



Land use and land cover classification using machine learning algorithms in google earth engine

Arpitha M¹ · S A Ahmed¹ · Harishnaika N¹

Received: 15 June 2023 / Accepted: 7 August 2023 / Published online: 19 August 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023, corrected publication 2023

Abstract

Natural resources are under tremendous amounts of threat as a result of the expanding human population, which over time intensifies changes in Land use and Land cover (LULC). Understanding how various machine learning classifiers function is crucial as the demand for an accurate estimate of LULC from satellite images. The purpose of this research was to classify the LULC in the entire Karnataka state, using three distinct methods on the Google Earth Engine (GEE) namely RF (Random Forest), SVM (Support Vector Machine) and CART (Classification Regression Trees), are examples of machine learning techniques. The LULC is classified by the training sets using supervised classification. The NDVI (Normalized difference vegetation index) was assessed and used to increase classification accuracy. The LULC classification for the years 2015 to 2021 utilizes multi-temporal images like Sentinel-2, Landsat-8, and MODIS data with spatial resolution of 10 m, 30 m, and 250 m. Agricultural land, Built-up land, Forest land, Fallow land, wasteland, water body and others, are major LULC classes, it lies on a level I classification. According to the findings, the change % of agricultural land is high from 2015 (64.03%) to 2021 (67.81%), this roughly increased about 3.78% during the study year. While water bodies increased by 5.25 to 6.3%. Based on the results, the largest LULC group is agricultural land (122,789.4 km² or 64.03%), followed by forest land (37,678.56 km² or 19.65%). Increased built-up land in the studied area indicates extraordinarily rapid urban growth in major cities of the state. This research offers a reliable approach for comprehensive, automated, and LULC classification in Karnataka State.

Keywords Google earth engine · Land use and land cover · Normalized difference vegetation index · Random Forest · Classification regression trees and support vector machine

Introduction

Land use and land cover maps are essential sources of information for land management and planning (Esfandeh et al. 2022; Yao et al. 2022) Earth Sciences and remote sensing organizations are traditionally interested in accurate and current LULC monitoring (Qian and Zhang 2022; Viana et al.

2019), Primarily because it offers useful data to comprehend relationships between humans and their environment (Praticò et al. 2021). Using single-source, single-temporal satellite images served as the foundation for LULC mapping (Steinhausen et al. 2018). An accurate and timely collection of LULC maps helps in understanding the development of human society and improves climate change and change in the environment modelling, It has a growingly important part in contemporary society (Kuang et al. 2018).

The Spatial map of land use and land cover as well as other earth surface features may now be done quickly and more accurately because of the use of GEE, remote sensing, and GIS technology (Pande 2022), Google Earth, and machine learning classifiers. The GEE platform has offered improved possibilities for simple large satellite data processing and analysis and provided time series data on land cover and usage (Yuan et al. 2005). The most popular technique for assessing land cover and monitoring changes over time

Communicated by: H. Babaie

✉ Harishnaika N
harishnaikn9844@gmail.com

Arpitha M
arpitham317@gmail.com

S A Ahmed
ashfaqa@hotmail.com

¹ Department of Applied Geology, Kuvempu University, Shivamogga, Karnataka 577451, India

is remote sensing photography (Gómez et al. 2016; Wulder et al. 2016), Due to population growth and the requirement to create additional areas to satisfy the demand for On hydrological and water resource modelling community (Roy et al. 2014), is eager to incorporate and assess shifting land use and its effects on the food production to support water budget, water security and energy production,(Sridhar and Wedin 2009).

For large-scale areas, high-resolution land cover mapping takes a massive amount of data for classification in traditional geospatial techniques, so choosing GEE for classification makes it simple to categorize an entire large region. As a result, enormous storage capabilities, powerful processing, and the ability to use a variety of techniques are all necessary (Xie et al. 2019). With the introduction of Google Earth Engine, these demands were met and this kind of technology was made available to everyone without charge (Gorelick et al. 2017; Kolli et al. 2020; Sidhu et al. 2018).

In recent decades, a large number of researchers have numerous researchers have employed various remote sensing imagery for LULC classification throughout the last few decades. For the classification and extraction of cultivated land, scrubland, agriculture, bare soil, and water body (Zhang et al. 2020), suggested an innovative combining classification using SAR and optical images in remote sensing. The overall accuracy was more than 85%. For the extraction and categorization of agriculture, scrubland, farmland water bodies, and bare soil, (Zhang and Roy 2017) suggested a unique fusion (feature-based) classification approach utilizing remote sensing pictures (SAR and Optical). The total accuracy was greater than 85% (Hu and Nacun 2018; Pan et al. 2022). Many investigations on LULC, but the standard approaches only extract a small number of land cover categories, and the classification outcomes with these results are frequently coarse. In reality, there are many different and intricate types of land cover in the fields. Therefore, creating a novel strategy for the specific forms of land cover will be quite useful (Batunacun et al. 2018). One of the important study regions for remotely sensed satellite images is how to execute LULC classification rapidly and reliably (Batunacun et al. 2018). The classic LULC method relies on remote sensing elements in addition to computer-aided image interpretation and visual interpretation as its key bases of support colour, texture, shape information, natural geography, and landforms are every instance of sensing imagery (Petit and Lambin 2001; Singh and Singh 2018). However, typical supervised classification involves substantial expertise and a time-consuming technique for obtaining an extensive amount of features (Chen et al. 2018). In addition, supervised classification (Traditional) using manually created spatial aspects frequently leads to

small sample sizes and weak generalization (Zhao and Du 2016), also restricting the small number of training samples' reliability and accuracy of the classification outcome (Chen et al. 2018). In summary, the manual selection of samples for training is a very lengthy, laborious, and unpredictable operation. To supply sufficient and effective samples in a timely way, substantial effort should therefore be put into choosing and extracting training samples in supervised classification (Attarchi and Gloaguen 2014).The benefits of machine learning (ML) algorithms for different fields of vulnerability studies like flood, drought landslide assessment (Chowdhuri et al. 2022; Saha et al. 2022a; Wen et al. 2022). In this regard, geographic information systems (RS-GIS) have been employed extensively over the past few decades (Saha et al. 2022b). RS-GIS techniques are frequently employed in the investigation of geohazards for their impact assessment studies (Islam et al. 2022; Saha et al. 2022b; Saha et al. 2022a, 2022b).

It is possible to complete thorough and automatic LULC classification inside study regions because of the large storage capacity, powerful processing, and self-programming classification algorithms(Zhao and Du 2016). All of the aforementioned criteria can be fulfilled using Google Earth Engine (GEE), and it's free and available to everyone (Stromann et al. 2020). Without being restricted to consuming (time) conversion, mosaics, resampling, projection and registration, procedures, GEE can quickly process multi-source satellite pictures (Attarchi and Gloaguen 2014; Chowdhuri et al. 2022; Saha et al. 2022a, 2022b). In the current study, LULC classification maps have been generated utilizing remote sensing and GIS and Three advanced machine learning approaches, including SVM, RF, and CART. The effectiveness of the intricate and LULC classification (automated) approach was assessed using the use of machine learning techniques in supervised classification. Except for cartography, all procedures in this study were carried out using the GEE cloud platform. The goal of this research is to suggest a LULC classification technique that can quickly and efficiently supply unmanned training samples with thorough LULC categories. The main objective of the study was to classify LULC using multispectral satellite images from Landsat-8 and Sentinel-2, compare existing machine learning approaches on the GEE platform, and thereby determine the satellite image source and the machine learning technique that result in the most accurate classification.

Study area

The total area of Karnataka is 191,761 Square kilometre. The state is situated between the Latitude 11°31' N to 18°45' N and Longitude 74°12' E to 78°40' E. The Karnataka is divided into 31 districts and 176 taluks.

The Krishna, Cauvery, and Godavari are the three major rivers of Karnataka (Harishnaika et al. 2022), Based on Physiography, there are four zones in the state. In addition, the eastern portion of the state is made up of mountainous areas, coastal regions west of the Western Ghats, and northern and southern plains (Harishnaika and Kumar 2022).

Figure 1 is indicated the Koppen-Geiger climate classification system, which employs 30 types and 5 major classes to categorize the global climate patterns (Moghbeli et al. 2020; Yang et al. 2017), According to this classification it divides the Karnataka state into 4 groups, Namely; Aw (Equatorial Winter Dry), which predominates the region, is distinguished from BSh (Arid Steppe Hot) and tropical monsoon (Am) by various swings in monthly rainfall and temperature (Harishnaika and Ahmed 2023; Harishnaika and Ahmed 2022; Wu et al. 2019). The average annual rainfall of Karnataka is about 1248 mm, and the state’s average annual temperature is about 27 °C. The X and XII agro-climatic zones of Karnataka state are made up of the southern plateau and hills, west coast plain, and Ghats areas (Harishnaika and Ahmed 2022). The different part of the state has little different climate and weather conditions based on their environment and morphology, which controls the tropical climate of the area, you may get one of the greatest descriptions of the monsoon seasons around the world (Kumar et al. 2022). It possesses a high degree of ecological endemism and diversity, making it one of the 8 “hottest hotspots” of biodiversity on Earth. The Western Ghats and its surrounding regions do, however, periodically

have little rainfall, which might cause forest fires in this area (Harishnaika and Ahmed 2022).

Dataset

A platform called Google Earth Engine (GEE) contains a sizable amount of Earth observation widely used systems like MODIS, Landsat, and Sentinel as well as different geospatial platforms including demographic and climate data (Xie et al. 2019; Zhang and Roy 2017) from Landsat and Sentinel can be accessible in GEE through USGS (United States Geological Survey). The Landsat-8 OLI (Operational Land Imager), data for 2015, LULC generation, and the data (2016–2021), all contain information about the land surface cover it showed in Table 1. Sentinel-2 Level-2 C data and Landsat-8 data with atmospheric correction utilizing the Landsat-8 Surface Reflectance Code (LASRC) were employed in the current investigation. Because the cloud cover is less than 5–10% the data were chosen, and all these images had been combined into one image. Landsat-8 (5 bands) and Sentinel-2 (8 bands) were used to categorize the data. For a Landsat-8, the total number of images used was seven in 2015, For Sentinel-2, six in 2016, and nine images in 2017 and 2018, ten images are in 2019–2021.

