Insights from Big Spatial Data through Machine Learning Techniques for Prudent Management of Natural Resources

T V Ramachandra[1,2,3,*] Paras Negi[1,4] and Bharath Setturu[1]

Journal of Resources, Energy, and Development **19** (1&2): 1–18

Abstract

Evaluation of Land Use Land Cover (LULC) changes play a vital role in understanding the landscape dynamics that have been influencing climate, biodiversity, hydrology, and ecology of a region. The information of temporal LULC aids decision-makers in framing sustainable land use policies for nature conservation. Anthropogenic pressure, especially unplanned developmental activities, has contributed towards fragmenting contiguous forests, thus affecting their structure and loss of habitat for endemic taxa. LULC changes in the Bellary district, Karnataka have been assessed through temporal remote sensing data. Classification of remote sensing data for estimating the spatial extent of land uses has been done through supervised machine learning algorithms namely random forest (RF), support vector machine (SVM), and parametric maximum likelihood classifier (MLC). The performance of these algorithms was evaluated through accuracy assessments. Results reveal that RF has the highest overall accuracy (88.94%) and Kappa value (0.76) compared to overall Kappa of MLC (85.51%, 0.74) and SVM (85.47%, 0.63). Based on this, RF was considered for temporal data analyses, which highlighted the decline of forest cover from 2.61% (1973) to 0.74% (2022). The built-up has increased from 0.27% (1973) to 2.43% (2022), and agriculture from 68.21% (1973) to 84.95% (2022). Fragmentation of contiguous forests is evident from the decline in the interior or intact forests from 6.73% (1973) to 2.41% (2022) and the increase in the non-forest areas such as built-up, agriculture, etc., amounting now to 89.81%. Results highlight the need for immediate policy interventions for the conservation and protection of the remnant forest patches

Keywords: LULC, Forest fragmentation, Supervised learning techniques, Machine learning

[1] Energy and Wetlands Research Group, Environmental Information System (EIACP), Center for Ecological Sciences, New Bioscience Building, Third Floor, E-Wing [Near D-Gate], Indian Institute of Science, Bangalore 560012, Karnataka, India
[2] Centre for Sustainable Technologies (astra), Indian Institute of Science, Bangalore 560 012, Karnataka, India
[3] Centre for Infrastructure, Sustainable Transportation and Urban Planning, Indian Institute of Science, Bangalore 560 012, Karnataka, India
[4] Department of Remote Sensing and GIS, Soban Singh Jeena University, Almora, Uttarakhand, India
* Email: tvr@iisc.ac.in

Introduction

A landscape is a mosaic of heterogenous ecosystem elements with dynamic interactions (Forman 1995). It is a physical system composed of biotic (flora, fauna, etc.) and abiotic (climate, soil, water) elements that are governed by natural processes and anthropogenic activities. Landscape dynamics are reflected through the changes in land use and land cover and operate on a temporal scale in the physical and biological properties of a landscape (Ramachandra, Setturu, Karthik, *et al.* 2022). A landscape structure gets altered by the forces driven by humans as well as the forces of nature. Human needs and the capacity of the biosphere to provide goods and services in the long term are big challenges, which necessitates prudent management of the ecosystem to maintain the trade-off (Foley, DeFries, Asner, *et al.* 2005).

Land Use Land Cover (LULC) refers to the categorization or classification of natural features and human activities on the landscape in a specific time frame (Mariye, Jianhua, and Maryo 2022). Land cover refers to the physical cover of the surface, and provides information on the extent of vegetation and non-vegetation (bare soil, hard surfaces including rocks and buildings, water bodies). Land use is an outcome of anthropogenic activities and provides information on the spatial extent of diverse land uses such as agriculture, wildlife habitat, recreation. Land use changes are amplified by the burgeoning human population, coupled with enhanced consumption, energy demand, agriculture intensification, and overuse of natural resources (Kobayashi, Higa, Higashiyama, *et al.* 2020). Knowledge of land use and land cover is essential to formulate appropriate land use policies that are aimed at mitigating land degradation and deforestation (Solomun, Barger, Cerda, *et al.* 2018).

Land use changes are influenced by a wide range of drivers, which are either proximate (direct) or underlying (indirect). The direct driving factors also have indirect impact (Geist and Lambin 2002). Anthropogenic activities lead to changes in land use fragments ecosystems because of loss of natural habitat, degradation of forest connectivity between landscapes, and decline in biodiversity. The LULC dynamics lead to an increase in the surface temperature of the earth (Tan, Yu, Li, *et al.* 2020). The increase in concentrated human activities and paved areas, and the decline of vegetation and water bodies lead to increased heat emission from land surface, which raises atmospheric temperature and land surface temperature (LST) as well (Bharath, Rajan, and Ramachandra 2013).

Fragmentation of forest ecosystem is the process in which large contiguous forests are divided into small and isolated forest patches either by natural phenomena or because of anthropogenic activities such as road construction, logging, conversion to agriculture, or wildfire. These cause drastic changes in forest patch sizes, shape, connectivity, and internal heterogeneity (Ramachandra and Bharath 2016; Wade, Riitters, Wickham, *et al.* 2003). Degradation of forest ecosystems is a direct threat to biodiversity and endangers the sustainability of ecological goods and services from forestland (Young and Boyle 2000). Land uses tend to expand over time, and forests that share a high proportion of their borders with anthropogenic uses are at higher risk of further degradation than forests that share a high proportion of their borders with non-forest, natural land cover (e.g., wetland). LULC changes leading to deforestation have altered landscape structure (Ramachandra, Setturu, and Aithal 2020). Deforestation is a key contributor to species extinction and climate change (Ramachandra, Setturu, and Gupta 2018), yet the extent of tropical forests and their rate of destruction and degradation through fragmentation remain poorly known (Harper, Steininger, Tucker, *et al.* 2007). Land-use maps offer the potential for assessing forest fragmentation and its impacts on greenhouse gas emissions, biodiversity, economics, and water quality (Riitters, Wickham, O'Neill, i. 2000).

Assessment of Land Use Land Cover Changes (LULCC) using temporal remote sensing data aid in understanding landscape dynamics, which

is crucial for social and economic development (Lambin, Turner, Geist, *et al.* 2001). LULCC information provides vital insight for planning, management, and monitoring of natural resources at local, regional, and national levels. It provides insights into land utilization, which aids in formulating policies and programmes towards sustainable management of natural resources.

Spatial data acquired through space-borne sensors since the 1970s at regular intervals (remote sensing data) constitute big data, which are useful for analysing landscape dynamics. Satellite remote sensing provides accurate, reliable, and timely geospatial information for describing changes in the landscape through the assessment of LULC (Foody 2003). Advancements in machine learning algorithms have aided in assessing spatio-temporal patterns by taking advantage of the availability of voluminous multiresolution data. Geoinformatics (Geographic Information System and remote sensing data) help in the quantitative evaluation of land cover dynamics. Monitoring spatial and temporal patterns of land use is very important in maintaining the ecosystem integrity and meeting the objective of sustainable development through conservation programmes (Anderson 1976).

