

Article

Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India

Kotapati Narayana Loukika ¹, Venkata Reddy Keesara ¹ and Venkataramana Sridhar ^{2,*}

¹ Department of Civil Engineering, National Institute of Technology Warangal, Warangal 506004, India; loukikak@student.nitw.ac.in (K.N.L.); kvreddy@nitw.ac.in (V.R.K.)

² Department of Biological Systems Engineering, Virginia Polytechnic Institute, State University, Blacksburg, VA 24061, USA

* Correspondence: vsri@vt.edu

Abstract: The growing human population accelerates alterations in land use and land cover (LULC) over time, putting tremendous strain on natural resources. Monitoring and assessing LULC change over large areas is critical in a variety of fields, including natural resource management and climate change research. LULC change has emerged as a critical concern for policymakers and environmentalists. As the need for the reliable estimation of LULC maps from remote sensing data grows, it is critical to comprehend how different machine learning classifiers perform. The primary goal of the present study was to classify LULC on the Google Earth Engine platform using three different machine learning algorithms—namely, support vector machine (SVM), random forest (RF), and classification and regression trees (CART)—and to compare their performance using accuracy assessments. The LULC of the study area was classified via supervised classification. For improved classification accuracy, NDVI (normalized difference vegetation index) and NDWI (normalized difference water index) indices were also derived and included. For the years 2016, 2018, and 2020, multitemporal Sentinel-2 and Landsat-8 data with spatial resolutions of 10 m and 30 m were used for the LULC classification. ‘Water bodies’, ‘forest’, ‘barren land’, ‘vegetation’, and ‘built-up’ were the major land use classes. The average overall accuracy of SVM, RF, and CART classifiers for Landsat-8 images was 90.88%, 94.85%, and 82.88%, respectively, and 93.8%, 95.8%, and 86.4% for Sentinel-2 images. These results indicate that RF classifiers outperform both SVM and CART classifiers in terms of accuracy.

Keywords: classification and regression trees; Google Earth Engine; land use land cover; normalized difference vegetation index; random forest; support vector machine



Citation: Loukika, K.N.; Keesara, V.R.; Sridhar, V. Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India. *Sustainability* **2021**, *13*, 13758. <https://doi.org/10.3390/su132413758>

Academic Editor: Hone-Jay Chu

Received: 20 November 2021

Accepted: 9 December 2021

Published: 13 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding land use and land cover at various scales will aid future studies into a variety of global phenomena, such as droughts, floods, erosion, migration, and climate change. The continuous and accurate analysis of LULC is an integral part of the sustainable development activities undertaken in any given area. Detailed LC maps are an important input for a variety of scientific studies involving climate change effects on streamflow and water budgets [1,2], geomorphology [3], groundwater management [4–7], social knowledge management of natural resources [8], and agricultural land monitoring [9–11]. LULC maps can help determine which types of lands are suitable for agriculture and which can be useful in watershed management in general [12,13]. Remotely sensed imagery is the most commonly used method for mapping land cover and tracking changes over time [14–17]. Due to population increases and the need to develop new regions to meet the demand for food production, energy generation, and water security, the hydrologic and water resources modeling community is keen to integrate and evaluate changing land use and its impact on the water budget [18–21]. Generating low-resolution land cover maps across large regions

involves massive amounts of data. As such, huge storage capacities, high processing power, and the flexibility to apply diverse approaches are all required [22]. These requirements were addressed and such technology was made freely available to anyone with the launch of the Google Earth Engine (GEE). GEE is a cloud-based platform which combines vast amounts of remote sensing data from multiple sources with a high-performance computer service, allowing for quick and easy satellite imagery computing [23–26].

GEE contains freely available satellite imagery from Landsat, Sentinel, and MODIS, among others. Client libraries are created in JavaScript and Python handles the code editing [27–30] GEE employs the MapReduce architecture for parallel processing, which is a technique for breaking down large amounts of data into smaller pieces and processing them across multiple devices. As a result, the data were recompiled for the final result after being processed as many separate parts. LULC classification results utilizing remote sensing-based imagery and using non-parametric machine learning methods—such as classification and regression trees (CART), support vector machine (SVM), and random forest (RF)—were found to be exceptionally accurate [31–34]. GEE is applied in a variety of LULC-based research fields due to its extensive capabilities [23]. Gong et al. (2013) [35] created a land cover global map of 30 m using cloud computing techniques. Midekisa et al. (2017) [36] produced maps of locations all over the African continent for 15 years using the GEE platform. Kolli et al. (2020) [26] mapped land use changes around Kolleru Lake, India, using an RF classifier and obtained an overall accuracy of 95.9% and a kappa coefficient of 0.94. Rahman et al. (2020) [10] analyzed the performance, via accuracy levels, of RF and SVM on the classification of urban and rural areas in Bangladesh. They achieved a maximum SVM accuracy of 96.9% for Bhola and 98.3% for Dhaka. Large-scale urban land GEE has also been used in agricultural areas for mapping of crops [27], as well as for comparative analysis with several machine learning methods and multitemporal datasets over larger regions [37,38]. Most studies using GEE have focused on the role of temperature in climate change, the analysis of LULC changes, and the monitoring of water resources using time series analysis [39–41]. GEE has some computational constraints in terms of time, storage, and memory. Tamiminia et al. (2020) [25] discussed some of these limitations, such as large computations; given the time constraints involved, it is better to use batch process. Furthermore, GEE encounters memory issues in some cases when processing large numbers of datasets.

With the growing demand for reliable LULC data from satellite images over large areas, it is more important than ever to understand machine learning methods and their performance in widely used cloud-based platforms, such as GEE. The vast majority of the available studies focus on comparisons between LULC classifiers. There are few studies on large-scale LULC mapping using machine learning methods which compare LULC maps created from different multispectral satellite images. The main aim of the present study was to use multispectral satellite images from Landsat-8 and Sentinel-2 for LULC classification, compare existing machine learning methods on the GEE platform, and thereby determine the satellite image source and machine learning algorithm which result in classification with the highest accuracy.

2. Study Area

The Munneru sub-basin is one of the most important agriculturally dominated sub-basins in Lower Krishna basin, India (Figure 1). This sub-basin drains areas in the districts of Khammam, Warangal (Telangana State), and Krishna (Andhra Pradesh State). The Munneru basin lies between latitudes of 16.6° N and 18.1° N and longitudes of 79.2° E and 80.8° E. The Munneru basin encompasses a total area of 9854 km². Paddy, cotton, and maize are the dominant crops in the Munneru basin, which also has deciduous and degraded/scrub forests and a large spread of plantations in the lower part. This sub-basin contains the tributaries Munneru, Akeru, Wyra, and Kattaleru. This basin's major water bodies include Pakhal Lake, Wyra Reservoir, Bhayyaram Cheruvu, and Lanka Sagar Reservoir. The dominant soils in this basin are red soils followed by black soils. The river

plays a major role in providing water for irrigation and for domestic purposes. Around 77% of the total area of the basin is cultivable with the main crops being rice, corn, cotton, sorghum, millet, sugar cane, and a variety of horticulture crops. Population increase has resulted in heightened demand for and consumption of water for both domestic and industrial purposes, putting a strain on water resources. Land use of the watershed is mainly dominated by cropland and irrigated land. The major changes are the conversion of barren land to built-up land, cropland to dryland, and urbanisation in key areas such as Khammam and Nadigama. As part of the ongoing 2017 BRICS-DST project, this basin continues to be studied for the development of an integrated water resources management model under climate change scenarios.