Each Landsat and Sentinel pixel represented (30 m x30m and 10 m x 10 m) respectively, as the statistical unit. Five main classes were used to LULC namely, Agricultural

Fig. 1 Location and climatic classification of the Karnataka state

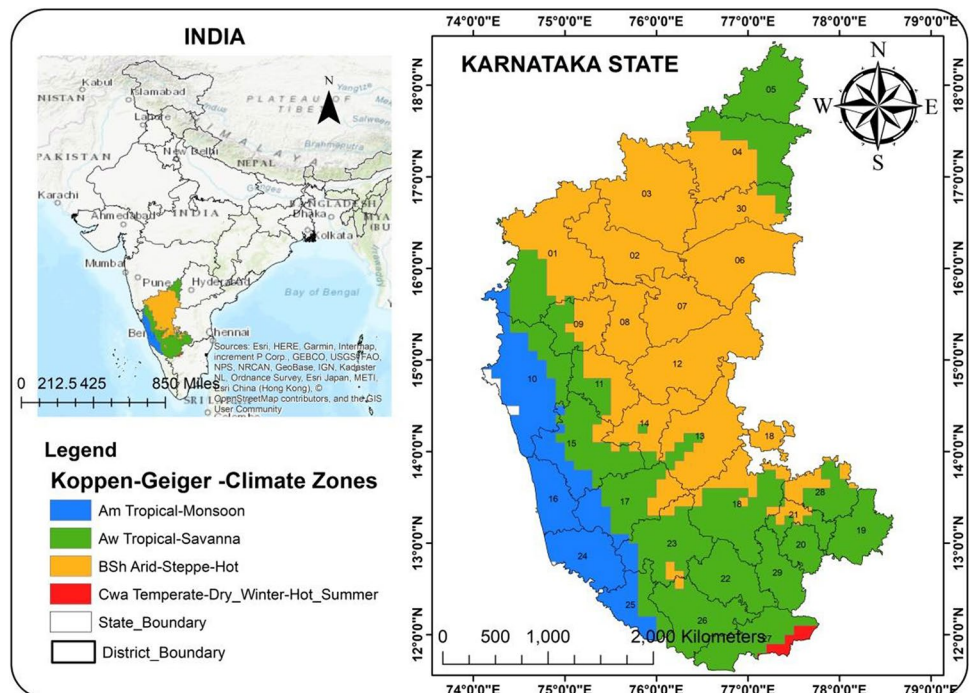


Table 1 Sentinel-2 and Landsat-8 band information, which were used for LULC classification

Data Layer	Source	Bands Used	Band Width (μm)	Central Wavelength (μm)	Spatial Resolution (m)
Landsat-8 (OLI)	Data acquired through the Google Earth Engine (GEE) United States Geological Survey	Blue (Band 2)	0.06	0.48	30
		Green (Band 3)	0.05	0.56	30
		Red (Band 4)	0.03	0.65	30
		Near-Infra-Red (Band 5)	0.02	0.86	30
		Short-Wave Infra-Red 1 (Band 6)	0.08	1.60	30
		Short-Wave Infra-Red 2 (Band 7)	0.18	2.20	30
		Sentinel-2A	Data acquired through the Google Earth Engine (GEE) United States Geological Survey	Blue (Band 2)	0.066
Green (Band 3)	0.036			0.560	10
Red (Band 4)	0.031			0.664	10
Red-Edge 1(Band 5)	0.015			0.704	20
Red-Edge 2 (Band 6)	0.015			0.740	20
Red-Edge 3 (Band 7)	0.020			0.782	20
Near-Infra-Red (Band 8)	0.106			0.835	10
Short-Wave Infra-Red 1 (Band 11)	0.091			1.610	20
Short-Wave Infra-Red 2 (Band 12)	0.175			2.202	20

land, Built-up land, Forest land, Fallow land, Wasteland, Water body and others. Forest and agricultural areas were represented as vegetation, while ponds and rivers were

represented as water areas. The investigation used spectral bands from Landsat-8 images 2 to 7, as well as from Sentinel-2 image bands 2 to 12 (Table 1).

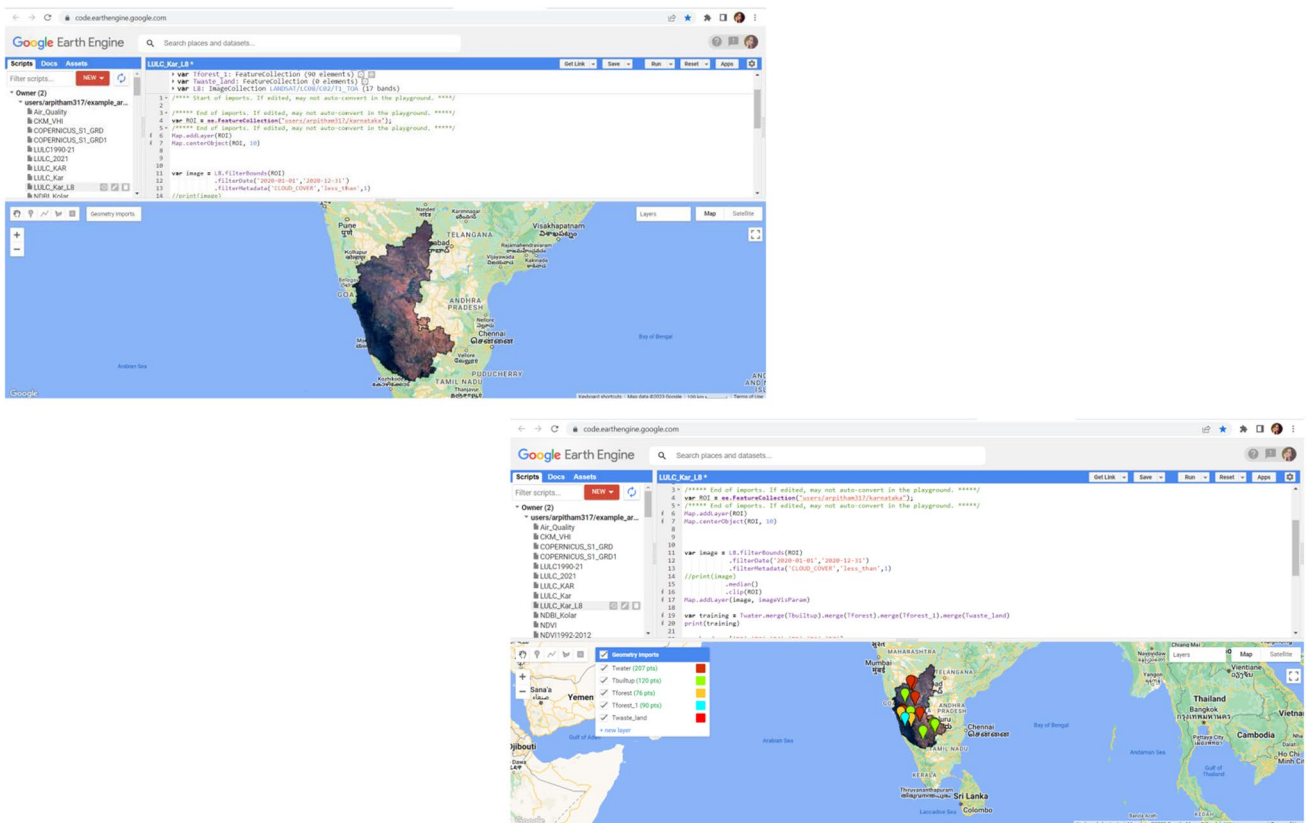


Fig. 2 Location and Training sample selected for LULC classification

Methodology

Our suggested approach, shown in Fig. 2, was implemented in an entire Karnataka state that was wholly dependent on the utilization of the GEE cloud computing setting. We created a consistent time series starting with the images that were readily available in the GEE collection. To identify the optimal input image that offered the maximum accuracy. The Classification of different images according to their period and finally, the multiple classifiers like Random Forest, Support Vector Machine and Classification Regression Trees method were tested and determine which one is best results regarding classification accuracy. The summary of our suggested approach is likely image pre-processing, which includes image picking, filtration, and time series data construction; image processing (Computation of yearly data, image reduction); categorization (acquisition of Validation and Training points, best image for input selection, machine learning methods and factors optimization); accuracy evaluation and comparison; and mapping.

CART (classification regression trees)

CART was a single-tree selection classifier, similar to RF. The defining characteristic of the split is produced by one attribute, which divides the set of data into subgroups at the node of the tree depending on the normalized knowledge gain (Loh 2011). Breiman (1984) created the CART binary decision classification tree, which enables straightforward decision-making in logical if-then scenarios. Based on a predetermined threshold, CART runs recursively by separating nodes until it reaches the terminal nodes. This method divides the input data into group sets, and the trees are built using all but one of those. The reduced tree with the lowest deviation is chosen once the tree has been evaluated using the group that was excluded. The sample size employed in each class will have a significant impact on CART. High dimensionality data, in particular, hinders CART efficacy. To reach a judgment, the character with the highest normalized information gain value is chosen. In GEE, the smallest leaf density and the maximum amount of connections are the only variables that can be changed.

RF (random Forest)

Exact categorization and output combining are used forecast the outcomes (Tumer and Ghosh 1996). To create a new label, the random forest classification model aggregates the results from many decision-making procedures with the highest number of votes (Loukika et al. 2021). To produce

a single tree, the random forest selects a randomly chosen subset of samples. Data from the initial, completed set of training data are sampled for every tree in this bagging procedure (Breiman 1996). RF randomly chooses the factors from training samples to calculate the best split for creating a tree. Although it can weaken individual trees, it weakens the link between trees, which causes a reduced misperception. Random forest gained popularity as a result of its dependability to sound and exceptions (Eisavi et al. 2015). It has been shown that RF can effectively classify various types of land use and land cover (Waske and Braun 2009). The two stronger algorithms, bagging and random, which are referred to as the method's "powerhouse," have helped the RF methodology. The R 'random Forest' package was utilized in our study to create the LULC map.