Data science needs to be aligned with machine learning techniques such that remote sensing data can be classified efficiently to derive land use information. Support vector machine (SVM), random forest (RF) method, Linear Discriminant Analysis (LDA), and K-nearest neighborhood (KNN) are some of the machine learning use algorithms that can assess patterns in the spatial data (Chen, Dewi, Huang, *et al.* 2020). Ensemble algorithms have fewer limitations, such as over-fitting, high collinearity, high sensitivity to outliers, high dimensionality, and large noise as compared to classification algorithms, including ISODATA and K-Means (Piao, Jeong, Park, *et al.* 2021). The non-parametric approach of ensemble algorithms such as RF, SVM, K-nearest neighbour, artificial neural network, and classification and regression tree based on nonlinear data improve classification accuracy by reducing collinearity and noise processing of time-series data without overfitting (Piao, Jeong, Park, *et al.* 2021).

One of the most widely used machine learning algorithms is RF. It is used with categorical and continuous variables to perform both classification and regression tasks (Abdi et al. 2019). RF algorithm creates decision trees on data samples randomly (Bagging) and then gets the prediction from each of them and finally selects the best solution through voting. RF had two main parameters to be optimized: the number of trees (*ntree*) and the number of features in each split (*mtry*). Several studies have stated that satisfactory results could be achieved with default parameters (Thanh Noi and Kappas 2017). The problem of overfitting is overcome by RF by averaging or combining the results of the number of decision trees and as the number of input features increases so does the accuracy significantly (Breiman 2001; Zhang, Su, Xu, *et al.* 2021). RF is not sensitive to the lower number of training sample sizes and is resistant to the presence of noise compared to other machine learning classifiers (Na, Zhang, Li, *et al.* 2010). It is positively correlated with spectral indices [normalized difference water index (NDWI), normalized difference vegetation index (NDVI), and normalized difference built-up index (NDBI)] (Talukdar, Singha, Mahato, *et al.* 2020).

The SVM is a non-parametric classifier (Huang, Xie, Tay, *et al.* 2009; Vapnik and Chervonenkis 1979; Chervonenkis 2013). It is a set of related learning algorithms used for classification and regression (Moguerza and Muñoz 2006). It is a linear binary classifier and finds an optimal hyperplane that separates the input training samples. SVM locates a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression, and other tasks. The dimension of the hyperplane depends upon the number of features. It is mostly used in linear and nonlinear separation problems and uses a subset of training points in the decision function called support vectors. SVM is a kernel-based classifier that projects low-dimensional input space and transforms it into higher-dimensional

space. There are several kernels, such as linear, polynomial, radial basis, and sigmoid kernels (Kecman 2005), and a category is assigned based on the most frequently assigned class.

SVM with the radial kernel used to classify Corine's data, Sentinel-2, and Landsat 8 data outperformed with better accuracy compared to RF applied in diverse areas of Catalonia, Poland, and Romania (Dabija, Kluczek, Zagajewski, *et al.* 2021).

Among the most popular traditional classifiers, widely used is the MLC (Huang *et al.* 2002). It is a parametric approach and uses a fixed number of parameters regardless of the amount of training data. Compared to this, artificial neural networks (ANN), decision trees, RF, and SVM have greater flexibility and offer a better fit to the data (Kavzoglu and Colkesen 2009).

The Google Earth Engine (GEE), a cloud-based spatial data computing platform with data archive, has revolutionized remote sensing-based applications with fast and easy computation (Phan, Kuch, and Lehnert 2020; Mutanga and Kumar 2019). It consists of a multi-petabyte analysis-ready data catalogue co-located with a high-performance, intrinsically parallel computation service. It is accessed and controlled through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment (IDE) that enables rapid prototyping and visualization of results (Gorelick 2017).

Objectives

The objectives of the current study are to (i) understand landscape dynamics in Bellary district, Karnataka, India, (ii) assess the condition of the forest ecosystem through computation of fragmentation metrics, and (iii) performance evaluation of the selected parametric (MLC) and non-parametric (RF and SVM) classifiers.

Study Area

Bellary district is a major district in Karnataka with a spatial extent of 9896 km². It is located in the northeastern part of Karnataka at between 14° 30' and 15° 50' N, and 75°40' and 77° 11' E (Figure 1). The district is bounded by Raichur district in the north, Anantapur and Kurnool districts of Andhra Pradesh in the east, Chitradurga and Davanagrere in the south, and Dharwad district in the west. River Tungabhadra flows eastwards and forms the natural boundary dividing the district on the west from Dharwar and on the north from Raichur. The population of the district as per the 2011 census was about 2,452,595 (density of 300 per km²) with a rural population of 1,320,290.

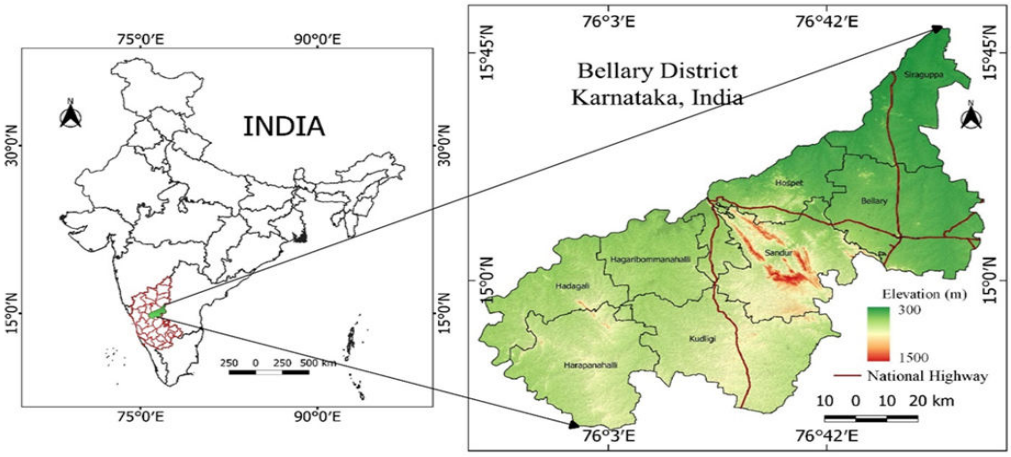

Figure 1 Study area

Bellary district comprises eight taluks: Kudligi, Hadagali, HB Halli, Hospet, Sandur, Siruguuppa, Harapanahalli, and Bellary. As per *Statistical Report 2020–21* (https://cdn.s3waas.gov.in/), there will be 11 taluks when Bellary is reorganized.

The district consists of plains and undulating terrain consisting of northern maidan region with a monotonous, treeless, and expansive plateau landscape. The central part is occupied by the hills and plain to the east and west. The Sandur hills run north-west of the district and divides the district into two equal parts. The average elevation of the district is 493.71 m above sea level. It has a rocky terrain with monolithic granite hills, and gneisses hills of a height of about 500 m to 1 km.