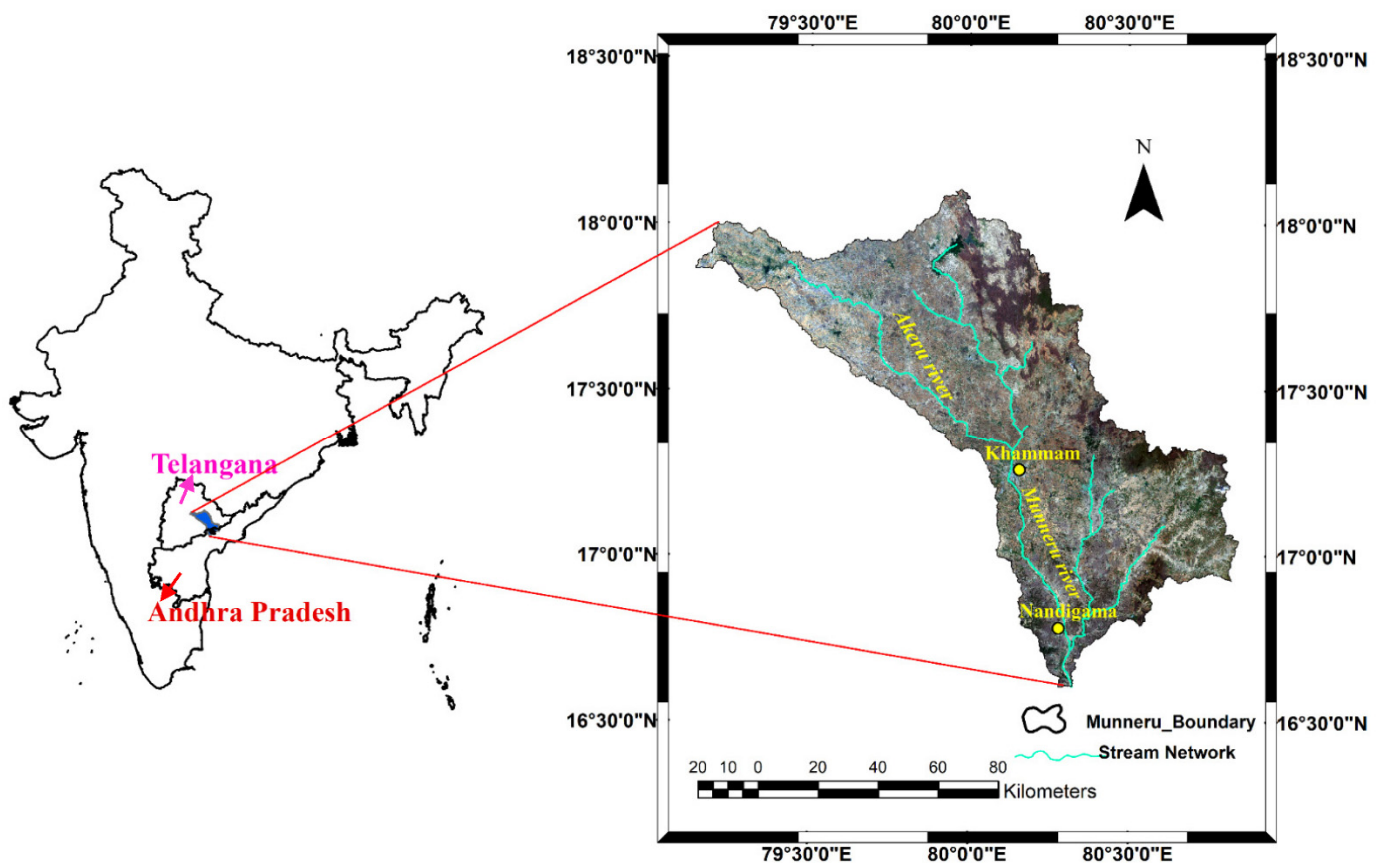


Figure 1. Location map of the Munneru sub-basin, India.

3. Data and Methods

3.1. Data

A massive amount of EOD (Earth observation data) from the previous four decades—encompassing satellite images from popular platforms such as Sentinel, Landsat, and MODIS, as well as other geographic data including climate and demographic data—are stored in the cloud-based GEE platform. Landsat and Sentinel data can be accessed via USGS (the United States Geological Survey) in GEE. In the current study, Landsat-8 surface reflectance Tier 1 data—atmospherically corrected using the LASRC (Landsat-8 Surface Reflectance Code) and Sentinel-2 Level-1C data—were used. Due to cloud cover, less than 10 percent of the datasets were selected for each year, and those images were combined into a single image. For classification of the images, six bands from Landsat-8 and nine bands from Sentinel-2 were used. For Landsat-8, the total number of images used was ten in 2016, fifteen in 2018, and fourteen in 2020. For Sentinel-2, 23 images were used in 2016, 44 images in 2018, and 36 images in 2020. The unit of analysis was the pixel, with each pixel in Landsat representing $30\text{ m} \times 30\text{ m}$, and $10\text{ m} \times 10\text{ m}$ in Sentinel. LULC was divided

into five major classes: water bodies, forest, barren land, vegetation, and built-up areas. Agriculture area and plantations were considered vegetation, while rivers and ponds were considered water bodies. The study made use of spectral bands from 1–7 of Landsat-8 images, as well as 2–8 and 11–12 of Sentinel-2 images (Table 1).

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

Data Layer	Source	Bands Used	Central Wavelength (µm)	Band Width (µm)	Spatial Resolution (m)
Landsat-8 Operational Land Imager surface reflectance Tier 1	Google Earth Engine (GEE), data accessed via the U.S. Geological Survey (USGS)	Blue (Band 2)	0.482	0.060	30
		Green (Band 3)	0.561	0.057	30
		Red (Band 4)	0.655	0.038	30
		Near-Infra-Red (Band 5)	0.865	0.028	30
		Short-Wave Infra-Red 1 (Band 6)	1.609	0.085	30
		Short-Wave Infra-Red 2 (Band 7)	2.200	0.186	30
		Sentinel-2 MSI: MultiSpectral Instrument, Level-1C	Google Earth Engine (GEE), data accessed via the U.S. Geological Survey (USGS)	Blue (Band 2)	0.496
Green (Band 3)	0.560			0.036	10
Red (Band 4)	0.664			0.031	10
Red-Edge 1(Band 5)	0.704			0.015	20
Red-Edge 2 (Band 6)	0.740			0.015	20
Red-Edge 3 (Band 7)	0.782			0.020	20
Near-Infra-Red (Band 8)	0.835			0.106	10
Short-Wave Infra-Red 1 (Band 11)	1.610			0.091	20
Short-Wave Infra-Red 2 (Band 12)	2.202			0.175	20

3.2. Methods

Figure 2 presents the methodology flowchart used in this study. Orthorectified images with the least amount of cloud cover served as the primary input for classification. The first step after importing the satellite data into GEE was to remove the cloud shadow and cloud cover. Contaminated pixels were removed from all available images due to cloudy or no-data conditions using cloud mask [29], a technique suggested by Simonetti et al. (2015) [42] and Zurqani et al. (2018) [43] and achievable on GEE. The yearly means of normalized difference vegetation (NDVI) and normalized difference water (NDWI) indices were calculated in the second phase.

To create a composite image, Landsat and Sentinel data from each year were combined into a single image using the median filter. A median value is assigned to each pixel for the entire stack of images, resulting in a single image for the entire image collection. To perform LULC classification, high-resolution Google Earth images were used to generate 575 training polygons for five land use classes. The generated polygons were evenly distributed throughout the study area. Next, the training data were loaded into the GEE as a feature collection table. For maximum classification accuracy, indices such as NDVI and NDWI were used.

The NDVI [44] is the normalized difference between the NIR and red bands and the NDWI [45] is the normalized difference between the NIR and SWIR bands, as shown in Equations (1) and (2):

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

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (2)$$

Machine learning algorithms available in GEE, such as RF, CART, and SVM, were used to train the classifiers for both Landsat-8 and Sentinel-2 images.

