-region in South west Cameroon. Environ Monit Assess 193:338. https://doi.org/10.1007/s10661-021-09077-z

FAO (2017) The future of food and agriculture—trends and challenges. http://www.fao.org/3/i6583e/i6583e.pdf

Forget Y, Linard C, Gilbert M (2018) Supervised classification of built-up areas in sub-saharan African cities using landsat imagery and OpenStreetMap. Remote Sens. https://doi.org/10.3390/rs10071145

Fox J, Weisberg S, An R (2019) Companion to applied regression. In: Salmon H, Heffernan M, DeRosa K and Dickens G (eds) 3rd edn. Sage, Thousand Oaks, pp 593–599

Giller KE, Delaune T, Silva JV, van Wijk M, Hammond J, Descheemaeker K, van de Ven G, Schut AGT, Taulya G, Chikowo R, Andersson JA (2021) Small farms and development in sub-Saharan Africa: farming for food, for income or for lack of better options? Food Secur 13:1431–1454. https://doi.org/10.1007/s12571-021-01209-0

GOK (1994) Kenya population census, 1989, Vol 1. https://www.knbs.or.ke/?wpdmpro=kenya-population-census-1989-vol-1-march-1994 Accessed 27 June 2022

GOK (2010) 2009 Kenya population and housing census: volume 1A population distribution by administrative units. https://www.knbs.or.ke/2009-kenya-population-and-housing-census-volume-1a-population-distribution-by-administrative-units/ Accessed 27 June 2022

GOK (2019) Trans Nzoia district environment action plan (2009–2013). https://www.nema.go.ke/images/Docs/Awarness%20Materials/NEAPS/trans%20nzoia.pdf Accessed 10 December 2021

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

Hassan Z, Shabbir R, Ahmad SS, Malik AH, Aziz N, Butt A, Erum S (2016) Dynamics of land use and land cover change (LULCC) using geospatial techniques: a case study of Islamabad Pakistan. Springerplus 5:812. https://doi.org/10.1186/s40064-016-2414-z

Hijbeek R, van Loon MP, Ouaret W, Boekelo B, van Ittersum MK (2021) Liming agricultural soils in Western Kenya: can long-term economic and environmental benefits pay off short term investments? Agric Syst 190:103095. https://doi.org/10.1016/j.agsy.2021.103095

Hoque MZ, Cui S, Islam I, Xu L, Tang J (2020) Future impact of land use/land cover changes on ecosystem services in the Lower Meghna River Estuary, Bangladesh. Sustainability. https://doi.org/10.3390/su12052112

Jayne TS, Muyanga M (2012) Land constraints in Kenya's densely populated rural areas: implications for food policy and institutional

reform. Food Secur 4:399–421. https://doi.org/10.1007/s12571-012-0174-3

Jellason NP, Robinson EJZ, Chapman ASA, Neina D, Devenish AJM, Po JYT, Adolph B (2021) A systematic review of drivers and constraints on agricultural expansion in Sub-Saharan Africa. Land. https://doi.org/10.3390/land10030332

Kabubo JM, Karanja FK (2007) The economic impact of climate change on Kenyan crop agriculture: a Ricardian approach. Glob Planet Change 57:319–330. https://doi.org/10.1016/j.gloplacha.2007.01.002

KEFRI (2017) Land tenure profiles in 'hotspots' and vulnerable areas on public and community lands in Mt. Elgon and Cherangany hills ecosystems. https://kefri.org/WaterTowers/PDF/KEFRI_EMC_Draft%20%20Report_August_2017_Land%20Tenure_kefri_comments.pdf

KEFRI (2018) Water quality in Mt. Elgon and Cherangany Forest Ecosystems/Catchment. http://197.248.75.118:8282/jspui/handle/123456789/883 Accessed 17 Jan 2022

Keijiro O, Frank P (2014) Changes in land tenure and agricultural intensification in Sub-Saharan Africa, vol 2014. UNU-WIDER, Helsinki

Kilama Luwa J, Bamutaze Y, Majaliwa Mwanjalolo J-G, Waiswa D, Pilesjö P, Mukengere EB (2021) Impacts of land use and land cover change in response to different driving forces in Uganda: evidence from a review. Af Geogr Rev 40:378–394. https://doi.org/10.1080/19376812.2020.1832547

Kindu M, Schneider T, Teketay D, Knoke T (2015) Drivers of land use/land cover changes in Munessa-Shashemene landscape of the south-central highlands of Ethiopia. Environ Monit Assess 187:452. https://doi.org/10.1007/s10661-015-4671-7

KNBS (2019) 2019 Kenya Population and Housing Census. http://www.knbs.or.ke/. Accessed 02 Nov 2021

Kogo BK, Kumar L, Koech R (2021) Analysis of spatio-temporal dynamics of land use and cover changes in Western Kenya. Geocarto Int 36:376–391. https://doi.org/10.1080/10106049.2019.1608594