The major type of soil in the district is red soil, which is found at the fringes of the hills. The soil of the district's stream bed is sandy loam soil mixed with black and grey soil. The thickness of the soil varies from 0.2 to 1 mm and red soil is of high permeability and neutral pH value. The black soil is found in the prolonged submerged area and canals areas having low permeability. The climate of the Bellary district is quite moderate or semi-arid. It receives rainfall from the southwest monsoon between June and September and from the northeast monsoon from October to December. The district receives an average rainfall of 502–600 mm with the highest in Sandur taluk at 783 mm.

The total Gross District Domestic Product (GDDP) of the district estimated during 2012-13 is Rs 10,618 crore. The per capita annual income of the district was Rs 74,554 (2012–13).

Data

Table 1 lists the data used for the analyses of landscape dynamics. The multi-spectral remote sensing data of Landsat 1 MSS, Landsat 5 TM, Landsat 8 OLI, and Landsat 9 OLI-2 with a minimum cloud cover of less than 10 % were used for spatial analyses. The data was preprocessed and rectified radiometrically and atmospherically. Accuracy assessment was done using training data compiled from the field and supplemented with the data from virtual data portals [Karnataka Geographic Information Science (K-GIS) Portal and Bhuvan Geo Portal].

Table 1 Data used for landscape dynamics

Data	Band combination	Spatial resolution	Radiometric resolution	Date	
Landsat Multi Spectral Scanner (MSS)	B4, B5, B6	60 m	6-bit	1973	
US Geological Survey				Scene 1	5 January
https://earthexplorer.usgs.gov/				Scene 2	10 February
				Scene 3	6 February
				Scene 4	6 February
Landsat Thematic Mapper (TM)	B1, B2, B3, B4, B5, B7		8-bit	2000	
US Geological Survey		30 m		Scene 1	26 December
https://earthexplorer.usgs.gov/				Scene 2	26 December
				Scene 3	21 December
Landsat Enhanced Thematic Mapper			9-bit	2007	
(ETM+)				Scene 1	28 April
US Geological Survey				Scene 2	28 April
https://earthexplorer.usgs.gov/				Scene 3	5 April
				2014	

Contd...

Table 1 *Contd...*

Data	Band combination	Spatial resolution	Radiometric resolution	Date	
Landsat Operational Land Imager (OLI and OLI-2) *US Geological Survey* https://earthexplorer.usgs.gov/	B2, B3, B4, B5, B6, B7		16-bit	Scene 1 Scene 2 Scene 3	7 April 22 March 5 April
				2022	
				Scene 1 Scene 2 Scene 3	24 February 12 March 1 February
Thematic Maps Karnataka Geographic Information Science Portal https://kgis.ksrsac.in/	LULC Reference Maps			LULC Reference Maps	
Thematic Maps (1:50000) Bhuvan Geo Portal https://bhuvan.nrsc.gov.in/					

Method

Assessment of landscape dynamics involved computation of (i) land cover and (ii) land uses with the help of supervised classifiers based on non-parametric and parametric algorithms. See Figure 2.

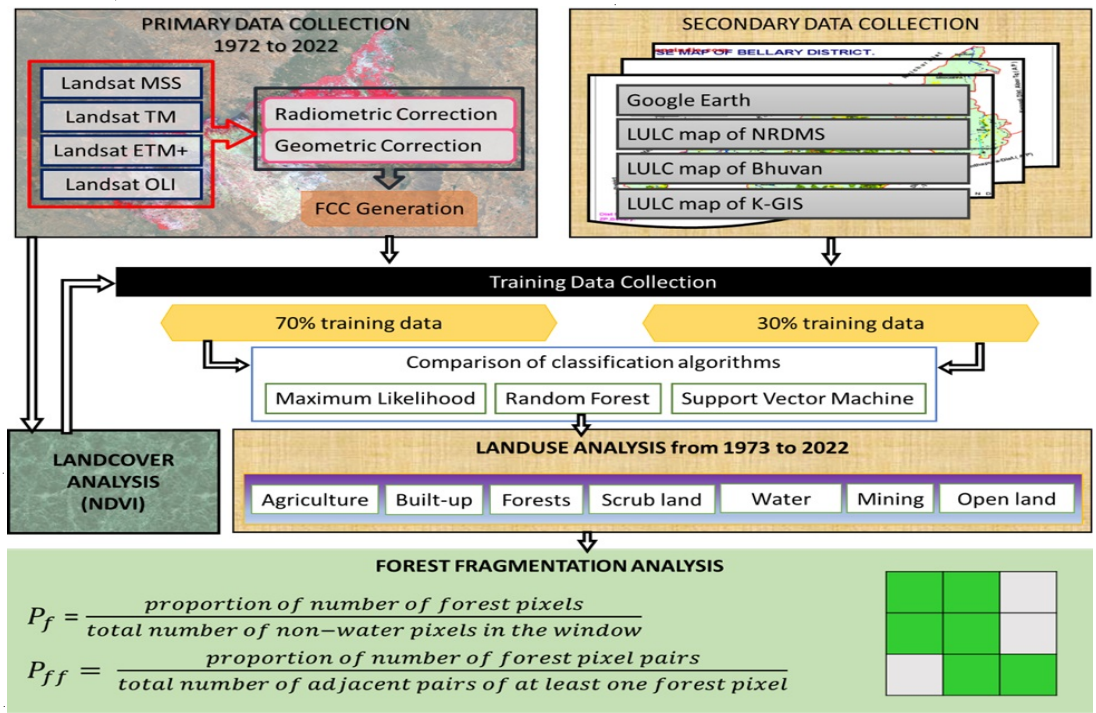

Figure 2 Method

Land cover classification

The land cover helped in delineating areas under vegetation and non-vegetation in the district, which is calculated by Normalized Difference Vegetation Index (NDVI) as per Equation (1) and the index value ranges from −1 to 1 in which values ranging from 0 to 0.3 depict the scrub land and agricultural fallow land and values above 0.3 indicate the area under vegetation or healthy vegetation land cover.

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

Land use classification

The spatio-temporal changes in land use were analysed using multi-temporal remote sensing data from 1973 to 2022. False Color Composite (FCC) aided in identifying heterogeneous features in the landscape. Training polygons were digitized by covering all heterogeneous features, which are uniformly distributed throughout the landscape, and the spatial extent of these training polygons covered 15% of the study region. The attribute information of these training polygons was obtained from field by using a pre-calibrated handheld global positioning system (GPS) and virtual high spatial resolution Google Earth data. About 70% of training data is used for classification and the remaining 30% is used for accuracy assessment. Classification of remote sensing data is done through a supervised learning algorithm (RF, SVM, MLC).

Random Forest: RF is an ensemble tree-based learning algorithm (Breiman 2001). RF classifier creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution through voting. RF overcomes the problem of overfitting by averaging or combining the results of several decision trees (Rodriguez-Galiano, Ghimire, Rogan, *et al.* 2012; Nguyen, Doan, and Radeloff 2018).

The decision tree uses the Gini index, which measures the inequality between each class in each node as shown in Equation (2). The lower the value of the Gini Index (nearer to 0), the purer the node or homogeneous, which is good for accurate classification.

$$Gini = 1-\sum_{(i=1)}^{n} P_i^2 \quad (2)$$

Bagging (an acronym for Bootstrap AGGregatING) is a regular ensemble classifier technique in which independently several predictions are made and then grouped using the weighted average (Breiman 1996). In bagging, the algorithm is built on bootstrap by replicating samples with replacement, and then each replica is used for classification. The classes are then assigned by taking the average (or voting principle) of the output of all iterations (Maclin and Opitz 1997).