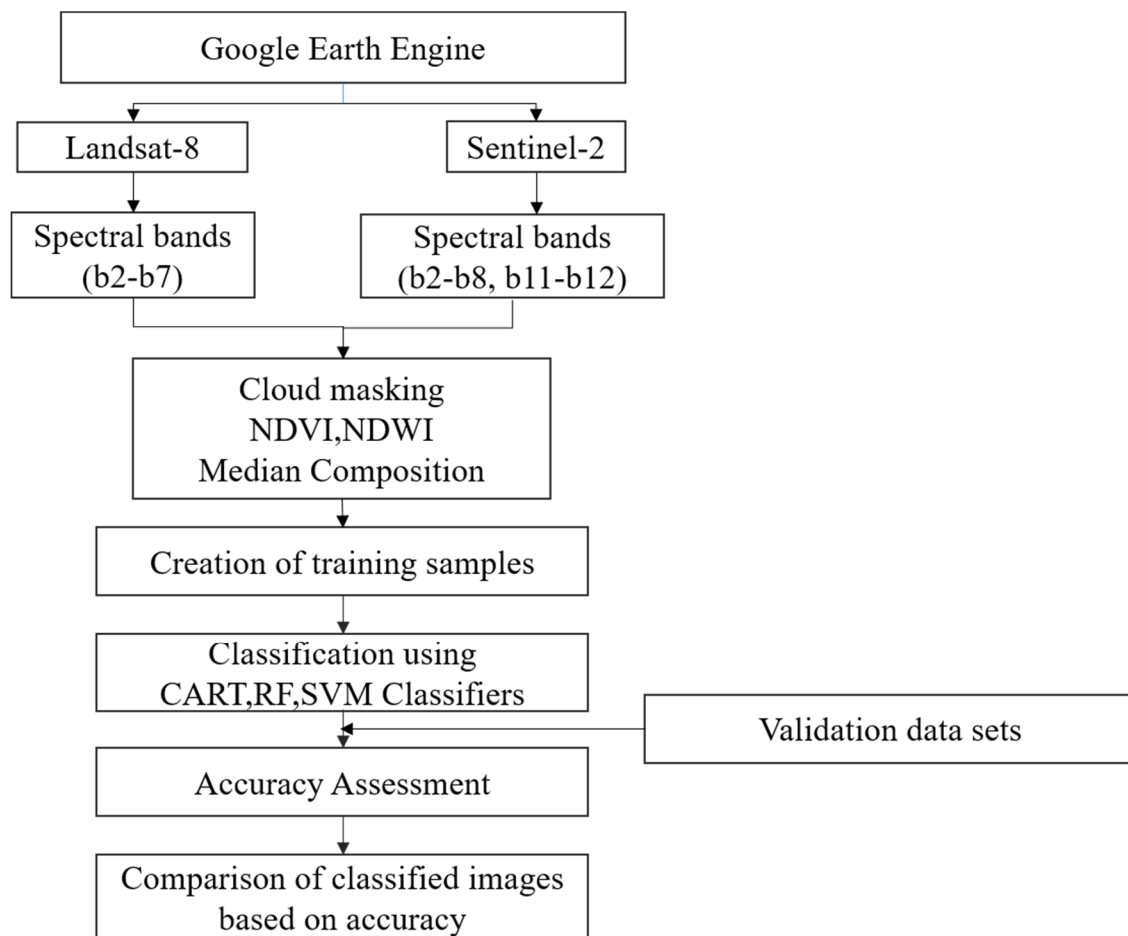


Figure 2. Methodology for LULC classification on the GEE platform.

Classification and Regression Tree (CART)

CART is a binary decision classification tree developed by Breiman et al. (1984) [46] that allows for simple decision making in logical if-then scenarios. CART operates recursively by splitting nodes until it reaches the terminal nodes, based on a predefined threshold. In this approach, input data are split into group sets and the trees are constructed utilizing all except one of those. The tree is validated using the left-out group, and the reduced tree with the lowest deviation is selected. CART is highly dependent on the sample size used in each class. The effectiveness of CART is hampered in particular by high dimensionality data, which result in complex tree architectures. The “classifier.smileCart” technique, which is included in the GEE library, was used in the current study to perform CART classification.

Random Forest Classifier (RF)

RF is the most commonly used classifier that builds an ensemble classifier [47] by combining many CART trees. Multiple decision trees are generated by RF utilizing a random selection of training datasets and variables. Internally, the non-training samples are used to evaluate the classifier’s performance and provide an unbiased assessment of the generalization error. To establish the appropriate split for building of a tree, RF selects variables at random from training samples at each node. The two most important input factors for RF are the number of parameters and trees, which are both user-defined parameters. According to the literature, the optimal number of trees to be counted ranges from 100 to 500, and the optimal number of variables counted is the square of the set of variables [48].

Support Vector Machine (SVM)

The support vector machine (SVM) is a type of supervised learning algorithm that is used to solve regression and classification issues. SVM classifiers create an ideal hyperplane in the training stage that separates multiple classes with the fewest misclassified pixels. SVM is used to select the extreme points/vectors that will help create the hyperplane. These extreme points are referred to as support vectors. The main parameters for selecting support vectors are the cost parameter C , Gamma, and kernel functions [49]. The grid search technique is used to define C and Gamma parameters, resulting in reliable prediction results. C , the cost parameter, has a significant impact on support vector selection and SVM performance. The linear kernel is preferred for training on large datasets.

Accuracy Assessment

Once the classification was completed using machine learning algorithms, an accuracy assessment was performed to determine the accuracy of the classified images. Training datasets were divided into training and validation sets. Of the total training datasets, 70 percent, or 402 polygons, were used for training and 30 percent, or 173 polygons, were used as testing sets. A confusion matrix is a built-in algorithm in GEE that validates and then evaluates the classification accuracy of the images. The kappa coefficient (k) and overall accuracy (OA) are calculated from the following equations:

$$OA = \left(\frac{P_c}{P_n} \right) * 100 \quad (3)$$

where P_c is the number of pixels classified correctly and P_n is the total number of pixels.

$$k = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad (4)$$

where r = the number of rows and columns in the error matrix, x_{ii} = the number of observations in row i and column i , x_{i+} = the marginal total of row i , x_{+i} = the marginal total of column i , and N = the total number of observations. The consumer accuracy for each class is determined by the ratio of properly categorized pixels in the class to the total number of classified pixels. Similarly, the producer accuracy is determined by the ratio of properly categorized pixels to the total number of pixels in the reference data in each class. The proportionate reduction in errors is determined by comparing the errors in a classification class with the errors in a totally random class. Typically, the magnitude ranges from -1 to $+1$. It agrees well with the categorization if the value is larger than $+0.5$ [50]. The best performing classifier will be selected for further classification of images and will be used to examine spatiotemporal change in the future.

4. Results and Discussion

4.1. LULC Classification Using GEE

This study examines the performance of various machine learning techniques on LULC classification using Landsat-8 surface reflectance Tier 1 and Sentinel-2 Level-1C data with 30 m and 10 m resolutions, respectively. Figures 3 and 4 demonstrate how machine learning algorithms such as RF, CART, and SVM were used for the classification of LULC maps from 2016, 2018, and 2020 using Landsat-8 and Sentinel-2 images on the GEE platform. As the primary input, orthorectified images with minimal cloud cover were used and contaminated pixels, due to cloudy conditions, were removed from all available images using the cloud mask algorithm available on the GEE platform. To fill the gaps in cloudy images, temporal aggregation methods such as median, mean, and minimum/maximum were used. In this study, the median was used to compose Landsat-8 and Sentinel-2 images for the entire year. Two widely used indices that were developed and used as additional inputs for the classification of LULC are normalized difference water index (NDWI) and normalized difference vegetation index (NDVI), which are representative of water bodies and vegetation characteristics, respectively. Training

and validation datasets were generated via image observations. A total of 575 training sites were used for classification. As a rule of thumb [50], each class should have at least 50 training samples for classification. Each class received 80–95 samples for training and 65–80 samples for validation. SVM, RF, and CART algorithms were used to classify the same training and validation data. LULC was divided into five major classes: vegetation, forest, water bodies, built-up, and barren land. From the studies of Kohavi (1995) [51], the best cross validation factor was determined to be 5 or 10 and was used as an input value for CART classifier. A number of trees in the 50–100 range exhibited higher accuracy and performed better for RF classification [48]. In the present study, a total of 100 trees yielded good results. Kernel type, gamma value, and cost are all important parameters in SVM. For large datasets, the linear kernel type is preferable [49]. For linear kernels, the gamma parameter is not required. The cost parameter determines the severity of the penalty for incorrectly classified data. A higher C value indicates less misclassified data. The C-SVC method is used for SVM classification, with a cost parameter of 10 and a linear kernel type.