SVM (support vector machine)

A supervised learning algorithm called an SVM- support vector machine is utilized for solving regression and classification problems. In the training stage, SVM classifiers build a perfect hyperplane that divides several classes with the least misclassified pixels. The extreme points and vectors needed to build the hyperplane are chosen using SVM. Support vectors are the names for these extreme positions. The price factor C, gamma, and parameters serve as the primary selection criteria for support vectors (Hepağuşlar et al. 2004). The C and Gamma variables are defined using the grid search technique, producing accurate prediction outcomes. The cost parameter C significantly affects the performance of SVM and support vector machines. For training on sizable datasets, the simple linear kernel is favoured. Since the goal of SVM is to discover the best separating hyper-plane from the available hyper-planes, the original SVM method was launched with a set of data and its objective was to locate the hyper-plane that could separate the datasets into several classes. Additionally, the SVM algorithm needs an appropriate kernel function to precisely establish the hyper-planes and reduce classification mistakes. The type of kernel that is employed is a crucial component of the SVM approach. The SVM's performance is primarily determined by the kernel size, and its resemblance to a smooth surface is primarily determined by the higher kernel density.

Accuracy assessment

Following the completion of the classification using machine learning techniques, the accuracy of the categorized images was assessed. New separate verification samples that were naturally extracted from excellent quality period data of the GEE platform were

used to assess the accuracy of the ensuing results. To make each land cover type's spatial distribution more clear, the System for Sites for terrestrial environment parameterization from Google Earth was also used.

Quantitatively evaluate the accuracy, the PA (producer's accuracy), the user's accuracy (UA), and OA (Overall accuracy) depending on the standard confusion measures utilized. Despite being a frequently used measure, the j coefficient was not included in the present investigation because it has a strong correlation with general precision (Olofsson et al. 2014). Likewise, an interval of confidence of 95% is shown for each accuracy index after it has been updated.

Results and discussion

Land use/land covers classification applying GEE

The present research uses Sentinel-2 Level-1C (10 m) and Landsat -8 (30 m). Landsat-8 level 1 data to assess the efficacy of several machine learning approaches on LULC and NDVI categorization. Figures 2 and 3 show how Sentinel-2 and Landsat-8 images on the GEE system were used as the preliminary input for machine learning techniques including RF, CART, and SVM (classification) for the LULC from 2015 to 2021. Disturbed pixels due to cloudy situations were removed from any obtainable data employing the cloud mask technique

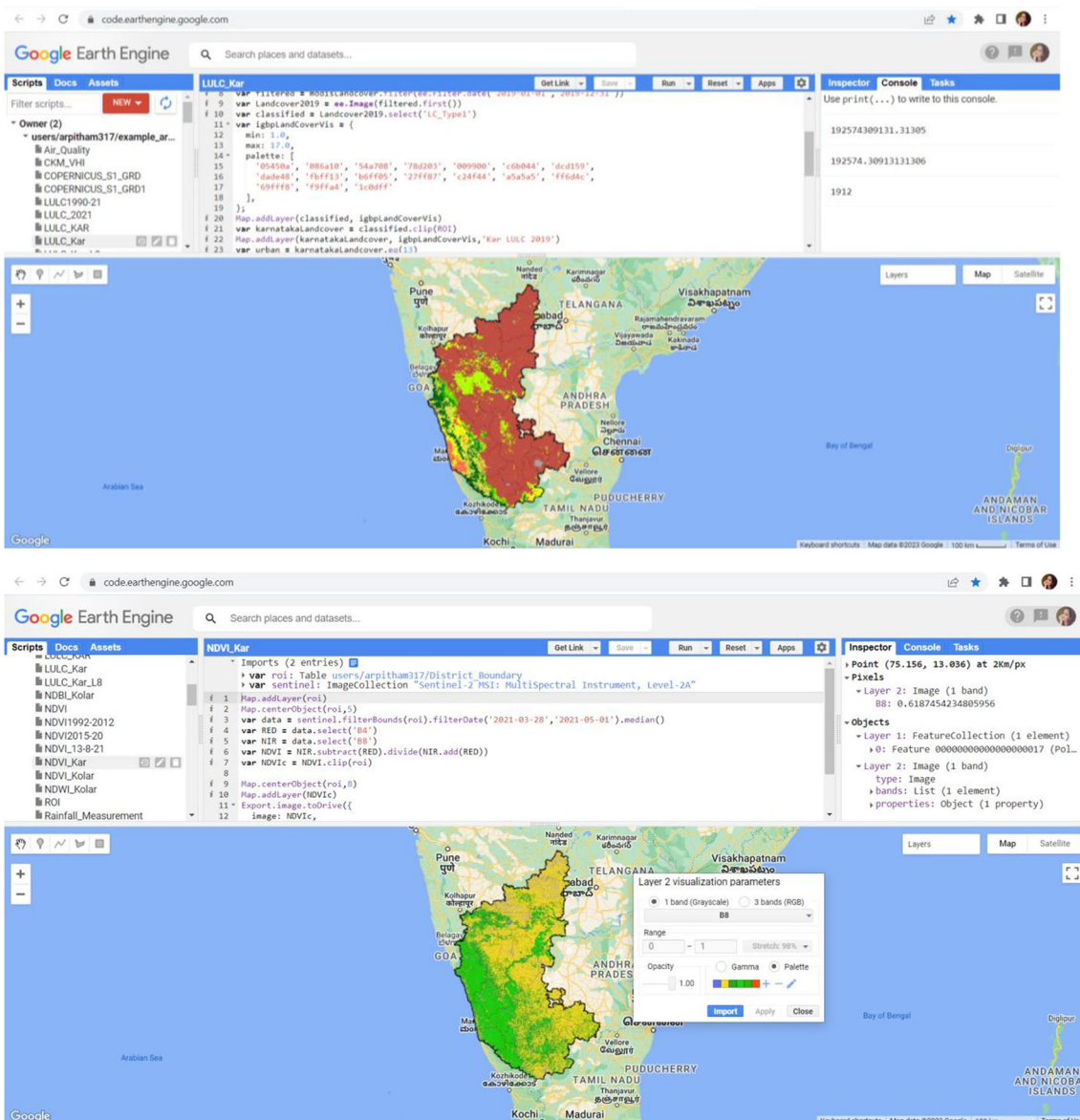


Fig. 3 LULC and NDVI at Different Machine learning Techniques

accessible on the GEE. Temporal aggregation techniques including median, imply that minimum/maximum was utilized to fill in the blanks in cloudy data. For the duration of the investigation, Sentinel-2 and Landsat-8 images were put together using the media NDVI (Normalized Difference Vegetation Index) are commonly employed index that was created and utilized as ad inputs for LULC classification to represent the properties of the vegetation and the water bodies, respectively.

Image observations were used to generate the training and validation datasets. For classification, a total of 1617 training Samples were utilized. Each class should have

at least 567 (RF), 521 (CART), and 527(SVM) training samples for classification, according to a rule of thumb (Loukika et al. 2021). Each class received 66 to 81 samples for verification and 81 to 97 samples for training samples shown in Fig. 2. The identical data used for training and validation were classified using SVM, RF, and CART algorithms. The five main categories of LULC were Agricultural, forest, water body, built-up land, fallow land, wasteland, and others. The optimal cross-validation factor is 5–10 from the experiments, which was utilized as a valuable input for the CART (Belgiu and Drăguț 2016).