Support Vector Machine: SVM is a kernel-based classifier that projects low-dimensional input space and transforms it into higher-dimensional space (Vapnik 1999). The training is linearly separated, where it projects x from the original space to projected space $\Phi(x)$. The training problem appears in the form of a dot product of two vectors $[\Phi(x_i), \Phi(x_j)]$.

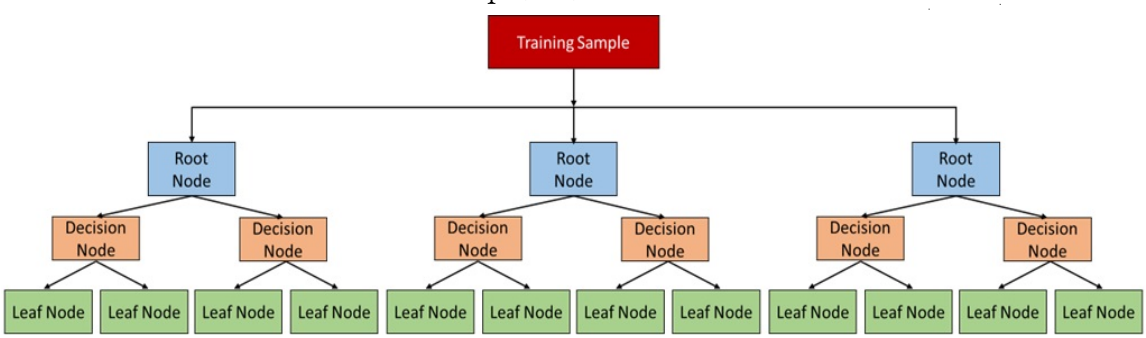

Figure 3 Random forest

The widely used approach is a one-against-one rule that trains each pairwise class combination and the class label is assigned based on the most frequently assigned class (Debnath, Takahide, and Takahashi 2004; Cherkassky and Ma 2004; Al-Mejibli, Alwan, and Abd Dhafar 2020). The computational cost in higher dimensional space is less because of the kernel transformation k as per Equation (3).

$$(\emptyset(\boldsymbol{x}_i),\emptyset(\boldsymbol{x}_j) = k(x_i,x_j) \quad (3)$$

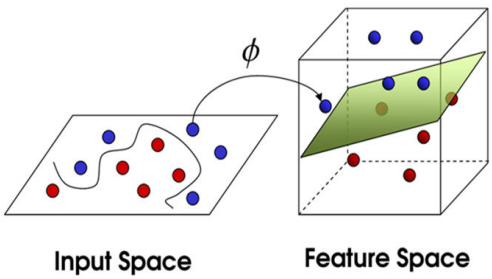

Figure 4 Data classification using kernel

Maximum Likelihood Classifier: MLC assigns a pixel to a class according to its probability of belonging to a particular class. Its mean and covariance are modelled as forming a normal distribution in multispectral feature space (Le Cam 1990). MLC is based on Bayes' classification (Sisodia, Tiwari, and Kumar 2014). The conditional probabilities are given as , $i = 1,2,3\ldots N$, where represents the classes, i represents the number of classes, and represents the measurement vector.

The Baye's theorem is given in Equation (4).

$$P(C_i \mid E) = \frac{(P(E|C_i)P(C_i)}{P(E)} \quad (4)$$

where $P(C_i)$ represents the probability from class C_i and $P(E)$ is the probability of finding a pixel with measurement vector E.

$$P(E) = \sum_{(i=1)}^{N} P(E|C_i)P(C_i) \quad (5)$$

Multivariate Gaussian distribution for N dimensional space is given in Equation (6).

$$P(E|C_i) = (2\pi)^{\frac{-N}{2}} |Y_i|^{\frac{-1}{2}} \exp\{-\frac{1}{2}(E-\mu_i)^T\}Y_i^{-1}(E-\mu_i)w \quad (6)$$

where μ_i and Y_i are the mean vector and variance-covariance matrix of class i.

Accuracy Assessment

Accuracy assessment of classification is done using training data through a confusion matrix or error matric, which compares, on a class-by-class basis, results of classification with the reference data or ground truth. It is assessed through computation of producer accuracy (Equation 7), user accuracy (Equation 8), overall accuracy (Equation 9), and Kappa Statistic or Cohen's Kappa (Equation 10).

$$\text{Producer accuracy} = \frac{C_{ij}}{C_{sum}} \times 100\% \quad (7)$$

where C_{ij} is the element at position i^{th} row and j^{th} column and C_{sum} is column sum.

$$\text{User accuracy} = \frac{C_{ii}}{R_{sum}} \times 100\% \quad (8)$$

where R_{sum} is row sum.

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{n} C_{ij}}{Q} \times 100\% \quad (9)$$

where n and Q are the total number of classes and pixels, respectively.

$$\text{Kappa statistics} = \frac{\text{(Total features} \times \text{No. of correct features)-(Sum of all rows total} \times \text{Column total)}}{\text{(Total no. of features squared)-(Sum of all rows total} \times \text{Column total)}} \quad (10)$$

Temporal changes in land uses are estimated as per Equation (11) by considering 1973 as the base year.

$$\text{Change rate} = \left(\frac{\text{Land use area of current year-Land use area of base year}}{\text{Current year-Base year}}\right) \times 100 \quad (11)$$

Forest Fragmentation

The condition of forest ecosystems is evaluated through fragmentation metrics considering proportion of forest cover (equation 12) and occurrence of forest in adjacent pixels (equation 13) using 3×3 kernel. Various levels of fragmentation are computed as per Table 2.

Results and Discussion

The temporal land cover of Bellary district from 1973 to 2022 is assessed through computation of NDVI using remote sensing data and is listed in Table 3. It can be seen that the vegetation cover had increased from 45.56% (in 1973) to 60.28% (in 2022) due to expansion of agricultural (croplands and horticulture) activities. Areas under non-vegetation decreased (Figure 5) from 54.24% (1973) to 39.75% (2022).