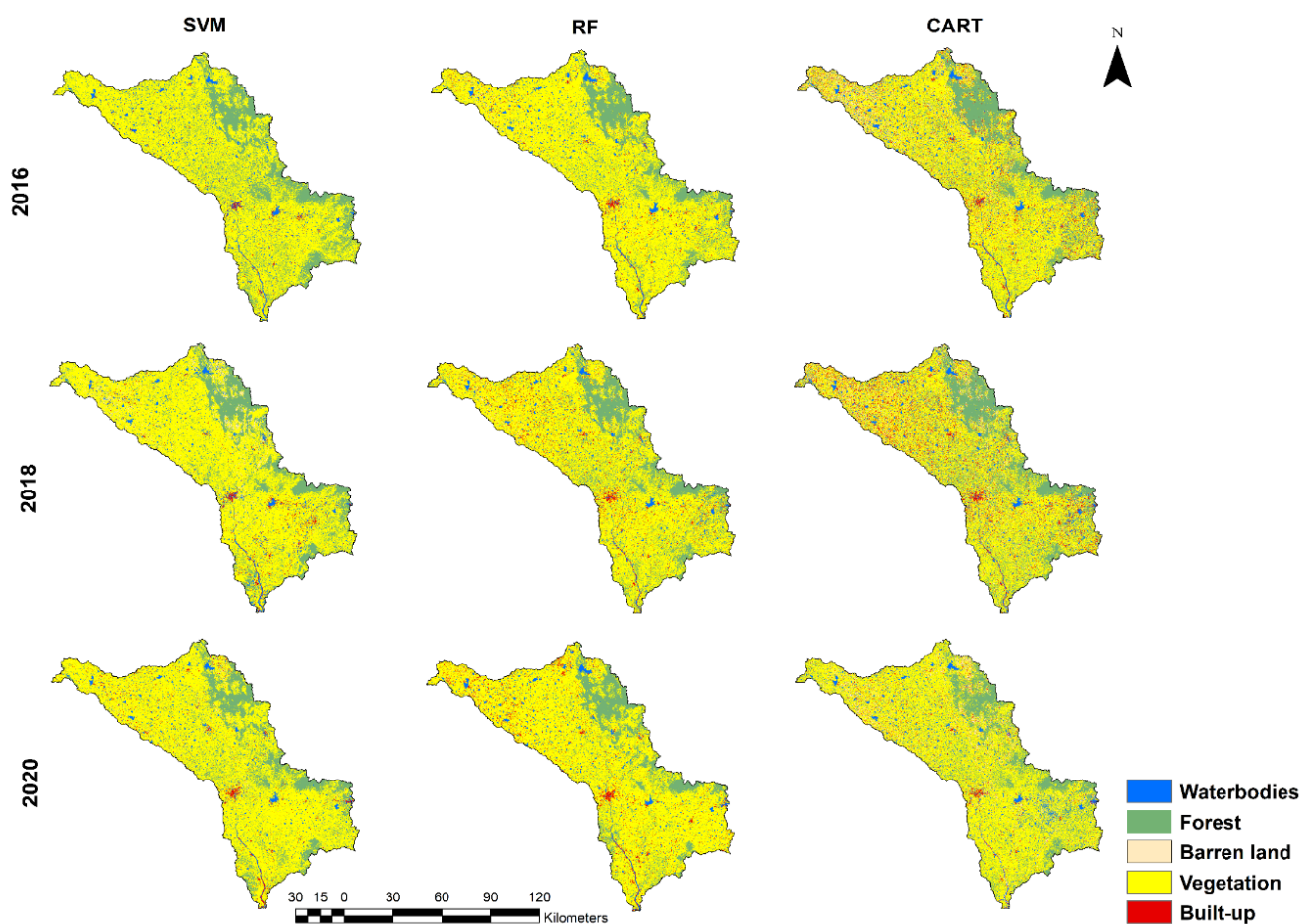


Figure 3. LULC maps of Landsat-8 images using SVM, RF, and CART classifiers for the years 2016, 2018, and 2020.

From Figure 3, we can see that classification by CART resulted in most of the vegetation being misclassified as built-up, water bodies, or forest in the years 2016 and 2018. Vegetation was also misclassified as barren land and, to a lesser extent, as water bodies in 2020. For SVM, vegetation was slightly misclassified as forest, built-up, or water bodies in 2016. Vegetation was slightly misclassified as built-up or forest in 2018. For 2020, SVM classified the image well except for forest and some built-up area. For all Landsat-8 images from the three years, the RF classifier performed well in comparison to the other two classifiers.

For the years 2016 and 2018, SVM misclassified vegetation as forest because dense plantations have identical reflectance to forests. There was also slight misclassification

around water bodies, as observed in Figure 4. Water bodies, forest, barren, and built-up areas are correctly classified using the RF algorithm. Vegetation was slightly misclassified as forest using the RF algorithm in 2016 and 2020. Using CART, the classification of water bodies, built-up areas, and forest was superior to that of vegetation and barren classes. Figure 5 depicts the changes in the LULC classification as determined from Sentinel-2 and Landsat-8 images using RF for the years 2016, 2018, and 2020. Figure 5 indicates that, for the period of 2016–2020, built-up, barren land, and vegetation increased by 0.6%, 0.78%, and 0.015%, respectively. Forest cover and water bodies decreased by 0.088% and 0.22%, respectively, for the period of 2016–2020. In 2018, vegetation cover decreased before increasing in 2020. Similarly, water bodies were reduced in 2018 and increased in 2020 as a result of heavy rains in some parts of the basin. Barren lands were more prevalent in 2018 than in 2016.