Kuemmerle T, Levers C, Erb K, Estel S, Jepsen MR, Müller D, Plutzar C, Stürck J, Verkerk PJ, Verburg PH, Reenberg A (2016) Hotspots of land use change in Europe. Environ Res Lett 11:064020. https://doi.org/10.1088/1748-9326/11/6/064020

Langat D (2018) Kenya's Water Towers Protection and Climate Change Mitigation and Adaptation (WaTER) Programme Guidelines for Establishing Payment for Ecosystem Services Schemes in Kenya. http://197.248.75.118:8282/jspui/handle/123456789/879. Accessed 02 Feb 2022

Lark TJ, Spawn SA, Bougie M, Gibbs HK (2020) Cropland expansion in the United States produces marginal yields at high costs to wildlife. Nat Commun 11:4295. https://doi.org/10.1038/s41467-020-18045-z

Li C, Kandel M, Anghileri D, Oloo F, Kambombe O, Chibarabada TP, Ngongondo C, Sheffield J, Dash J (2021) Recent changes in cropland area and productivity indicate unsustainable cropland expansion in Malawi. Environ Res Lett 16:084052. https://doi.org/10.1088/1748-9326/ac162a

Maeda E (2011) Agricultural expansion and climate change in the Taita Hills, Kenya: an assessment of potential environmental impacts. Dissertation, University of Helsinki

Mas J-F, Soares Filho B, Pontius RG, Farfán Gutiérrez M, Rodrigues H (2013) A suite of tools for ROC analysis of spatial models. ISPRS Int J Geo Inf. https://doi.org/10.3390/ijgi2030869

Masayi NN, Omondi P, Tsingalia M (2021) Assessment of land use and land cover changes in Kenya's Mt. Elgon forest ecosystem. Afr J Ecol 59:988–1003. https://doi.org/10.1111/aje.12886

Maua JO, Mbuvi MTE, Matiku P, Munguti S, Mateche E, Owili M (2022) The difficult choice - to conserve the living filters or utilizing the full potential of wetlands: Insights from the Yala swamp,

Kenya. Environ Chall 6:100427. https://doi.org/10.1016/j.envc.2021.100427

Msoffe FU, Kifugo SC, Said MY, Neselle MO, Van Gardingen P, Reid RS, Ogutu JO, Herero M, de Leeuw J (2011) Drivers and impacts of land-use change in the Maasai Steppe of northern Tanzania: an ecological, social and political analysis. J Land Use Sci 6:261–281. https://doi.org/10.1080/1747423X.2010.511682

Mulinge W, Gicheru P, Murithi F, Maingi P, Kihiu E, Kirui OK, Mirzabaev A (2016) Economics of land degradation and improvement in Kenya. In: Nkonya E, Mirzabaev A, von Braun J (eds) Economics of land degradation and improvement—a global assessment for sustainable development. Springer International Publishing, Cham, pp 471–498

Munthali MG, Davis N, Adeola AM, Botai JO, Kamwi JM, Chisale HLW, Orimoogunje OOI (2019) Local perception of drivers of land-use and land-cover change dynamics across Dedza district, Central Malawi Region. Sustainability. https://doi.org/10.3390/su11030832

Mutoko MC, Hein L, Bartholomeus H (2014) Integrated analysis of land use changes and their impacts on agrarian livelihoods in the western highlands of Kenya. Agric Syst 128:1–12. https://doi.org/10.1016/j.agsy.2014.04.001

Mwangi HM, Lariu P, Julich S, Patil SD, McDonald MA, Feger K-H (2018) Characterizing the intensity and dynamics of land-use change in the Mara River Basin, East Africa. Forests. https://doi.org/10.3390/f9010008

Mwangi N, Waithaka H, Mundia C, Kinyanjui M, Mutua F (2020) Assessment of drivers of forest changes using multi-temporal analysis and boosted regression trees model: a case study of Nyeri County, Central Region of Kenya. Model Earth Syst Environ 6:1657–1670. https://doi.org/10.1007/s40808-020-00781-2

Mwaniki WM, Möller SM (2015) Knowledge based multi-source, time series classification: a case study of central region of Kenya. Appl Geogr 60:58–68. https://doi.org/10.1016/j.apgeog.2015.03.005

Mwaura JI, Kenduiywo BK (2021) County level maize yield estimation using artificial neural network. Model Earth Syst Environ 7:1417–1424. https://doi.org/10.1007/s40808-020-00943-2

Nakalembe C, Dempewolf J, Justice C (2017) Agricultural land use change in Karamoja Region, Uganda. Land Use Policy 62:2–12. https://doi.org/10.1016/j.landusepol.2016.11.029

Nyberg Y, Jonsson M, Laszlo Ambjörnsson E, Wetterlind J, Öborn I (2020) Smallholders' awareness of adaptation and coping measures to deal with rainfall variability in Western Kenya. Agroecol Sustain Food Syst 44:1280–1308. https://doi.org/10.1080/21683565.2020.1782305

Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, Wulder MA (2014) Good practices for estimating area and assessing accuracy of land change. Remote Sens Environ 148:42–57. https://doi.org/10.1016/j.rse.2014.02.015

Ondiek RA, Vuolo F, Kipkemboi J, Kitaka N, Lautsch E, Hein T, Schmid E (2020) Socio-Economic Determinants of land use/cover change in wetlands in East Africa: a case study analysis of the Anyiko Wetland, Kenya. Front Environ Sci 7:207

Ongugo P, Langat D, Oeba VO, Kimondo JM, Owuor B, Njuguna J, Okwaro G, Russell AJ (2014) A review of Kenya's National Policies relevant to climate change adaptation and mitigation: insights from Mount Elgon. CIFOR. https://www.cifor.org/publications/pdf_files/WPapers/WP155Russell.pdf. Accessed 15 Dec 2021

Park N-W (2011) Application of Dempster-Shafer theory of evidence to GIS-based landslide susceptibility analysis. Environ Earth Sci 62:367–376. https://doi.org/10.1007/s12665-010-0531-5

Pellikka PKE, Clark BJF, Gosa AG, Himberg N, Hurskainen P, Maeda E, Mwang'ombe J, Omoro LMA, Siljander M (2013) Chapter 13—agricultural expansion and its consequences in the Taita

Hills, Kenya. In: Paron P, Olago DO, Omuto CT (eds) Developments in earth surface processes. Elsevier, pp. 165–179

Peng C-YJ, Lee KL, Ingersoll GM (2002) An introduction to logistic regression analysis and reporting. J Educ Res 96:3–14. https://doi.org/10.1080/00220670209598786

Pham LV, Smith C (2014) Drivers of agricultural sustainability in developing countries: a review. Environ Syst Decis 34:326–341. https://doi.org/10.1007/s10669-014-9494-5

R Core Team (2020) A language and environment for statistical computing. Foundation for Statistical Computing

Ridgeway G (2005) Generalized boosted models: a guide to the gbm package

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform 12:77. https://doi.org/10.1186/1471-2105-12-77

Rotich B, Ojwang D (2021) Trends and drivers of forest cover change in the Cherangany hills forest ecosystem, western Kenya. Glob Ecol Conserv 30:e01755. https://doi.org/10.1016/j.gecco.2021.e01755

Sarkar A, Chouhan P (2020) Modeling spatial determinants of urban expansion of Siliguri a metropolitan city of India using logistic regression. Model Earth Syst Environ 6:2317–2331. https://doi.org/10.1007/s40808-020-00815-9

Schapire RE (2003) The boosting approach to machine learning: an overview. In: Denison DD, Hansen MH, Holmes CC, Mallick B, Yu B (eds) Nonlinear estimation and classification. Springer, New York, pp 149–171

Schürmann A, Kleemann J, Fürst C, Teucher M (2020) Assessing the relationship between land tenure issues and land cover changes around the Arabuko Sokoke Forest in Kenya. Land Use Policy 95:104625. https://doi.org/10.1016/j.landusepol.2020.104625

Serneels S, Lambin EF (2001) Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. Agr Ecosyst Environ 85:65–81. https://doi.org/10.1016/S0167-8809(01)00188-8

Song W, Deng X (2017) Land-use/land-cover change and ecosystem service provision in China. Sci Total Environ 576:705–719. https://doi.org/10.1016/j.scitotenv.2016.07.078

Tilman D, Balzer C, Hill J, Befort Belinda L (2011) Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci 108:20260–20264. https://doi.org/10.1073/pnas.1116437108

Twesigye CK, Onywere SM, Getenga ZM, Mwakalila SS, Nakiranda JK (2011) The impact of land use activities on vegetation cover and water quality in the Lake Victoria watershed. Open Environ Eng J 4:66–77

Were K, Dick ØB, Singh BR (2014) Exploring the geophysical and socio-economic determinants of land cover changes in Eastern Mau forest reserve and Lake Nakuru drainage basin, Kenya. GeoJournal 79:775–790. https://doi.org/10.1007/s10708-014-9525-2

Winkler K, Fuchs R, Rounsevell M, Herold M (2021) Global land use changes are four times greater than previously estimated. Nat Commun 12:2501. https://doi.org/10.1038/s41467-021-22702-2

Zaveri E, Russ J, Damania R (2020) Rainfall anomalies are a significant driver of cropland expansion. Proc Natl Acad Sci 117:10225–10233. https://doi.org/10.1073/pnas.1910719117

Zeng Z, Estes L, Ziegler AD, Chen A, Searchinger T, Hua F, Guan K, Jintrawet A, Wood EF (2018) Highland cropland expansion and forest loss in Southeast Asia in the twenty-first century. Nat Geosci 11:556–562. https://doi.org/10.1038/s41561-018-0166-9

Zhai R, Zhang C, Li W, Zhang X, Li X (2020) Evaluation of driving forces of land use and land cover change in New England area by a mixed method. ISPRS Int J Geo Inf. https://doi.org/10.3390/ijgi9060350