Table 2 Forest fragmentation

Fragmentation classes	Description	Computation
Patch	Forested pixels surrounded by non-forested pixel	$P_f < 0.4$
Transitional	Pixels between edge and non-forest pixel	$0.4 < P_f < 0.6$
Edge	Forest pixels that act as the boundary between interior forest pixels and non-forested pixels	$P_f > 0.6$ and $P_f - P_{ff} < 0$
Perforated	Forest pixels that act as the boundary between and interior forest pixels and perforations pixels	$P_f > 0.6$ and $P_f - P_{ff} > 0$
Interior	Forest pixels surrounded by non-forested pixels and the pixels are far from the boundary of forest pixels and non-forest pixels	$P_f = 1$

$$P_f = \frac{\text{Proportion of number of forest pixels}}{\text{Total number of non-water pixels in the window}} \quad (12)$$

$$P_{ff} = \frac{\text{Proportion of number of forest pixel pairs}}{\text{Total number of adjacent pairs of at least one forest pixel}} \quad (13)$$

Table 3 Land cover analysis of Bellary district from 1973 to 2022

Land cover	Non-vegetation		Vegetation	
	km²	%	km²	%
1973	5382.27	54.24	4515.11	45.56
2000	4390.96	44.34	5505.03	55.61
2007	5015.21	50.71	4880.78	49.33
2014	4359.74	44.09	5523.22	55.83
2022	3930.2	39.75	5965.79	60.28

Land use analysis

Land uses presented in Figure 6 show a decline of dry deciduous forest from 258.23 km² (2.61% in 1973) to 73.14 km² (0.74% in 2022), with an increase in the built-up area from 26.33 km² (0.27% in 1973) to 240.47 km² (2.43% in 2022). The water body showed fluctuation due to seasonal variation, especially near the Tungabhadra dam. The scrub land had decreased from 1923.82 km² (19.44% in 1973) to 762.24 km² (7.70% in 2022). Agricultural (cropland and horticulture) activities are very dominant in the district and expansion has led to the decline of scrub land. Mining has increased from 0.89 km² (0.01%) in 2000 to 19.39 km² (0.20%) in 2022. The district has only a thick forest cover in the Sandur taluk and the extraction of iron ore here by the industries had degraded the dry deciduous forest in the Sandur forest range. The open and rocky surfaces have decreased from 760.14 km² (7.68%) in 1973 to 221.34 km² (2.24%) in 2022.

The agricultural land has increased from 6750.66 km² (68.21% in 1973) to 8407.86 km² (84.95% in 2022) with a decline of scrub lands (1260.28 km²). Built-up has increased by 214.16 km², and agriculture by 1657.20 km². Dry deciduous has reduced by 185.10 km², specifically in Sandur Forest range because of increase in mining to 19.39 km² since 2000. Figure 7 illustrates land uses in Bellary district from 1973 to 2022.

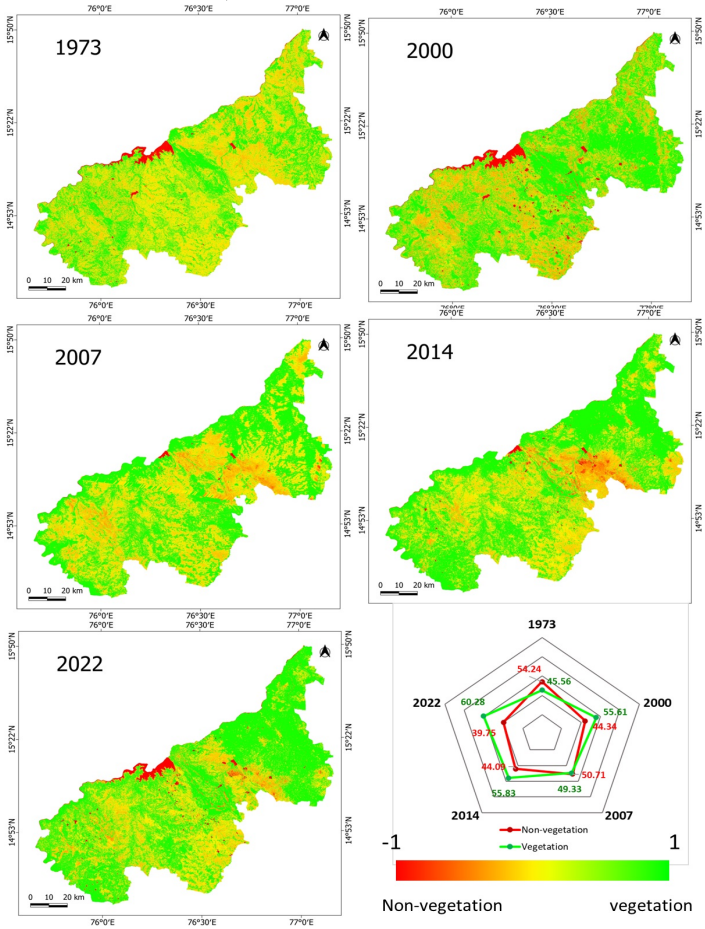

Figure 5 Land cover of Bellary district from 1973 to 2022

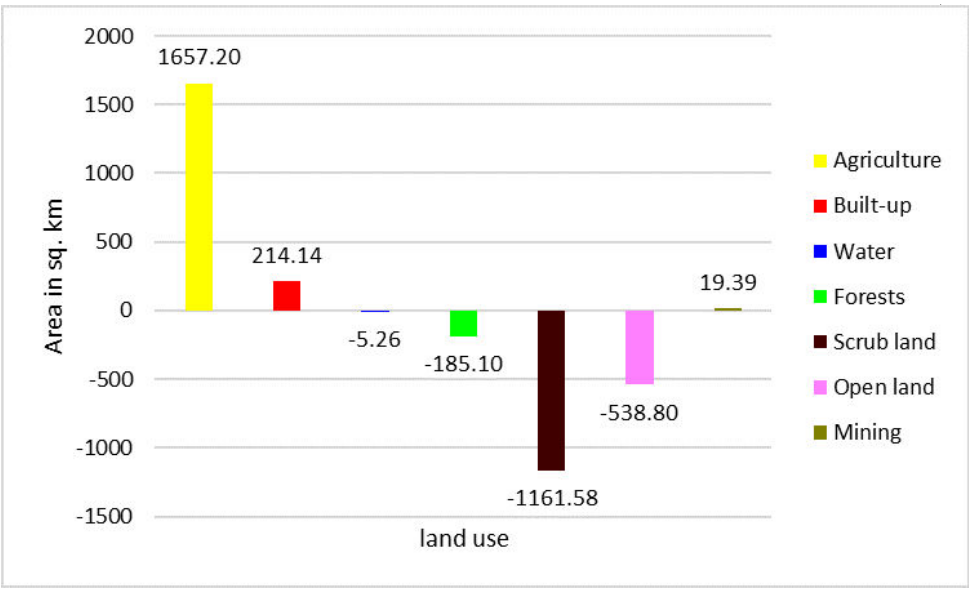

Figure 6 Land use change analysis for each class from 1973 to 2022

Table 4 Land use analysis of Bellary district from 1973 to 2022

Land use		1973	2000	2007	2014	2022
Agriculture	km²	6750.66	7620.69	8125.12	8152.75	8407.86
	%	68.21	77.00	82.09	82.37	84.95
Built-up	km²	26.33	81.73	108.42	165.22	240.47
	%	0.27	0.83	1.10	1.67	2.43
Water	km²	178.30	154.52	173.90	177.61	173.04
	%	1.80	1.56	1.76	1.79	1.75
Forests	km²	258.23	87.84	85.54	82.04	73.14
	%	2.61	0.89	0.86	0.83	0.74
Scrub land	km²	1923.82	1329.76	984.37	987.04	762.24
	%	19.44	13.44	9.95	9.97	7.70
Open land	km²	760.14	622.08	416.97	324.93	221.34
	%	7.68	6.29	4.21	3.28	2.24
Mining	km²	0.00	0.89	3.17	7.90	19.39
	%	0.00	0.01	0.03	0.08	0.20