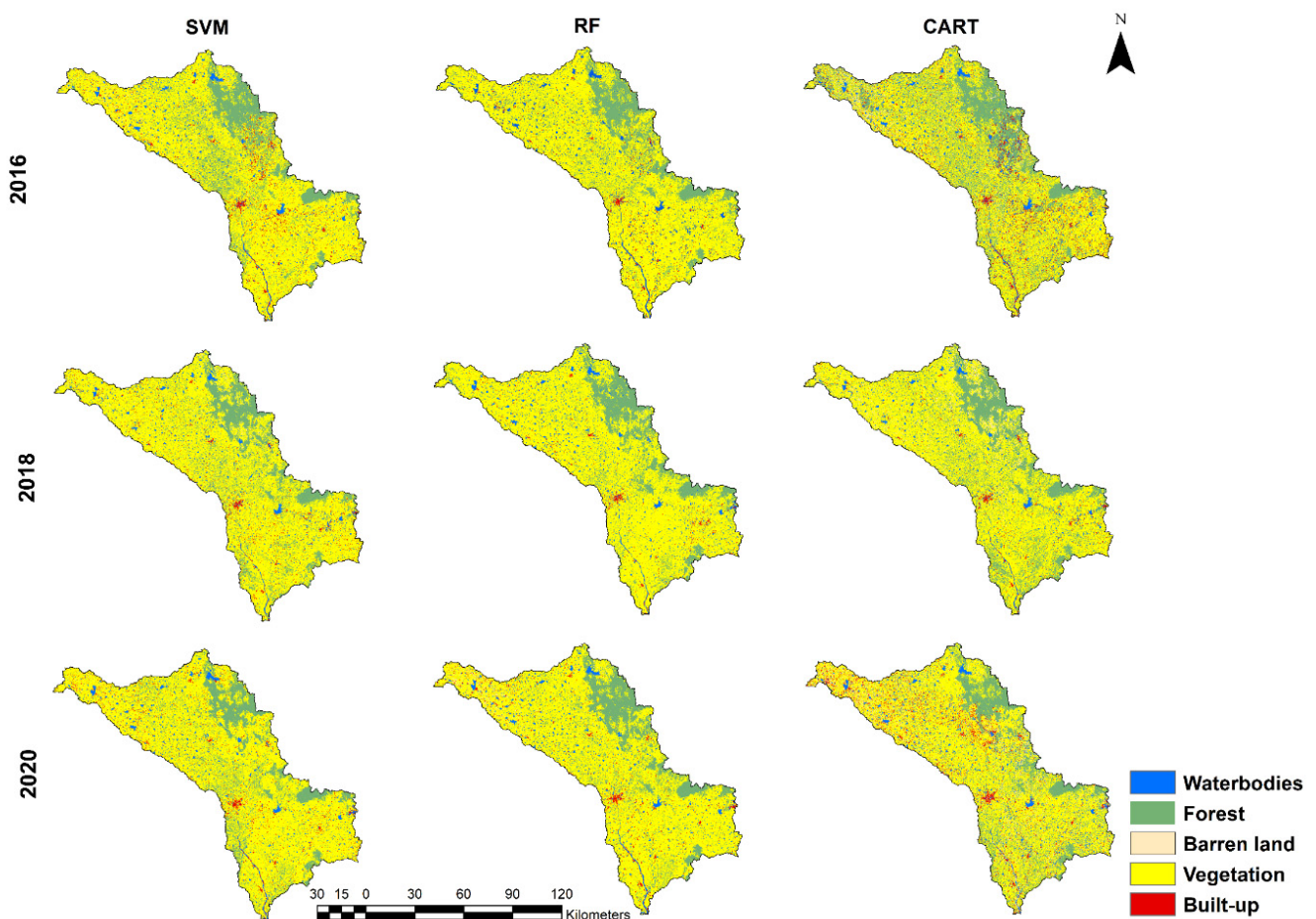


Figure 4. LULC maps of Sentinel-2 images using SVM, RF, and CART classifiers for the years 2016, 2018, and 2020.

4.2. Comparison of Classification Performances

Accuracy assessment was used to assess the efficacy of various classifiers. The most commonly used metric for evaluating the accuracy and effectiveness of all classifiers is overall accuracy (OA), which represents the amount of test data correctly classified by the classifier as a percentage. Furthermore, the confusion matrix and user and producer accuracy were used to assess the class-wise performance of each classifier. GEE has methods for determining the correctness of several classifiers, and accuracy assessments were used to determine the accuracy of each classified image for the years 2014, 2016, 2018, and 2020. To obtain the ground truth data, user interpretation was employed to select testing sites. The performance of RF, SVM, and CART classifiers are compared in Table 2 in terms of the overall accuracy and kappa coefficient.

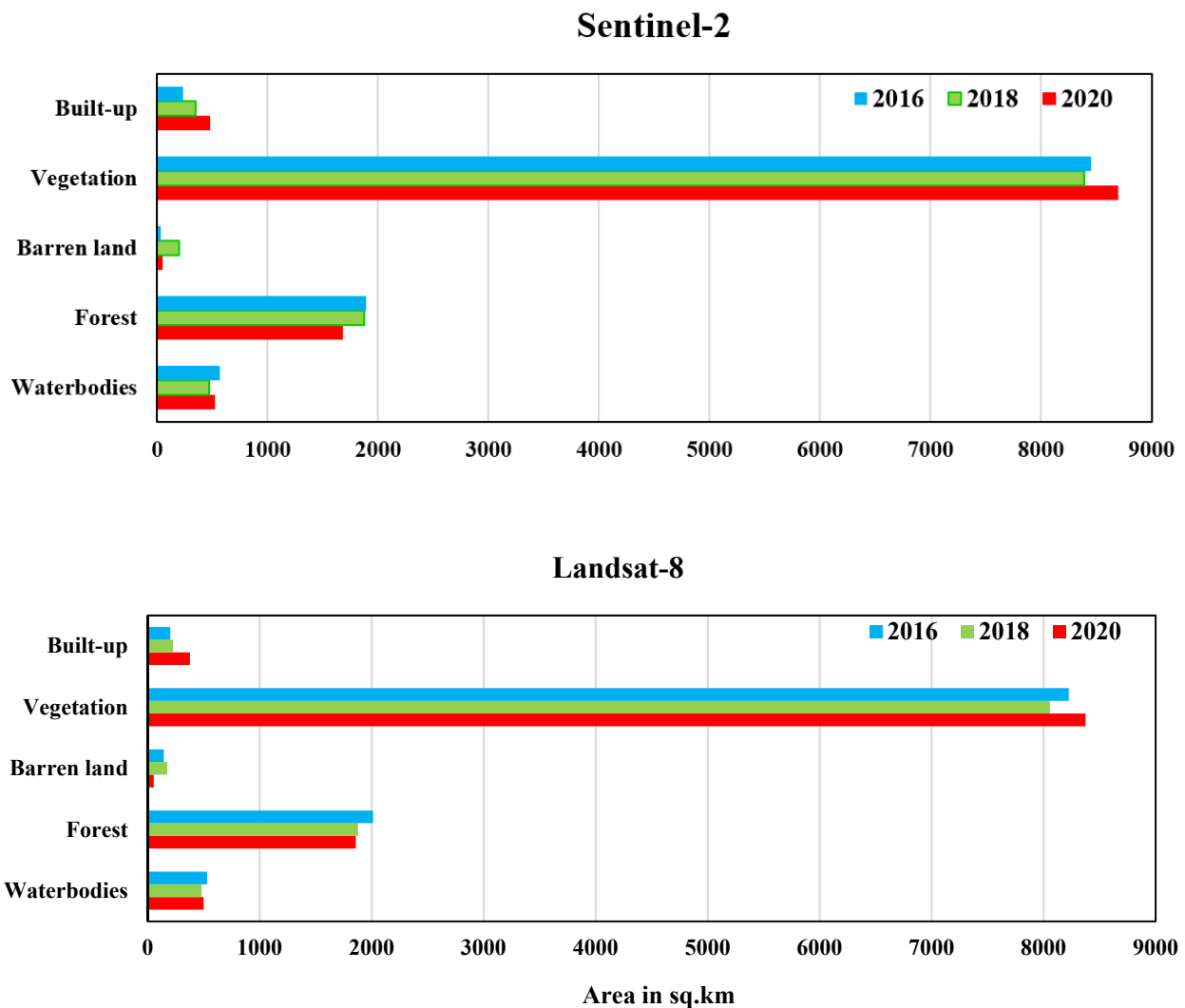


Figure 5. LULC changes using RF classifier of Sentinel-2 and Landsat-8 for the years 2016, 2018, and 2020.

Table 2. Kappa coefficient and overall accuracy of Landsat-8 and Sentinel-2 for various machine learning classifiers.

Year	Classifier	Landsat-8		Sentinel-2	
		Overall Accuracy (%)	Kappa Coefficient	Overall Accuracy (%)	Kappa Coefficient
2016	SVM	88.99	0.81	92.37	0.868
	RF	93.93	0.89	94.65	0.904
	CART	81.61	0.72	84.75	0.747
2018	SVM	91.62	0.85	93.05	0.878
	RF	94.86	0.91	95.85	0.928
	CART	82.59	0.73	85.88	0.772
2020	SVM	92.95	0.87	95.54	0.918
	RF	95.84	0.92	97.04	0.947
	CART	86.58	0.79	88.81	0.814

From Table 2, the RF classifier outperformed the SVM and CART classifiers. It can also be observed that Sentinel-2 images were more accurate than Landsat images. The overall accuracy of RF, SVM, and CART classifiers for Landsat-8 was 94.85%, 90.88%, and 82.88%, respectively. The average overall accuracy of RF, SVM, and CART classifiers for Sentinel-2 was 95.84%, 93.65%, and 86.48%, respectively. For Landsat-8 data with RF, SVM,

and CART classifiers, the average kappa coefficients were 0.90, 0.84, and 0.74, respectively. The average kappa coefficients for RF, SVM, and CART classifiers on Sentinel-2 data were 0.92, 0.88, and 0.77, respectively. When compared with SVM and CART, the RF classifier achieved the highest producer and user accuracy for both Landsat-8 and Sentinel-2.