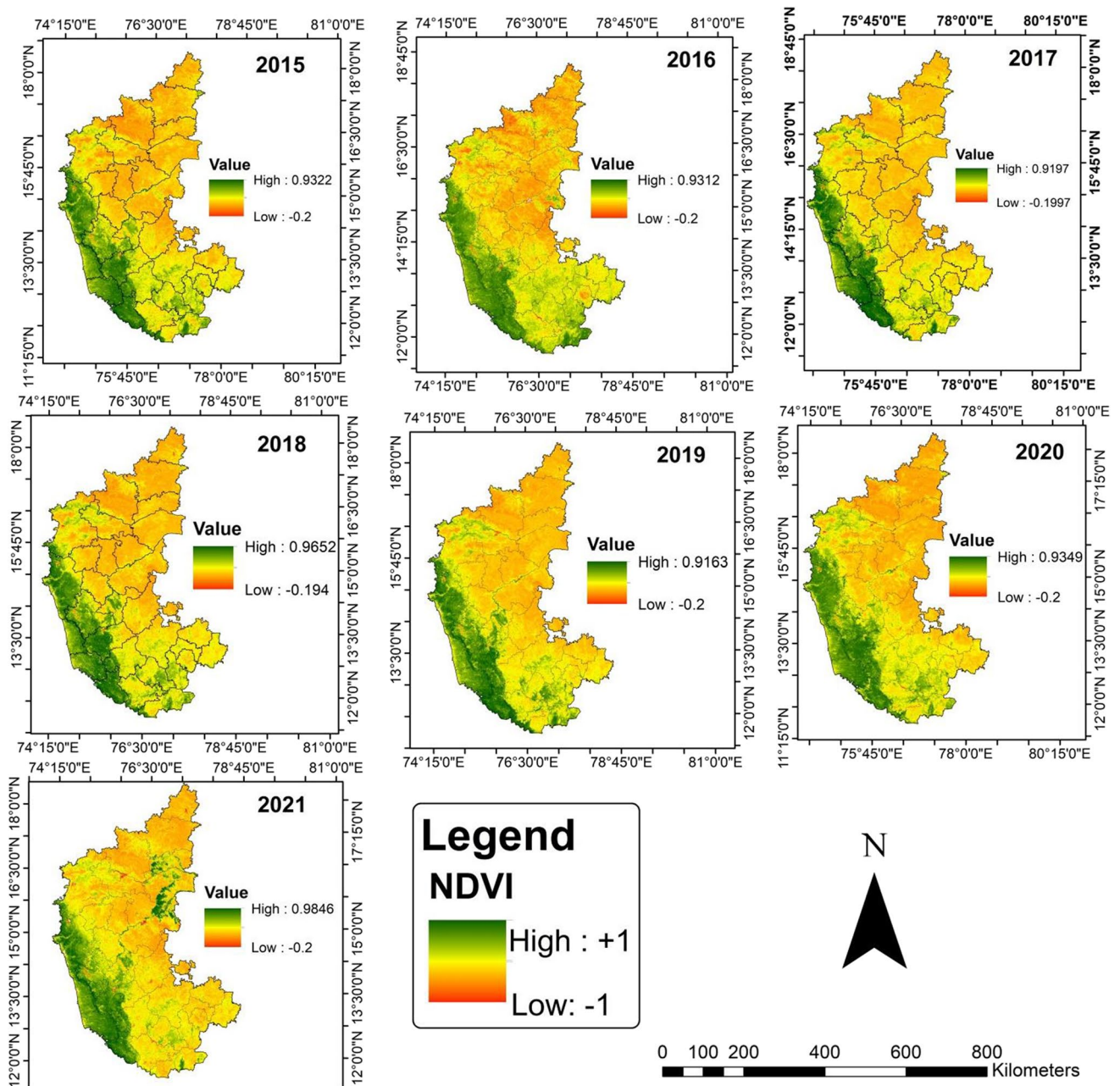


Fig. 4 Spatial pattern of NDVI from 2015 to 2021

For the years 2015 to 2021, maps of land use and cover were created using the random forest (RF) CART and SVM models. To identify LULC classes, therefore, the random forest strategy was effective in terms of results and accuracy. Tables 1 for data sets Sentinel-2 and Landsat-8 band information and Table 3 provide appropriate results of the LULC in Km².

The NDVI method for time-series classification

The suitability of the NDVI approach for overtime classification using the Landsat time series was also demonstrated in this study. Using MODIS imagery, the NDVI had previously achieved extremely precise values for the satellite time series indicated in Fig. 3. Demonstrating the classifier's ability to handle multidimensional analysis to some extent. The

Fig. 5 Temporal pattern of NDVI (mean, maximum and minimum) from 2015 to 2021

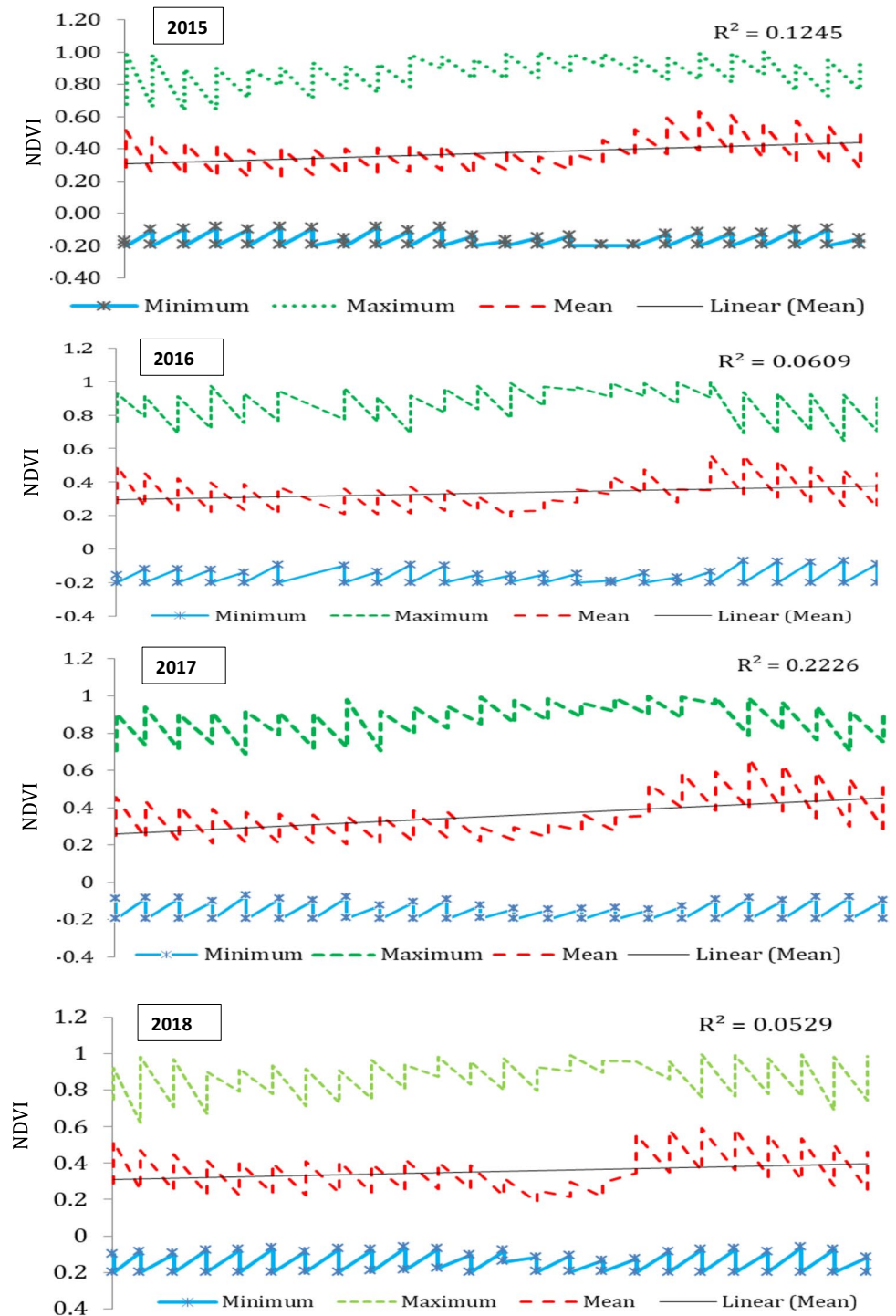
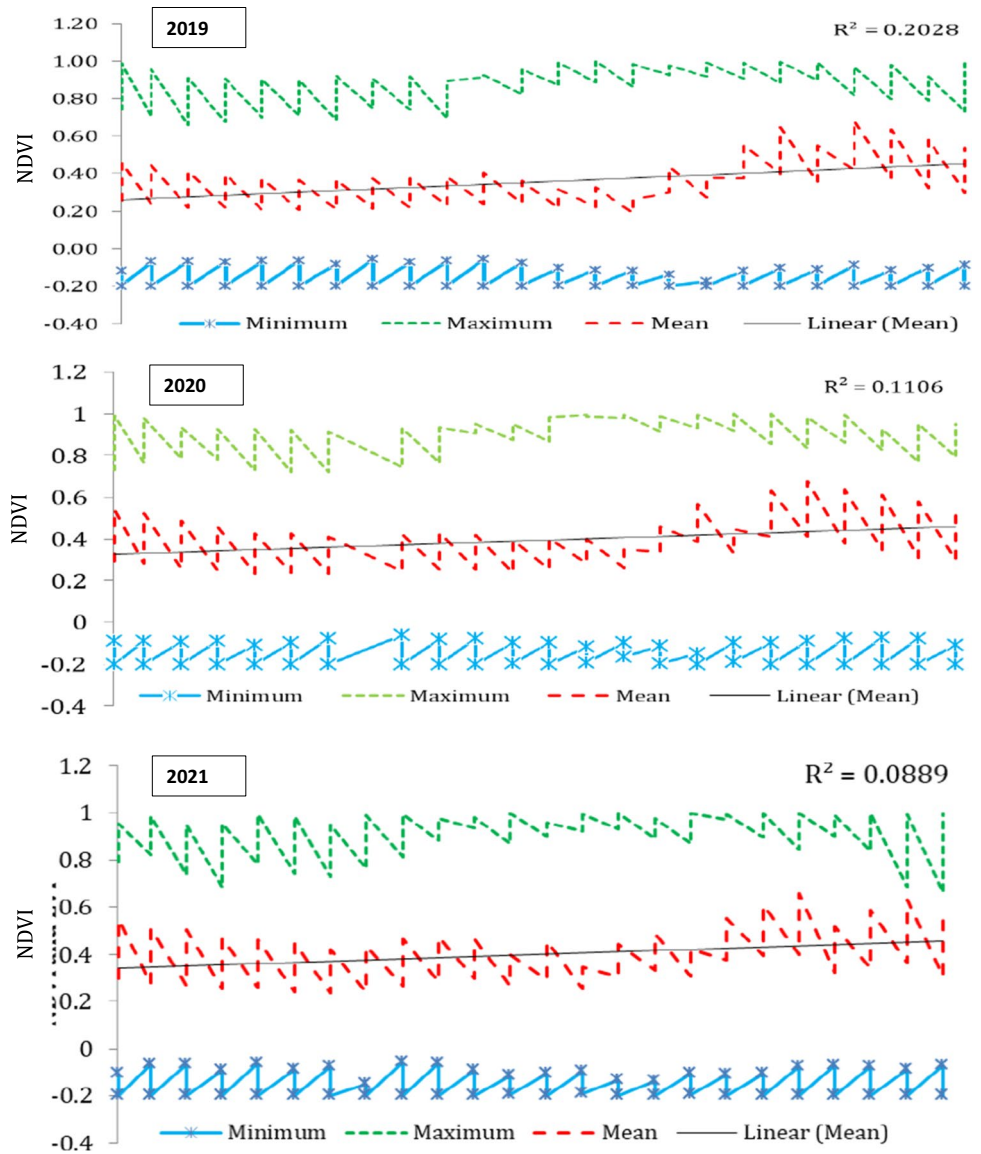


Fig. 5 (continued)



assessment in this study showed a good overall accuracy of roughly 74%. The accuracy values recorded for each of the agricultural land categories (the agricultural classes were prevalent in the research area, representing roughly 78% of the total area) were significant, even though the water