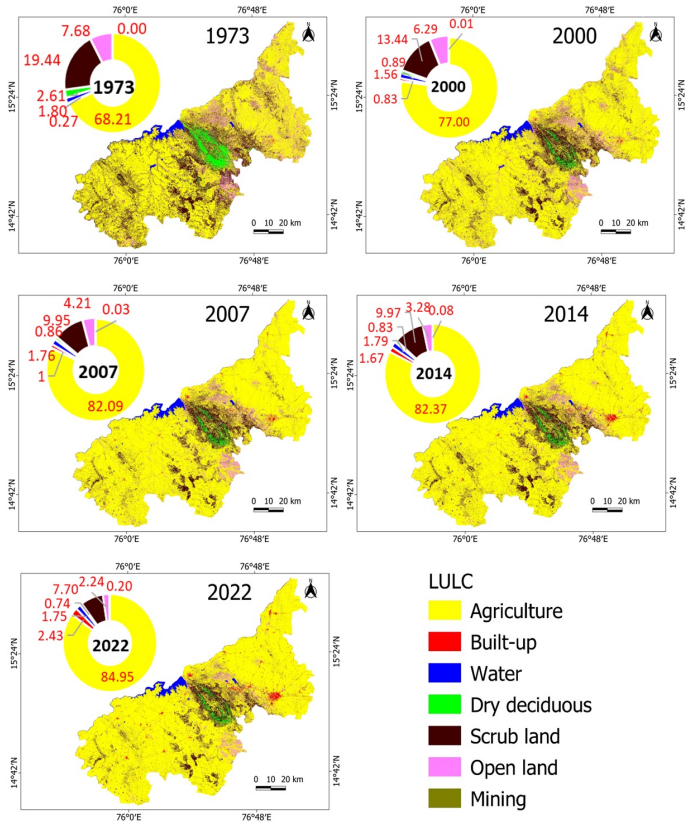

Figure 7 Land use analysis from 1973 to 2022

Performance evaluation

Classification of temporal remote sensing data for Bellary district was done through supervised non-parametric classifiers RF and SVM using Google Earth Engine cloud platform, and a parametric supervised classifier based on maximum likelihood algorithm using GRASS (http://wgbis.ces.iisc.ac.in/grass).

Performance of these classifiers was evaluated through accuracy assessments. The overall accuracy and Kappa of three supervised learning techniques showed that RF (88.94%, 0.76) performed comparatively better compared to MLC (86.51%, 0.74) and SVM (85.47%, 0.63). Classifier-wise results are listed in Table 5. Category-wise, user's and producer's accuracy produced by MLC is maximum for water (94.72%, 95.83%) followed by agriculture (89.88%, 95.34%), while open land/rocky surface and mining have lower values: (73.99%, 53.87%) and (87.95%, 27.17%).

RF produced the highest user's and producer's accuracy for agriculture class (94.52%, 93.69%), followed by water (88.27%, 91.49%), while open land or rocky surface (55.95%, 75.63%) and mining (78.61%, 66.62%) have lower accuracies. SVM performed maximum user's and producer's accuracy for water (96.76%, 97.96%), followed by agriculture (97.03%, 89.28%), while scrub land (56.40%, 55.71%) and open land or rocky surface (38.35%, 62.88%) produced lower accuracy, respectively.

Table 5 Performance evaluation of classifiers through accuracy of land use classification

Accuracy Assessment	Maximum Likelihood		Random Forest		SVM	
	User's accuracy	Producer's accuracy	User's accuracy	Producer's accuracy	User's accuracy	Producer's accuracy
Agriculture	89.88	95.34	94.52	93.69	97.03	89.28
Built-up	77.01	61.64	84.07	64.51	82.92	68.97
Water	94.72	95.83	88.27	91.49	96.76	97.96
Forests	81.28	67.01	86.29	66.64	40.73	99.39
Scrub land	79.34	83.82	79.14	77.25	56.40	55.71
Open land	73.99	53.87	55.95	75.63	38.35	62.88
Mining	87.95	27.17	78.61	66.62	77.31	64.23
Kappa	0.74		0.76		0.63	
Overall accuracy	86.51%		88.94%		85.47%	

Among the chosen three supervised classifiers, RF performed relatively better compared to MLC and SVM, which corroborates with the earlier studies (Adam, Mutanga, Odindi, *et al.* 2014), ML, Decision Trees (DT) (Breiman *et al.* 1984, Friedl and Brodley 1997), SVM (Vapnik 1995, Vapnik 1998, Rodriguez-Galiano, Ghimire, Rogan, *et al.* 2012; Amini, Saber, Rabiei-Dastjerdi, *et al.* 2022, Talukdar, Singha, Mahato, *et al.* 2020, Waske and Braun 2009, Pelletier, Valero, Inglada, *et al.* 2016, Rodriguez-Galiano and Chica-Rivas 2014,Wagle, Acharya, Kolluru, *et al.* 2020, Piao, Jeong, Park, *et al.* 2021).

Forest fragmentation analysis

Fragmentation of forest ecosystems from 1973 to 2022 listed in Table 6 and Figure 8 shows that the area under interior forest has decreased from 665.92 km² (6.73% in 1973) to 238.15 km² (2.41% in 2022). The patch forest has increased from 17.74 km² (0.18 in 1973) to 143.30 km² (1.45% in 2022). The transitional, edge, and perforated forests have decreased from 497.24 km² (5.02%), 403.84 km² (4.08%), and 597.32 km² (6.04%) in 1973 to 149.98 km² (1.52%), 53.86 km² (0.54%), and 250.09 km² (2.53%) in 2022.

Table 6 Forest fragmentation analysis of Bellary district from 1973 to 2022

Forest fragmentation		1973	2000	2007	2014	2022
Non-forest	km²	7537.14	8325.39	8653.69	8650.82	8889.07
	%	76.15	84.12	87.43	87.40	89.81
Patch	km²	17.74	292.09	188.90	193.40	143.30
	%	0.18	2.95	1.91	1.95	1.45
Transitional	km²	497.24	291.78	197.27	198.08	149.98
	%	5.02	2.95	1.99	2.00	1.52
Edge	km²	403.84	93.15	73.37	71.60	53.86
	%	4.08	0.94	0.74	0.72	0.54
Perforated	km²	597.32	394.41	309.43	309.00	250.09
	%	6.04	3.98	3.13	3.12	2.53
Interior	km²	665.92	346.17	300.94	297.00	238.15
	%	6.73	3.50	3.04	3.00	2.41
Water	km²	178.30	154.52	173.90	177.61	173.04
	%	1.80	1.56	1.76	1.79	1.75