The producer and user accuracy of Landsat-8 and Sentinel-2 for each land class are presented in Figures 6 and 7. When compared with other classes, forest and water bodies performed well, with more than 90% user and producer accuracy for both Sentinel and Landsat data. For both satellites, RF outperformed the other classifiers in terms of producer and user accuracy.

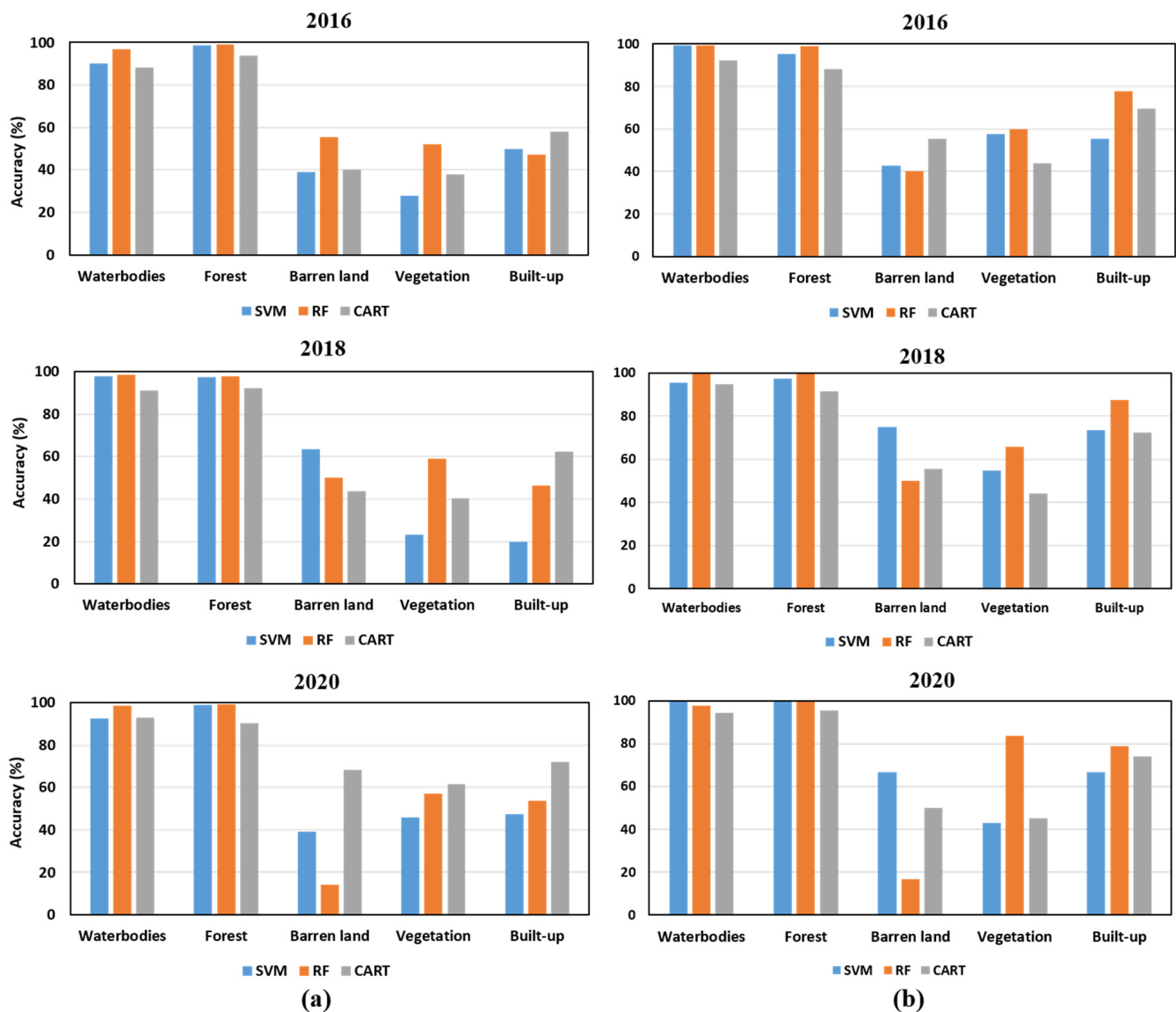


Figure 6. Accuracy of the user for each land class using SVM, RF, and CART classifiers: (a) Landsat-8, (b) Sentinel-2.

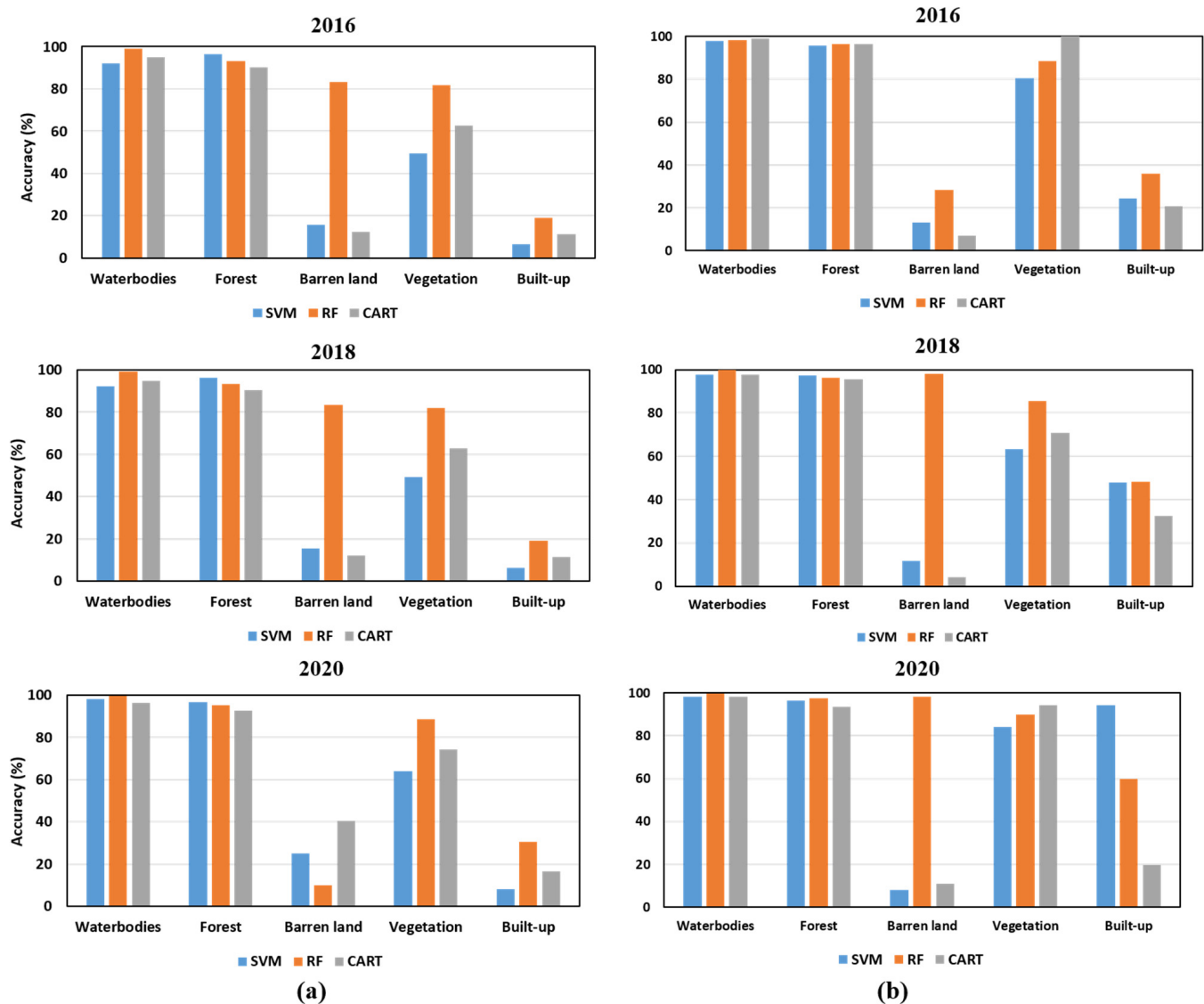


Figure 7. Accuracy of the producer for each land class using SVM, RF, and CART classifiers: (a) Landsat-8, (b) Sentinel-2.