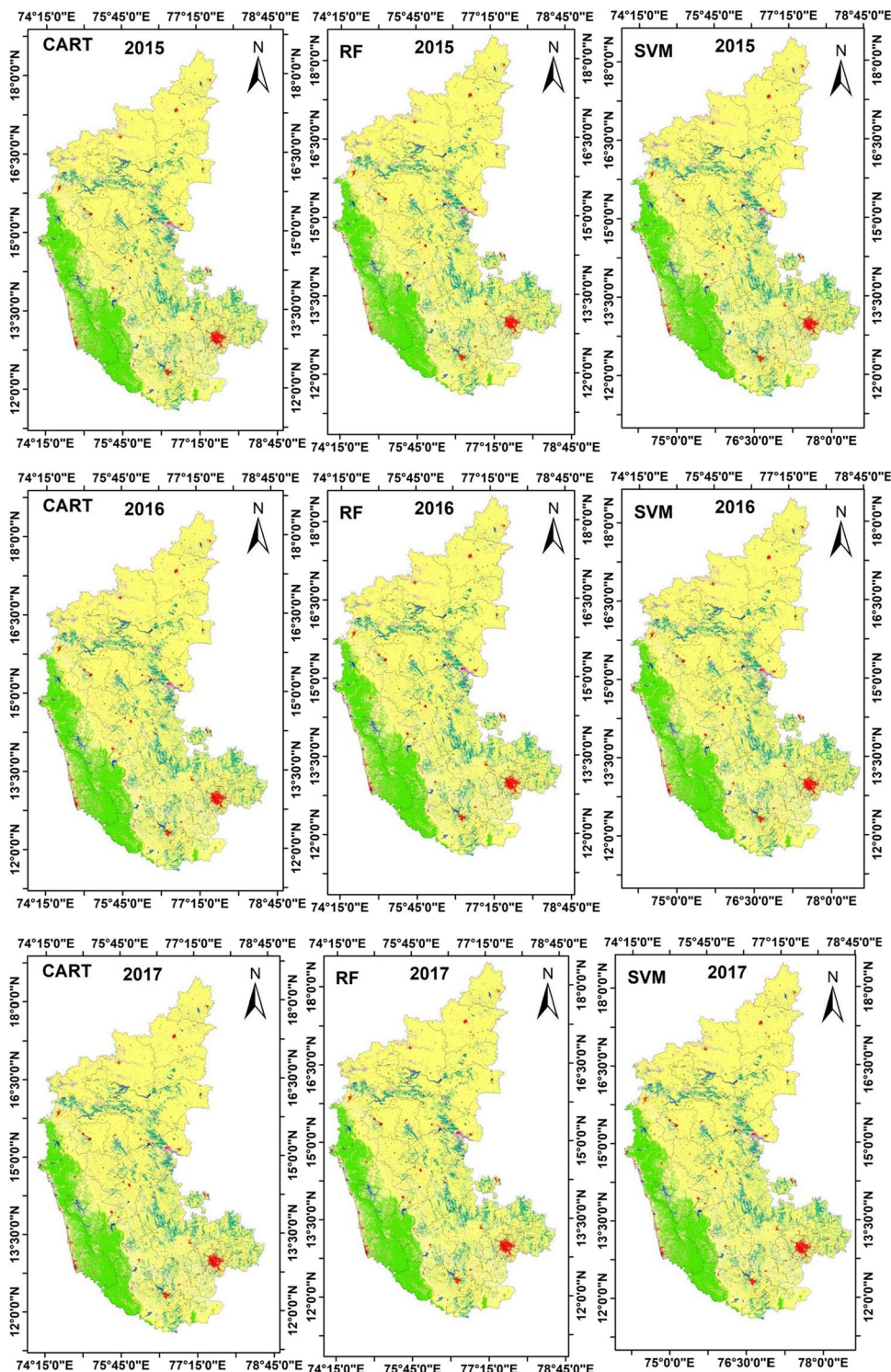
bodies and forest categories contributed to this total value. We choose to calculate the series index (NDVI) and carry out the cluster analysis utilizing the spectral characteristics of the training data for all LULC classes to enhance the recognition of each LULC type and boost the classification

Table 2 Results of the Simple linear regression (SLR) For NDVI

Description	Year	R^2	Adjusted R^2	MSE	RMSE	p Value
NDVI	2015	0.12	0.172	0.002	0.046	<0.001
	2016	0.06	0.014	0.007	0.086	<0.001
	2017	0.22	0.169	0.010	0.101	<0.001
	2018	0.05	0.069	0.009	0.096	<0.001
	2019	0.20	0.047	0.012	0.108	<0.001
	2020	0.11	0.117	0.004	0.061	<0.001
	2021	0.08	0.144	0.005	0.073	<0.001
Total Mean		0.076	-0.155	0.009	0.094	<0.001

Annual Negative trend of NDVI in SLR

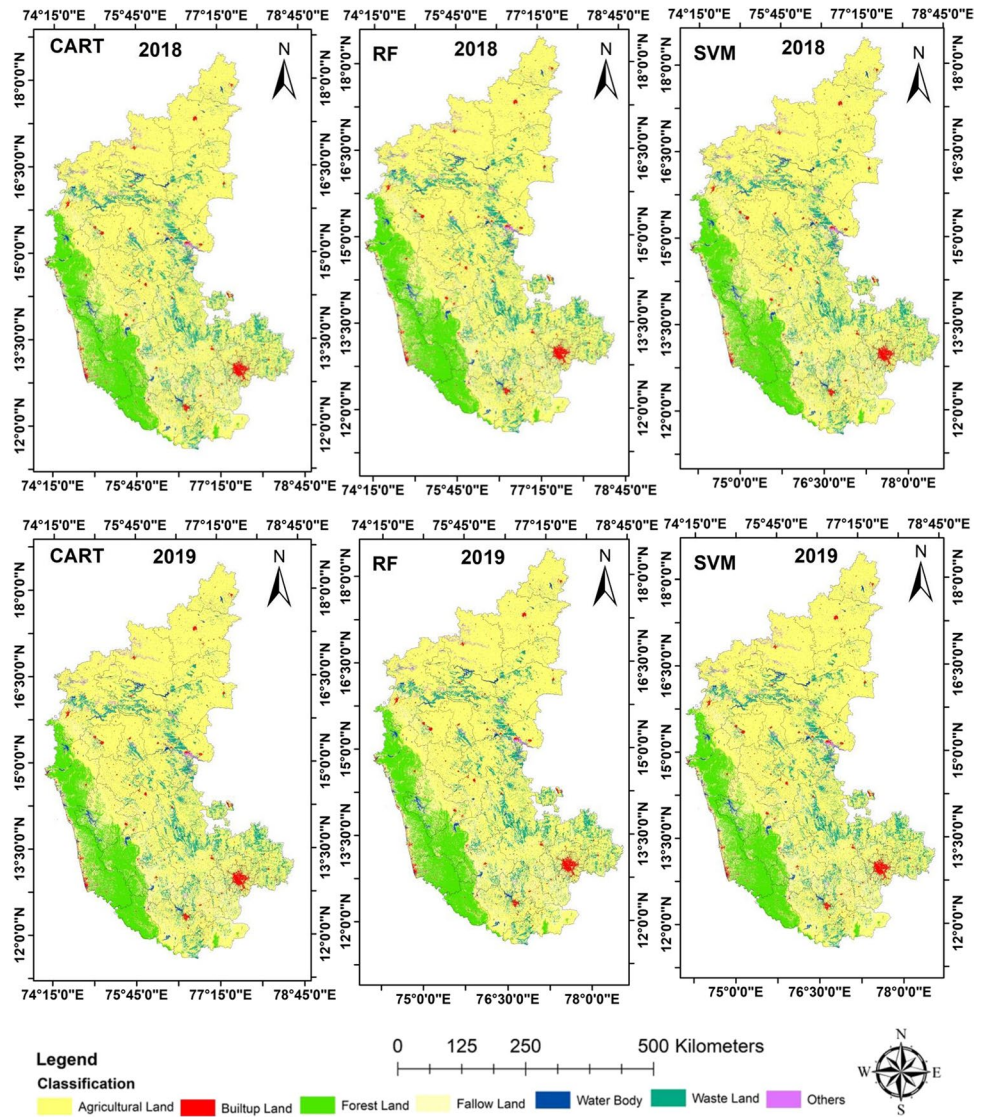
Fig. 6 Spatial pattern of LULC with respect to CART, RF and SVM



accuracy of the relationship between the NDVI and LULC classes were indicated in Fig. 4. Surprisingly, the spectral characteristics of the different vegetation types observed in agriculture classes meant such confusion within agriculture land classification. This is particularly true for short-term

crops in agroforestry settings. They possess a Complex agricultural pattern with seasonal production and harvest of short-term irrigated and rain-fed crops, followed by bare soil that is readily mistaken for long-term farmland and permanent pastures.

Fig. 6 (continued)