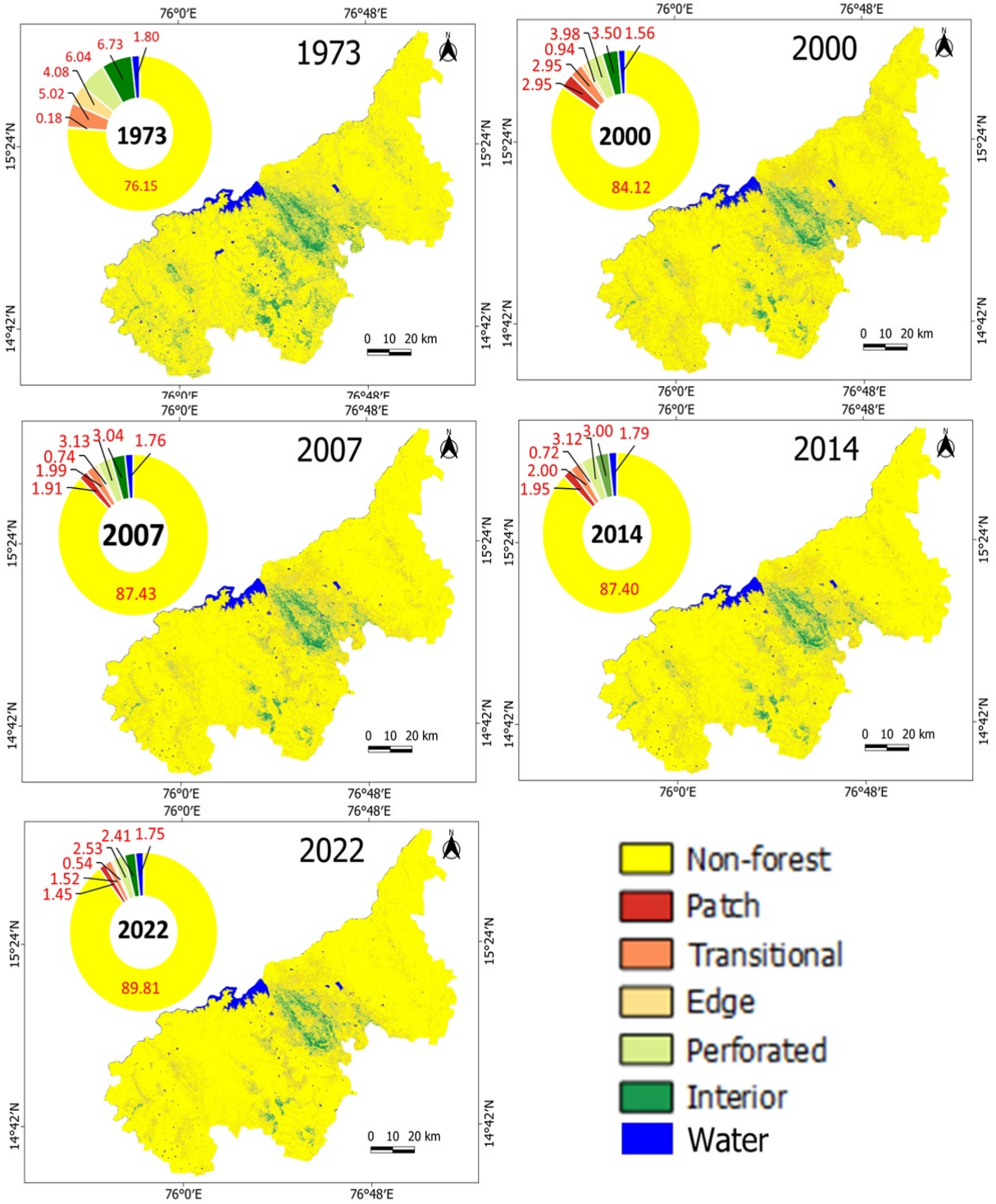

Figure 8. Forest fragmentation analysis of Bellary district from 1973 to 2022

Conclusion

The analyses of LULC dynamics over a temporal scale provided insights to the landscape composition and configuration in the Bellary district in Karnataka. Temporal land use analyses reveal that built-up has increased from 26.33 km² (1973) to 240.47 km² (2022) with a decline in forest cover by 185.10 km² from 1973 to 2022. Agricultural land, the dominant occupation in the district, has increased significantly from 6750.66 km² (1973) to 8407.86 km² (2022). Anthropogenic activities involving conversion of forest to agriculture and industrial developments are the major driving forces of degradation. Forest fragmentation analyses highlight the reduction of interior forests from 6.73% (1973) to 2.41% (2022), with the formation of non-forest patches and increase in patch forests. Land use analyses reveal an increase in urban and agriculture category and degradation on forests during the past three decades. Comparative assessment of performance of parametric and non-parametric supervised classifiers namely RF, SVM, and MLCs reveal highest classification accuracy with RF. It required two user-defined parameters and does not overfit. The advanced techniques based on non-parametric approach of classification of land use has greater flexibility and better fit to the data because of no limitation of selecting the number of parameters to train the data.

Acknowledgement

We are grateful to the ENVIS Division, the Ministry of Environment, Forest and Climate Change, Government of India (grant: CES/TVR/DE007).

References

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

Adam, E, Mutanga, O, Odindi, J, and Abdel-Rahman, E M. 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing* **35**(10): 3440–58.

Al-Mejibli, I S, Alwan, J K, and Abd Dhafar, H. 2020. The effect of gamma value on support vector machine performance with different kernels. *International Journal of Electrical and Computer Engineering* **10**(5): 5497.

Amini, S, Saber, M, Rabiei-Dastjerdi, H, and Homayouni, S. 2022. Urban land use and land cover change analysis using Random Forest classification of Landsat Time Series. *Remote Sens.* **14**(11): 2654.

Anderson, J R. 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data* (Vol. 964). US Government Printing Office.

Bharath, S., Rajan, K. S., and Ramachandra, T. V. 2013. Land surface temperature responses to land use land cover dynamics. *Geoinfor Geostat: An Overview* 54: 50–78.

Breiman, L. 1996. Bagging predictors. *Machine Learning* **24**(2): 123–40.

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

Breiman, L, Friedman, J H, Olshen, R A, and Stone, C J. 2017. *Classification and Regression Trees*. London: Routledge.

Census of India 2011 - Karnataka - Series 30 - Part XII B - District Census Handbook, Bellary

Chen, R C, Dewi, C, Huang, S W, and Caraka, R E. 2020. Selecting critical features for data classification based on machine learning methods. *Journal of Big Data* **7**(1): 1–26.

Cherkassky, V. and Ma, Y. 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* **17**(1): 113–26.

Dabija, A, Kluczek, M, Zagajewski, B, Raczko, E, Kycko, M, Al-Sulttani, A H, Tardà, A, Pineda, L, and Corbera, J. 2021. Comparison of Support Vector Machines and Random Forests for Corine land cover mapping. *Remote Sens.* 13: 777

Debnath, R, Takahide, N, and Takahashi, H. 2004. A decision based one-against-one method for multi-class support vector machine. *Pattern Analysis and Applications* 7(2): 164–75.

Digital India. nd. District at Glance 2017-2018, Bellary district. Details available at https://ballari.nic.in/en/document/district-at-glance-index/

Digital India. nd. District at Glance 2020-2021, Bellary district. Details available at https://ballari.nic.in/en/document/district-at-glance-2020-2021/.