5. Discussion

In the current study, different machine learning methods were applied to determine the accuracy of LULC classifications using multispectral Sentinel-2 and Landsat imagery. From Figures 3 and 4, as the reflectance of plantations coincides with that of forest, the majority of the vegetation was misclassified and confused with forest. As there is a river in the study area, no flow was observed during the non-monsoon season, and the area was slightly misclassified as built-up and barren in some parts of the river due to reflectance matches. Overall accuracy (OA) is the most extensively used metric for estimating accuracy. It represents the percentage of the testing set that was correctly classified by the classifier. Additionally, confusion matrix, user accuracy, and producer accuracy are utilized to further evaluate the class-level performance of a given classifier [52]. The best performing model is chosen based on accuracy and kappa coefficient. RF classifies well in all classes, as evidenced by other studies [53,54]. The accuracy of barren land was lower than that of other land use classes, as observed in Figures 6 and 7. Because some parts of plantations are classified as forest rather than vegetation, the accuracy of vegetation was reduced. Built-up areas were also mistaken for water bodies because their reflectance values match during the non-monsoon season. Barren land had very few pixels, which were insufficient to train the classifier efficiently, resulting in poor performance when compared with the other land use

classes. In terms of producer and user accuracy, RF outperformed the other classifiers for both satellites; however, SVM and CART classifiers performed better for water bodies and forest land cover. Forest, water bodies, and barren land were misclassified as vegetation and built-up areas by SVM and CART classifiers. RF outperformed the other two classifiers in classifying all five classes for both the Sentinel-2 and Landsat-8 datasets.

It is difficult to distinguish between built-up, vegetation, and barren land classes in 30 m resolution Landsat-8 images due to mixed pixels. The Sentinel-2 image, on the other hand—in which multiple land-use classifications are combined together at the same time—allows for superior classification of tiny regions and diverse land use systems. A resolution of 10 m is preferred to others in the situation of scattered classes and rapidly shattered area categorization. In this scenario, the Sentinel-2 image performed better. When the results obtained from Landsat-8 and Sentinel-2 imagery are compared, the Sentinel-2 dataset yielded the highest accuracy results due to its higher spatial resolution and greater number of band combinations used for classification. When compared with Landsat data, the Sentinel red-edge band combination is best suited for accurately classifying vegetation.

Random forests, in general, combine numerous soft linear boundaries at the surface of the decision. In SVM and CART, misclassification occurs between some classes and SVM performs well if the input training data are sparse, making it a better choice when less data are available [54]. Each algorithm has its own set of benefits and drawbacks. RF is more resilient and less impacted by parameters, whereas SVM is sensitive to hyper-parameters [55]. RF outperformed all of the other classifiers, regardless of training data size, followed by SVM and then CART. Some existing literature claims that SVM outperforms CART [56], which was observed in the present study. However, few studies claim that CART outperforms SVM [57]. It is best to use the least sensitive, most complex, and fastest method for classification [58]. Through the use of multispectral satellite images, GEE simplifies the process of classifying large study areas. With the available methods and algorithms applied here, the performance of image pre-processing tasks is made more flexible.

6. Conclusions

The performance of RF, CART, and SVM machine learning methods for the classification of LULC on the GEE platform, using Landsat-8 and Sentinel-2 datasets over three years, was analyzed. The classifier type used influenced the accuracy of LULC data classification from satellite images. Accuracy assessment of each individual class can be used to evaluate the performance of each classifier with respect to each class. The accuracy of the classifications was assessed using an error matrix. For both Sentinel-2 and Landsat-8 images, RF outperformed CART and SVM. The combination of band data was important and affected classification accuracy. Sentinel data have red-edge bands, which allow for better vegetation classification than Landsat data. Because of its high resolution, the Sentinel-2 dataset outperformed the Landsat-8 in terms of accuracy. The most suitable classifier for any given scenario may also be affected by the study region, thematic accuracy, training sample quality, and map necessity.