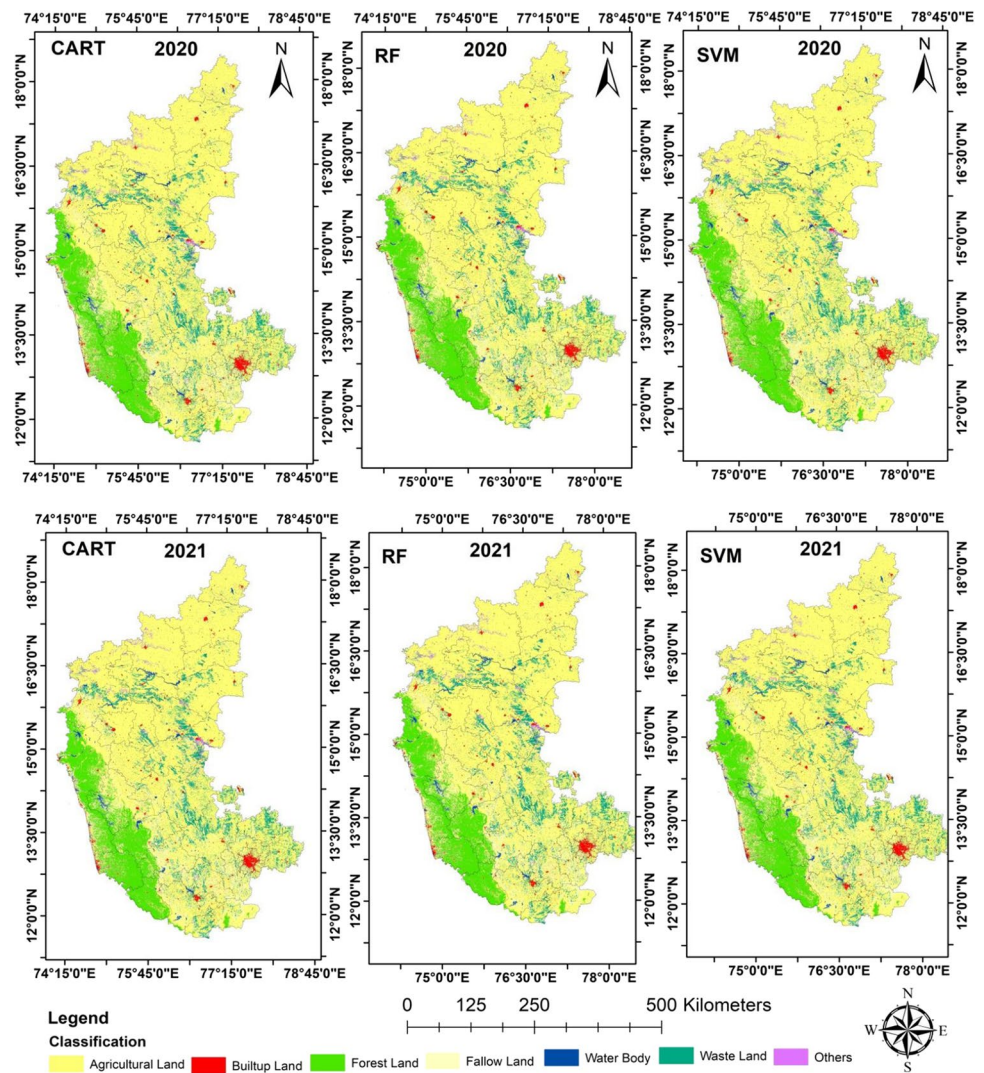


Linear regression model for NDVI

The measure of how closely two datasets are associated is the correlation coefficient (R^2), often known as the Pearson correlation coefficient. Figure 5 and Table 2 Indicated the temporal pattern of NDVI concerning LULC. In the current investigation, R^2 was used to demonstrate the level of correlation between the modeled and observed data sets. The current study also used the root mean square error (RMSE) to examine the discrepancy between forecasts and observations. The RMSE was used to assess the relative NDVI variability as well as the error of the predicted in Table 2. Using the Root Mean Square Error (RMSE), Mean square of the errors (MSE), Mean Absolute Percentage Error (MAPE), and R^2 values displayed in Table 2, a detailed description of the

geographic variation of the modelled concentration of NDVI is given. From 2015 to 2021, the coefficient value (R^2) has a maximum of $R^2=0.22$ and a lowest value of $R^2=0.05$ for the years 2017 and 2018, respectively. The yearly mean, maximum variation, and minimum variance in NDVI were shown in Fig. 5 for the entire state. In 2015 and 2020, an average linear trend with R^2 values of 0.12 and 0.11 was observed. In addition, 2021 and 2016 will show a median linear trend with a coefficient of determination of 0.08, which will be comparable to 2018. Karnataka had the greatest values for the NDVI maximum, minimum, and mean correlations ($R^2=0.169$, $RMSE=0.101$, $MSE=0.010$, and $P=0.001$). NDVI and ($R^2=0.014$ and $RMSE=0.086$, $MSE=0.007$, and $P=0.001$ respectively) have modest correlations, but they nevertheless fulfill the model indices in Table 2.

Fig. 6 (continued)



LULC pattern from 2015 to 2021

Figure 6 displays the Landsat 8 and sentinels –2 derived LULC map layout for 2015–2021 together with a table showing the percentages of various land use groups offer data for the year 2015 on the producer’s overall

accuracy, the kappa coefficient, and the arrangement of land types by Square kilometre. Table 3 showed the Results of the LULC in Square Kilometre, the results of this study show that Agricultural land (122,789.4 km² or 64.03 of the total investigated area), is the biggest LULC group, Followed by Forest land (37,678.56 km²

Table 3 Results of the LULC in Km²

Area in Km ²	2015	2016	2017	2018	2019	2020	2021	Results
Agricultural land	122,789.4	122,567.9	127,900.9	127,600.2	126,788.8	125,976.9	126,193.78	Increases
Built-up land	4023.46	4027.88	4071.47	4076.75	4091.53	4133.61	4208.31	Increases
Fallow land	251.02	343.94	303.91	324.47	321.44	279.75	231.11	Decreases
Forest land	37,678.56	37,456.45	37,347.89	37,876.65	38,019.11	38,115.23	38,453.98	Increases
Others	2463.86	2424.33	2389.06	2348.75	2341.33	2298.55	2107.84	Decreases
Waste land	13,897.63	13,884.68	13,859.85	13,835.03	13,819.41	13,776.37	13,646.45	Decreases
Water body	10,060.7	10,530.8	10,450.7	10,790.7	11,230.7	11,654.7	11,789.7	Increases

Table 4 Results of the LULC Change Percentage %

Area in Km ²	LULC Percentage Change (%)						
	2015	2016	2017	2018	2019	2020	2021
Agricultural land	64.03	63.92	66.70	66.54	66.12	65.69	67.81
Built-up land	2.10	2.10	2.12	2.13	2.13	2.16	2.19
Fallow land	0.13	0.18	0.16	0.17	0.17	0.15	0.12
Forest land	19.65	19.53	19.48	19.75	19.83	19.88	20.05
Others	1.28	1.26	1.25	1.22	1.22	1.20	1.15
Waste land	7.25	7.24	7.23	7.21	7.21	7.18	6.12
Water body	5.25	5.49	5.45	5.63	5.86	6.08	6.34

or 19.65% of the total region) is the second-biggest land use group. The Wasteland (13,897.63 km², or 7.25% of the total land), water bodies (10,060.7 km², or 5.25% of the total land), and development area (4023.46 km²) make up the other three land use categories, the Table 4 indicated that Results of the LULC Change Percentage %. Throughout the study period, agriculture and forest areas predominate; however, there is an increase in water bodies land due to the construction of new canals and water bodies from the government plan; these increases are about 10,060.7, 10,530.8, 10,450.7, 11,230.7, 11,654.7, and 11,789.7 from 2015 to 2021, respectively in Table 3. The area of built-up and agricultural land is expanding, while the area of wasteland is decreasing from 2015 to 2021. On an annual basis from 2015 to 2021, LULC maps were produced for the entirety of Karnataka. Nearly 7.5% of the land region in Karnataka had a change in LULC, which served as a defining characteristic of the changing patterns in Table 4. As demonstrated in these case studies, time-series maps of LULC were created using RS images and then used to remove the veil of mystery around the changes that the surface of the planet had undergone. Figure 6 notified the spatial pattern of LULC concerning CART, RF, and SVM.

The accuracy assessment

In this investigation, stratified random sampling was the method of choice, and it was carried out individually for each type of LULC. Using the RF, CART, And SVM classification methods, an overall of 567, 521, and 527 validating samples and training samples 221, 203 and 199 respectively from Karnataka districts were extracted, respectively. For the Landsat 8 data in 2015–2021, the overall accuracy in the CART Model is 84.01%, and the Kappa Coefficient is 0.704%. For the classification of Sentinel-2 images (Fig. 7

and Table 5), the overall accuracy is about 85.17%, and the Kappa Coefficient is 0.723%. The average overall accuracy in the RF Model for the Landsat data is 93.98%, and the kappa Coefficient is 0.902%. For the classification of sentinel 2 images, the overall accuracy is around 95.157%, and the kappa Coefficient is 0.916%. The mean value of the overall precision in the SVM Model for the Landsat data is 89.00%, and the Kappa Coefficient is 0.842%. For the classification of sentinel 2 images, the overall accuracy is approximately 92.13%, and the Kappa Coefficient is 0.88%. The study area and classifiers' overall accuracy (%) Kappa coefficient can be seen in the figure. The Overall Accuracy (%) based on the RF classifier for the state of Karnataka is consistently greater compared to CART and SVM. Figure 7 showed the temporal variations of accuracy in %. Table 5 indicated the Overall accuracy in the Kappa coefficient of Sentinel and Landsat for machine learning methods from 2015 to 2021.

As a result, for effective land use management and planning, which can be supported by RS (remote sensing) and GIS-based technologies, it is essential to have a greater comprehension of the many connections between changing land use/cover and rural inhabitants' livelihoods. Today, scientists and researchers are working to improve already-existing methods, cloud platforms, and tools for geographic information systems. It is a valuable setting for performing advanced analysis and classification procedures as well as producing multi-temporal composite images. The NDVI and land use maps of various times have been estimated using the Google Earth engine algorithm in the region. Using the GEE machine learning system, Erdas Imagine 2014, and ArcGIS 10.8 software, a change detection mapping has been created. Because of this, change detection mapping is a very popular and effective application of GIS software. There has been a significant change in many courses of LULC over the past five years.