Foley, J A, DeFries, R, Asner, G P, Barford, C, Bonan, G, Carpenter, S R, and Snyder, P K. 2005. Global consequences of land use. *Science* 309(5734): 570–74.

Foody, G M. 2003. Remote sensing of tropical forest environments: towards the monitoring of environmental resources for sustainable development. *International Journal of Remote Sensing* 24(20): 4035–46.

Forman, R T. 1995. Some general principles of landscape and regional ecology. *Landscape Ecology* **10**(3): 133–42.

Friedl, M A and Brodley, C E. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* **61**(3): 399–409.

Geist, H J and Lambin, E F. 2002. Proximate causes and underlying driving forces of tropical deforestation: tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. *BioScience* **52**(2): 143–50.

Gorelick, N, Hancher, M, Dixon, M, Ilyushchenko, S, Thau, D, and Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment* **202**: 18–27.

Harper, G J, Steininger, M K, Tucker, C J, Juhn, D, and Hawkins, F. 2007. Fifty years of deforestation and forest fragmentation in Madagascar. *Environmental Conservation* **34**(4): 325–33.

Huang, B, Xie, C, Tay, R, and Wu, B. 2009. Land-use-change modeling using unbalanced support-vector machines. *Environment and Planning B: Planning and Design* **36**(3): 398–416.

Kavzoglu, T. and Colkesen, I. 2009. A kernel functions analysis for support vector machines for land cover classification. International Journal of Applied Earth Observation and Geoinformation **11**(5): 352–59.

Kecman, V. 2005. Support vector machines–an introduction. In *Support Vector Machines: Theory and Applications,* pp. 1-47. Berlin, Heidelberg: Springer.

Kobayashi, Y, Higa, M, Higashiyama, K, and Nakamura, F. 2020. Drivers of land-use changes in societies with decreasing populations: A comparison of the factors affecting farmland abandonment in a food production area in Japan. *PLoS One* 15(7): e0235846.

Lambin, E F, Turner, B L, Geist, H J, Agbola, S B, Angelsen, A, Bruce, J W, and Xu, J. 2001. The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change* **11**(4): 261–69.

Le Cam, L. 1990. Maximum likelihood: an introduction. *International Statistical Review/Revue Internationale de Statistique* 153–171.

Maclin, R and Opitz, D. 1997. An empirical evaluation of bagging and boosting. *AAAI/IAAI* 546–51.

Mariye, M, Jianhua, L, and Maryo, M. 2022. Land use and land cover change, and analysis of its drivers in Ojoje watershed, Southern Ethiopia. *Heliyon* 8(4): e09267.

Moguerza, J M and Muñoz, A. 2006. Support vector machines with applications. *Statistical Science* 21(3): 322–336.

Mutanga, O, and Kumar, L. 2019. Google earth engine applications. *Remote Sensing* **11**(5): 591.

Na, X, Zhang, S, Li, X, Yu, H, and Liu, C. 2010. Improved land cover mapping using random forests combined with Landsat thematic mapper imagery and ancillary geographic data. *Photogrammetric Engineering and Remote Sensing* **76**(7): 833–840.

Nguyen, H T T, Doan, T M, and Radeloff, V. 2018. Applying random forest classification to map land use/land cover using Landsat 8 OLI. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42(3): W4.

Otukei, J R, and Blaschke, T. 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation* **12**: S27–S31.

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

Pelletier, C, Valero, S, Inglada, J, Champion, N, and Dedieu, G. 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment* 187: 156–168.

Phan, T N, Kuch, V, and Lehnert, L W. 2020. Land cover classification using Google Earth Engine and Random Forest classifier—the role of image composition. *Remote Sensing* 12(15): 2411.

Piao, Y, Jeong, S, Park, S, and Lee, D. 2021. Analysis of land use and land cover change using time-series data and random forest in North Korea. Remote Sensing 13(17): 3501.

Ramachandra, T V, Setturu, B, and Aithal, B H. 2020. Insights of forest dynamics for the regional ecological fragility assessment. *Journal of the Indian Society of Remote Sensing* **48**(8): 1169–1189.

Ramachandra, T V, Setturu, B, and Gupta, N. 2018. Modelling landscape dynamics with LST in protected areas of Western Ghats, Karnataka. *Journal of Environmental Management* 206: 1253–1262.

Ramachandra, T V, Setturu, B, and Chandran, S. 2016. Geospatial analysis of forest fragmentation in Uttara Kannada District, India. *Forest Ecosystems* 3(1): 1–15.

Ramachandra, T V, Setturu, B, Karthik, R N, and Jagadessha, B P. 2022. Conservation prioritization of ecologically susceptible zones at disaggregated levels. *Advances in Environmental and Engineering Research* **3**(2): 1–1.

Riitters, K, Wickham, J, O'Neill, R, Jones, B, and Smith, E. 2000. Global-scale patterns of forest fragmentation. *Conservation Ecology* **4**(2): 3.

Rodriguez-Galiano, V F, and Chica-Rivas, M. 2014. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *International Journal of Digital Earth* 7(6): 492–509.

Rodriguez-Galiano, V F, Ghimire, B, Rogan, J, Chica-Olmo, M, and Rigol-Sanchez, J P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93–104.

Saini, R, and Ghosh, S K. 2018. Crop classification on single date sentinel-2 imagery using random forest and support vector machine. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **42**: 683–688.

Sisodia, P S, Tiwari, V, and Kumar, A. 2014. Analysis of supervised maximum likelihood classification for remote sensing image. *International Conference on Recent Advances and Innovations in Engineering* (ICRAIE-2014) (pp. 1-4).

Solomun, M K, Barger, N, Cerda, A, Keesstra, S, and Marković, M. 2018. Assessing land condition as a first step to achieving land degradation neutrality: A case study of the Republic of Srpska. *Environmental Science and Policy* 90: 19–27.

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

Tan, J, Yu, D, Li, Q, Tan, X, and Zhou, W. 2020. Spatial relationship between land-use/land-cover changes and land surface temperature in the Dongting Lake area, China. *Scientific Reports* 10(1): 1–9.

Thanh Noi, P, and Kappas, M. 2017. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* 18(1): 18.

Vapnik, V. 1999. *The Nature of Statistical Learning Theory*. Berlin: Springer.

Wade, T G, Riitters, K H, Wickham, J D, and Jones, K B. 2003. Distribution and causes of global forest fragmentation. *Conservation Ecology* 7(2): 7.

Wagle, N, Acharya, T D, Kolluru, V, Huang, H, and Lee, D H. 2020. Multi-temporal land cover change mapping using google earth engine and ensemble learning methods. *Applied Sciences* 10(22): 8083.

Waske, B, and Braun, M. 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **64**(5): 450–457.

Young, A G and Boyle, T J. 2000. Forest fragmentation. *Forest Conservation Genetics: Principles and Practice* 123–134.

Zhang, T, Su, J, Xu, Z, Luo, Y, and Li, J. 2021. Sentinel-2 satellite imagery for urban land cover classification by optimized Random Forest classifier. *Applied Sciences* 11: 543.