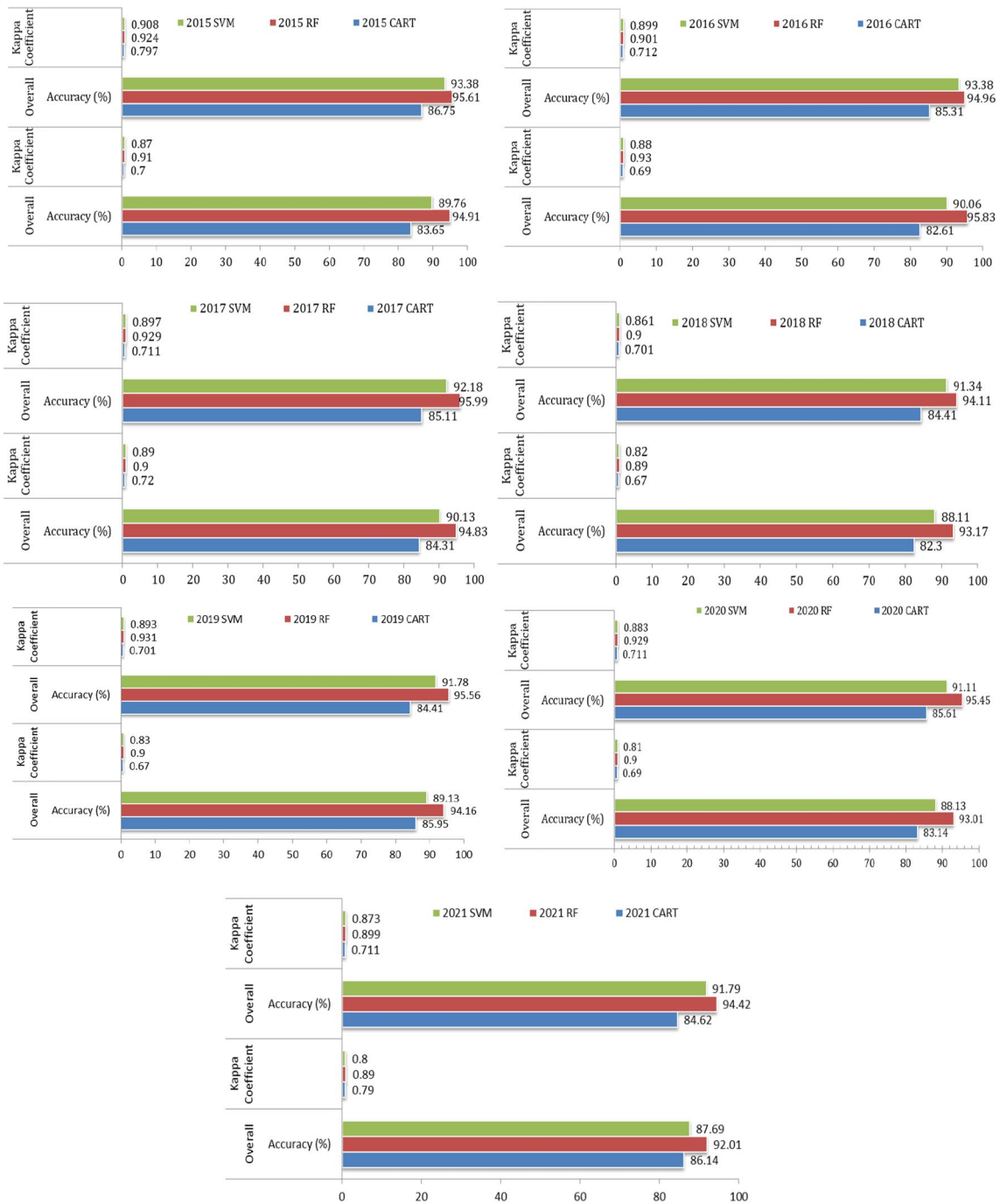


Fig. 7 Accuracy Assessment of LULC (SVM, RF and CART) in %

Conclusions

The state of Karnataka area’s LULC class change has been detected over six years via remote sensing and GIS technology. The study was carried out in a Karnataka state for the classification of LULC using advanced machine learning techniques in the GEE platform.

1. Landsat-8 satellite and Sentinel-2 were used for study over 6 years, and the effectiveness of RF, SVM, and CART techniques for the classification of LULC on the GEE platform was examined. The reliability of LULC data classification from satellite photos was influenced by the type of classifier that was employed. The efficacy of each classification concerning each

Table 5 Overall accuracy in Kappa coefficient of Sentinel-2 and Landsat-8 for machine learning methods

Year	Method	Landsat-8		Sentinal-2	
		Overall Accuracy (%)	Kappa Coefficient	Overall Accuracy (%)	Kappa Coefficient
2015	CART	83.65	0.70	86.75	0.797
	RF	94.91	0.91	95.61	0.924
	SVM	89.76	0.87	93.38	0.908
2016	CART	82.61	0.69	85.31	0.712
	RF	95.83	0.93	94.96	0.901
	SVM	90.06	0.88	93.38	0.899
2017	CART	84.31	0.72	85.11	0.711
	RF	94.83	0.90	95.99	0.929
	SVM	90.13	0.89	92.18	0.897
2018	CART	82.30	0.67	84.41	0.701
	RF	93.17	0.89	94.11	0.900
	SVM	88.11	0.82	91.34	0.861
2019	CART	85.95	0.67	84.41	0.701
	RF	94.16	0.90	95.56	0.931
	SVM	89.13	0.83	91.78	0.893
2020	CART	83.14	0.69	85.61	0.711
	RF	93.01	0.90	95.45	0.929
	SVM	88.13	0.81	91.11	0.883
2021	CART	86.14	0.79	84.62	0.711
	RF	92.01	0.89	94.42	0.899
	SVM	87.69	0.80	91.79	0.873

class can be assessed using the accuracy evaluation of each class.

2. A conservative approach would not be likely to try to compute both temporal and spatial phenomena without the use of multitemporal satellite data. The results show that in six years, agricultural land increased by 64.03–67.81 (3.78%), while water bodies increased by 5.25 to 6.3. The increased built-up land in the study region implies extraordinary spatial development of the area. A total of 567, 521, and 527 validating samples from Karnataka districts were extracted, respectively, using the RF, CART, and SVM classification algorithms.
3. According to the findings of this study, agricultural land (122,789.4 km² or 64.03%) is the largest LULC group, followed by forest land (37,678.56 km² or 19.65%).
4. The government’s new development strategies and plans, the LULC and NDVI images show that the quantity of forest and agricultural land increased between 2015 and 2021. The NDVI coefficient value (R²) for the years 2017 and 2018 has a maximum value of R²=0.22 and a lowest value of R²=0.05 from 2015 to 2021.
5. The linear trend in NDVI with R² values of 0.12 and 0.11 was detected between 2015 and 2020. Furthermore, 2021

and 2016 will exhibit an average linear trend with a coefficient of determination of 0.08, comparable to 2018.

For LULC classification, the benefits of the GEE (cloud platform) are convenient and adaptable, especially for a large number of input features. We were able to overcome the limitations of desktop systems by using the GEE to speed up geospatial large data collecting and processing the datasets.

Acknowledgments The authors would like to thank the Kuvempu University, Shankaraghatta for providing the research fellowship during this research. The authors are grateful to the Department of Applied Geology, Kuvempu University for technical and moral support.

Authors contribution 1. Arpitha M- conceptualization, formal analysis, investigation, software, writing original draft, figures, and tables preparation

Gmail- arpitham317@gmail.com

2. S A Ahmed- guided, investigation, writing—Review, validation, and editing

Gmail- ashfaqsa@hotmail.com

3. Harishnaika N* - data interpretation, manuscript writing, software handling, writing original draft, figures, and tables preparation, conceptualization, formal analysis and editing

Gmail- harishnaikan9844@gmail.com

Funding No funding (not applicable).

Data availability The data that support the findings of this study are available on request from the corresponding author.

Declarations

Conflict of interest The authors declare no competing interests.

Consent for publications Not applicable.

References

- Attarchi S, Gloaguen R (2014) Classifying complex mountainous forests with L-band SAR and Landsat data integration: A comparison among different machine learning methods in the Hyrcanian Forest. *Remote Sens* 6(5):3624–3647. <https://doi.org/10.3390/rs6053624>
- Batunacun N, Hu Y, Lakes T (2018) Land-use change and land degradation on the Mongolian plateau from 1975 to 2015—A case study from Xilingol, China. *Land Degrad Dev* 29(6):1595–1606. <https://doi.org/10.1002/ldr.2948>
- Belgiu M, Drăguț 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>
- Breiman, L (1984) Classification and regression trees (1st ed.). Routledge. <https://doi.org/10.1201/9781315139470>
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140. <https://doi.org/10.1007/BF00058655>
- Chen Y, Zhou Y, Ge Y, An R, Chen Y (2018) Enhancing land cover mapping through integration of pixel-based and object-based classifications from remotely sensed imagery. *Remote Sens* 10(1). <https://doi.org/10.3390/rs10010077>
- Chowdhuri I, Pal SC, Saha A, Roy P, Chakraborty R, Shit M (2022) Application of novel framework approach for assessing rainfall induced future landslide hazard to world heritage sites in Indo-Nepal-Bhutan Himalayan region. *Geocarto Int* 37(27):17742–17776. <https://doi.org/10.1080/10106049.2022.2134464>
- Eisavi V, Homayouni S, Yazdi AM, Alimohammadi A (2015) Land cover mapping based on random forest classification of multi-temporal spectral and thermal images. *Environ Monit Assess* 187(5):291. <https://doi.org/10.1007/s10661-015-4489-3>
- Esfandeh, S, Danehkar, A, Salmanmahiny, A, Sadeghi, SMM, Marcu, MV (2022) Climate change risk of urban growth and land use/land cover conversion: An In-Depth Review of the Recent Research in Iran Sustainability, 14(1). <https://doi.org/10.3390/su14010338>
- Gómez C, White JC, Wulder MA (2016) Optical remotely sensed time series data for land cover classification: A review. *ISPRS J Photogramm Remote Sens* 116:55–72. <https://doi.org/10.1016/j.isprsjprs.2016.03.008>
- 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>
- Harishnaika, A, Kumar, S (2022) Spatio-temporal rainfall trend assessment over a semi-arid region of Karnataka state, using non-parametric techniques. *Arab J Geosci*, 15(16). <https://doi.org/10.1007/s12517-022-10665-7>
- Harishnaika, Ahmed, SA (2022) Analysis of drought severity and vegetation condition prediction using satellite remote sensing indices in Kolar and Chikkaballapura districts, Karnataka state. *Research Square*. <https://doi.org/10.21203/rs.3.rs-1764783/v1>
- Harishnaika, N, Ashfaq Ahmed, S, Kumar, S, Arpitha M (2022) Computation of the spatio-temporal extent of rainfall and long-term meteorological drought assessment using standardized precipitation index over Kolar and Chikkaballapura districts, Karnataka during 1951–2019. *Remote Sens Appl Soc Environ*, 100768. <https://doi.org/10.1016/j.rsase.2022.100768>
- Harishnaika SN, Ahmed SA (2023) Detection of spatiotemporal patterns of rainfall trends, using non-parametric statistical techniques, in Karnataka state. *India Environ Monit Assess* 195(7):909. <https://doi.org/10.1007/s10661-023-11466-5>
- Hepağuşlar H, Özzeybek D, Özkardeşler S, Taşdoğan A, Duru S, Elar Z (2004) Propofol and sevoflurane during epidural/general anesthesia: comparison of early recovery characteristics and pain relief. *Middle East J Anesthesiol* 17(5):819–832
- Hu, Y, Nacun, B (2018) An analysis of land-use change and grassland degradation from a policy perspective in Inner Mongolia, China, 1990–2015. *Sustain*, 10(11). <https://doi.org/10.3390/su10114048>
- Islam ARMT, Saha A, Ghose B, Pal SC, Chowdhuri I, Mallick J (2022) Landslide susceptibility modeling in a complex mountainous region of Sikkim Himalaya using new hybrid data mining approach. *Geocarto Int* 37(25):9021–9046. <https://doi.org/10.1080/10106049.2021.2009920>
- Kolli, MK, Opp, C, Karthe, D, Groll, M (2020) Mapping of major land-use changes in the Kolleru Lake freshwater ecosystem by using Landsat satellite images in Google earth engine. *Water*, 12(9). <https://doi.org/10.3390/w12092493>
- Kuang W, Yang T, Yan F (2018) Examining urban land-cover characteristics and ecological regulation during the construction of Xiong’ an New District, Hebei Province. *China J Geograph Sci* 28(1):109–123. <https://doi.org/10.1007/s11442-018-1462-4>
- Kumar S, Ahmed SA, Harishnaika N et al (2022) Spatial and temporal pattern assessment of meteorological drought in Tumakuru District of Karnataka during 1951–2019 using standardized precipitation index. *J Geol Soc India* 98:822–830. <https://doi.org/10.1007/s12594-022-2073-3>
- Loh W-Y (2011) Classification and regression trees. *WIREs Data Min Knowl Discov* 1(1):14–23. <https://doi.org/10.1002/widm.8>
- Loukika, KN, Keesara, VR, Sridhar, V (2021) Analysis of land use and land cover using machine learning algorithms on google earth engine for Munneru river basin, India. *Sustain (Switzerland)*, 13(24). <https://doi.org/10.3390/su132413758>
- Moghbeli, A, Delbari, M, Amiri, M (2020) Application of a standardized precipitation index for mapping drought severity in an arid climate region, southeastern Iran. *Arab J Geosci*, 13(5). <https://doi.org/10.1007/s12517-020-5201-7>
- 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>
- Pan X, Wang Z, Gao Y, Dang X, Han Y (2022) Detailed and automated classification of land use/land cover using machine learning algorithms in Google earth engine. *Geocarto Int* 37(18):5415–5432. <https://doi.org/10.1080/10106049.2021.1917005>
- Pande CB (2022) Land use/land cover and change detection mapping in Rahuri watershed area (MS), India using the google earth engine and machine learning approach. *Geocarto Int* 37(26):13860–13880. <https://doi.org/10.1080/10106049.2022.2086622>
- Petit CC, Lambin EF (2001) Integration of multi-source remote sensing data for land cover change detection. *Int J Geogr Inf Sci* 15(8):785–803. <https://doi.org/10.1080/13658810110074483>
- Praticò, S, Solano, F, Di Fazio, S, Modica, G (2021) Machine learning classification of Mediterranean Forest habitats in Google earth engine based on seasonal Sentinel-2 time-series and input image

- composition optimisation. *Remote Sens*, 13(4). <https://doi.org/10.3390/rs13040586>
- Qian X, Zhang L (2022) An integration method to improve the quality of global land cover. *Adv Space Res* 69(3):1427–1438. <https://doi.org/10.1016/j.asr.2021.11.002>
- Roy DP, Wulder MA, Loveland TR, Woodcock CE, Allen RG, Anderson MC et al (2014) Landsat-8: science and product vision for terrestrial global change research. *Remote Sens Environ* 145:154–172. <https://doi.org/10.1016/j.rse.2014.02.001>
- Saha A, Pal SC, Chowdhuri I, Chakraborty R, Roy P (2022a) Understanding the scale effects of topographical variables on landslide susceptibility mapping in Sikkim Himalaya using deep learning approaches. *Geocarto Int* 37(27):17826–17852. <https://doi.org/10.1080/10106049.2022.2136255>
- Saha A, Pal SC, Chowdhuri I, Islam ARMT, Chakraborty R, Roy P (2022c) Threats of soil erosion under CMIP6 SSPs scenarios: an integrated data mining techniques and geospatial approaches. *Geocarto Int* 37(27):17307–17339. <https://doi.org/10.1080/10106049.2022.2127925>
- Saha A, Pal SC, Chowdhuri I, Islam ARMT, Chakraborty R, Roy P (2022d) Application of neural network model-based framework approach to identify gully erosion potential hotspot zones in subtropical environment. *Geocarto Int* 37(26):14758–14784. <https://doi.org/10.1080/10106049.2022.2091042>
- Saha A, Pal SC, Chowdhuri I, Roy P, Chakraborty R (2022b) Effect of hydrogeochemical behavior on groundwater resources in Holocene aquifers of moribund Ganges Delta, India: infusing data-driven algorithms. *Environ Pollut* 314:120203. <https://doi.org/10.1016/j.envpol.2022.120203>
- Sidhu N, Pebesma E, Câmara G (2018) Using Google earth engine to detect land cover change: Singapore as a use case. *Eur J Remote Sens* 51(1):486–500. <https://doi.org/10.1080/22797254.2018.1451782>
- Singh A, Singh KK (2018) Unsupervised change detection in remote sensing images using fusion of spectral and statistical indices. *Egypt J Remote Sens Space Sci* 21(3):345–351. <https://doi.org/10.1016/j.ejrs.2018.01.006>
- Sridhar V, Wedin DA (2009) Hydrological behaviour of grasslands of the Sandhills of Nebraska: water and energy-balance assessment from measurements, treatments, and modelling. *Ecohydrology* 2(2):195–212. <https://doi.org/10.1002/eco.61>
- Steinhausen MJ, Wagner PD, Narasimhan B, Waske B (2018) Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. *Int J Appl Earth Obs Geoinf* 73:595–604. <https://doi.org/10.1016/j.jag.2018.08.011>
- Stromann, O, Nascetti, A, Yousif, O, Ban, Y (2020) Dimensionality reduction and feature selection for object-based land cover classification based on Sentinel-1 and Sentinel-2 time series using Google earth engine. *Remote Sens*, 12(1). <https://doi.org/10.3390/rs12010076>
- Tumer K, Ghosh J (1996) Analysis of decision boundaries in linearly combined neural classifiers. *Pattern Recogn* 29(2):341–348. [https://doi.org/10.1016/0031-3203\(95\)00085-2](https://doi.org/10.1016/0031-3203(95)00085-2)
- Viana CM, Girão I, Rocha J (2019) Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. *Remote Sens* 11(9). <https://doi.org/10.3390/rs11091104>
- Waske B, Braun M (2009) Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS J Photogramm Remote Sens* 64(5):450–457. <https://doi.org/10.1016/j.isprsjprs.2009.01.003>
- Wen T, Tiewang W, Arabameri A, Nalivan OA, Pal SC, Saha A, Costache R (2022) Land-subsidence susceptibility mapping: assessment of an adaptive neuro-fuzzy inference system–genetic algorithm hybrid model. *Geocarto Int* 37(26):12194–12218. <https://doi.org/10.1080/10106049.2022.2066198>
- Wu W, Li Y, Luo X, Zhang Y, Ji X, Li X (2019) Performance evaluation of the CHIRPS precipitation dataset and its utility in drought monitoring over Yunnan Province, China. *Geomat Nat Hazards Risk* 10(1):2145–2162. <https://doi.org/10.1080/19475705.2019.1683082>
- Wulder MA, White JC, Loveland TR, Woodcock CE, Belward AS, Cohen WB et al (2016) The global Landsat archive: status, consolidation, and direction. *Remote Sens Environ* 185:271–283. <https://doi.org/10.1016/j.rse.2015.11.032>
- Xie, S, Liu, L, Zhang, X, Yang, J, Chen, X, Gao, Y (2019) Automatic land-cover mapping using Landsat time-series data based on Google earth engine. *Remote Sens*, 11(24). <https://doi.org/10.3390/rs11243023>
- Yang P, Ren G, Yan P (2017) Evidence for a strong association of short-duration intense rainfall with urbanization in the Beijing urban area. *J Clim* 30(15):5851–5870. <https://doi.org/10.1175/JCLI-D-16-0671.1>
- Yao Y, Yan X, Luo P, Liang Y, Ren S, Hu Y et al (2022) Classifying land-use patterns by integrating time-series electricity data and high-spatial resolution remote sensing imagery. *Int J Appl Earth Obs Geoinf* 106:102664. <https://doi.org/10.1016/j.jag.2021.102664>
- Yuan F, Sawaya KE, Loeffelholz BC, Bauer ME (2005) Land cover classification and change analysis of the twin cities (Minnesota) metropolitan area by multitemporal Landsat remote sensing. *Remote Sens Environ* 98(2):317–328. <https://doi.org/10.1016/j.rse.2005.08.006>
- Zhang HK, Roy DP (2017) Using the 500m MODIS land cover product to derive a consistent continental scale 30m Landsat land cover classification. *Remote Sens Environ* 197:15–34. <https://doi.org/10.1016/j.rse.2017.05.024>
- Zhang, R, Tang, X, You, S, Duan, K, Xiang, H, Luo, H (2020) A novel feature-level fusion framework using optical and SAR remote sensing images for land use/land cover (LULC) classification in cloudy mountainous area. *Appl Sci*, 10(8). <https://doi.org/10.3390/app10082928>
- Zhao W, Du S (2016) Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS J Photogramm Remote Sens* 113:155–165. <https://doi.org/10.1016/j.isprsjprs.2016.01.004>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.