Author Contributions: K.N.L.: Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review and editing. V.R.K.: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review and editing. V.S.: Conceptualization, Investigation, Software, Supervision, Validation, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Department of Science and Technology (DST), Government of India, under the BRICS—DST project with grant no. DST/IMRCD/BRICS/PilotCall2/IWMM-BIS/2018 (G). The authors acknowledge the support provided by the Virginia Agricultural Experiment Station (Blacksburg) and the Hatch Program of the National Institute of Food and Agriculture, U.S. Department of Agriculture (Washington, D.C.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Sridhar, V.; Kang, H.; Ali, S.A. Human-Induced Alterations to Land Use and Climate and Their Responses for Hydrology and Water Management in the Mekong River Basin. *Water* **2019**, *11*, 1307. [[CrossRef](#)]
- Sridhar, V.; Jin, X.; Jaks, W.T.A. Explaining the Hydroclimatic Variability and Change in the Salmon River Basin. *Clim. Dyn.* **2013**, *40*, 1921–1937. [[CrossRef](#)]
- Sujatha, E.R.; Sridhar, V. Spatial Prediction of Erosion Risk of a Small Mountainous Watershed Using RUSLE: A Case-Study of the Palar Sub-Watershed in Kodaikanal, South India. *Water* **2018**, *10*, 1608. [[CrossRef](#)]
- Sridhar, V.; Billah, M.M.; Hildreth, J.W. Coupled Surface and Groundwater Hydrological Modeling in a Changing Climate. *Groundwater* **2018**, *56*, 618–635. [[CrossRef](#)] [[PubMed](#)]
- Xiao, Y.; Liu, K.; Yan, H.; Zhou, B.; Huang, X.; Hao, Q.; Zhang, Y.; Zhang, Y.; Liao, X.; Yin, S. Hydrogeochemical Constraints on Groundwater Resource Sustainable Development in the Arid Golmud Alluvial Fan Plain on Tibetan Plateau. *Environ. Earth Sci.* **2021**, *80*, 1–17. [[CrossRef](#)]
- Xiao, Y.; Xiao, D.; Hao, Q.; Liu, K.; Wang, R.; Huang, X.; Liao, X.; Zhang, Y. Accessible Phreatic Groundwater Resources in the Central Shijiazhuang of North China Plain: Perspective From the Hydrogeochemical Constraints. *Front. Environ. Sci.* **2021**, *9*, 475. [[CrossRef](#)]
- Xiao, Y.; Hao, Q.; Zhang, Y.; Zhu, Y.; Yin, S.; Qin, L.; Li, X. Investigating Sources, Driving Forces and Potential Health Risks of Nitrate and Fluoride in Groundwater of a Typical Alluvial Fan Plain. *Sci. Total Environ.* **2022**, *802*, 149909. [[CrossRef](#)]
- Sridhar, V.; Ali, S.A.; Sample, D.J. Systems Analysis of Coupled Natural and Human Processes in the Mekong River Basin. *Hydrology* **2021**, *8*, 140. [[CrossRef](#)]
- Jamali, B.; Bach, P.M.; Cunningham, L.; Deletic, A. A Cellular Automata Fast Flood Evaluation (CA-Ffé) Model. *Water Resour. Res.* **2019**, *55*, 4936–4953. [[CrossRef](#)]
- Rahman, A.; Abdullah, H.M.; Tanzir, M.T.; Hossain, M.J.; Khan, B.M.; Miah, M.G.; Islam, I. Performance of Different Machine Learning Algorithms on Satellite Image Classification in Rural and Urban Setup. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100410. [[CrossRef](#)]
- Sridhar, V.; Anderson, K.A. Human-Induced Modifications to Land Surface Fluxes and Their Implications on Water Management under Past and Future Climate Change Conditions. *Agric. For. Meteorol.* **2017**, *234–235*, 66–79. [[CrossRef](#)]
- Cihlar, J. Land Cover Mapping of Large Areas from Satellites: Status and Research Priorities. *Int. J. Remote Sens.* **2000**, *21*, 1093–1114. [[CrossRef](#)]
- Renschler, C.S.; Harbor, J. Soil Erosion Assessment Tools from Point to Regional Scales—The Role of Geomorphologists in Land Management Research and Implementation. *Geomorphology* **2002**, *47*, 189–209. [[CrossRef](#)]
- Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and Product Vision for Terrestrial Global Change Research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [[CrossRef](#)]
- Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The Global Landsat Archive: Status, Consolidation, and Direction. *Remote Sens. Environ.* **2016**, *185*, 271–283. [[CrossRef](#)]
- Gómez, C.; White, J.C.; Wulder, M.A. Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [[CrossRef](#)]
- Noi Phan, T.; Kuch, V.; Lehnert, L.W. Land Cover Classification Using Google Earth Engine and Random Forest Classifier—the Role of Image Composition. *Remote Sens.* **2020**, *12*, 2411. [[CrossRef](#)]
- Sridhar, V.; Hubbard, K.G.; Wedin, D.A. Assessment of soil moisture dynamics of the Nebraska Sandhills using Long-Term measurements and a hydrology model. *ASCE J. Irrig. Drain. Engg* **2006**, *132*, 463–473. [[CrossRef](#)]
- Sridhar, V.; Wedin, D.A. Hydrological behaviour of grasslands of the Sandhills of Nebraska: Water and energy-balance assessment from measurements, treatments, and modelling. *Ecohydrol. Ecosyst. Land Water Process Interact. Ecohydrogeomorphol.* **2009**, *2*, 195–212. [[CrossRef](#)]
- Kang, H.; Sridhar, V.; Mills, B.F.; Hession, W.C.; Ogejo, J.A. Economy-Wide Climate Change Impacts on Green Water Droughts Based on the Hydrologic Simulations. *Agric. Syst.* **2019**, *171*, 76–88. [[CrossRef](#)]
- Setti, S.; Maheswaran, R.; Radha, D.; Sridhar, V.; Barik, K.K.; Narasimham, M.L. Attribution of Hydrologic Changes in a Tropical River Basin to Rainfall Variability and Land-Use Change: Case Study from India. *J. Hydrol. Eng.* **2020**, *25*, 05020015. [[CrossRef](#)]
- Xie, S.; Liu, L.; Zhang, X.; Yang, J.; Chen, X.; Gao, Y. Automatic Land-Cover Mapping Using Landsat Time-Series Data Based on Google Earth Engine. *Remote Sens.* **2019**, *11*, 23. [[CrossRef](#)]
- Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]

24. Sidhu, N.; Pebesma, E.; Câmara, G. Using Google Earth Engine to Detect Land Cover Change: Singapore as a Use Case. *Eur. J. Remote Sens.* **2018**, *51*, 486–500. [[CrossRef](#)]
25. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Quackenbush, L.; Adeli, S.; Brisco, B. Google Earth Engine for Geo-Big Data Applications: A Meta-Analysis and Systematic Review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 152–170. [[CrossRef](#)]
26. Kolli, M.K.; Opp, C.; Karthe, D.; Groll, M. Mapping of Major Land-Use Changes in the Kolleru Lake Freshwater Ecosystem by Using Landsat Satellite Images in Google Earth Engine. *Water* **2020**, *12*, 2493. [[CrossRef](#)]
27. Shelestov, A.; Lavreniuk, M.; Kussul, N.; Novikov, A.; Skakun, S. Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping. *Front. Earth Sci.* **2017**, *5*, 17. [[CrossRef](#)]
28. Patela, N.N.; Angiuli, E.; Gamba, P.; Gaughan, A.; Lisini, G.; Stevens, F.R.; Tatem, A.J.; Trianni, G. Multitemporal Settlement and Population Mapping from Landsat using Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *35*, 199–208. [[CrossRef](#)]
29. Mateo-García, G.; Gómez-Chova, L.; Amorós-López, J.; Muñoz-Marí, J.; Camps-Valls, G. Multitemporal Cloud Masking in the Google Earth Engine. *Remote Sens.* **2018**, *10*, 1079. [[CrossRef](#)]
30. Pimple, U.; Simonetti, D.; Sitthi, A.; Pungkul, S.; Leadprathom, K.; Skupek, H.; Som-ard, J.; Gond, V.; Towprayoon, S. Google Earth Engine Based Three Decadal Landsat Imagery Analysis for Mapping of Mangrove Forests and Its Surroundings in the Trat Province of Thailand. *J. Comput. Commun.* **2018**, *6*, 247–264. [[CrossRef](#)]
31. Nery, T.; Sadler, R.; Solis-Aulestia, M.; White, B.; Polyakov, M.; Chalak, M. Comparing Supervised Algorithms in Land Use and Land Cover Classification of a Landsat Time-Series. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5165–5168. [[CrossRef](#)]
32. Bar, S.; Parida, B.R.; Pandey, A.C. Landsat-8 and Sentinel-2 Based Forest Fire Burn Area Mapping Using Machine Learning Algorithms on GEE Cloud Platform over Uttarakhand, Western Himalaya. *Remote Sens. Appl. Soc. Environ.* **2020**, *18*, 100324. [[CrossRef](#)]
33. Liu, D.; Chen, N.; Zhang, X.; Wang, C.; Du, W. Annual Large-Scale Urban Land Mapping Based on Landsat Time Series in Google Earth Engine and OpenStreetMap Data: A Case Study in the Middle Yangtze River Basin. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 337–351. [[CrossRef](#)]
34. Tassi, A.; Vizzari, M. Object-Oriented LULC Classification in Google Earth Learning Algorithms. *Remote Sens.* **2020**, *12*, 3776. [[CrossRef](#)]
35. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654. [[CrossRef](#)]
36. Midekisa, A.; Holl, F.; Savory, D.J.; Andrade-Pacheco, R.; Gething, P.W.; Bennett, A.; Sturrock, H.J.W. Mapping land cover change over continental Africa using Landsat and Google Earth Engine cloud computing. *PLoS ONE* **2017**, *12*, e0184926. [[CrossRef](#)]
37. Dong, J.; Xiao, X.; Menarguez, M.A.; Zhang, G.; Qin, Y.; Thau, D.; Biradar, C.; Moore, B. Mapping Paddy Rice Planting Area in Northeastern Asia with Landsat 8 Images, Phenology-Based Algorithm and Google Earth Engine. *Remote Sens. Environ.* **2016**, *185*, 142–154. [[CrossRef](#)]
38. Aguilar, R.; Zurita-Milla, R.; Izquierdo-Verdiguier, E.; de By, R.A. A Cloud-Based Multi-Temporal Ensemble Classifier to Map Smallholder Farming Systems. *Remote Sens.* **2018**, *10*, 729. [[CrossRef](#)]
39. Workie, T.G.; Debella, H.J. Climate Change and Its Effects on Vegetation Phenology across Ecoregions of Ethiopia. *Glob. Ecol. Conserv.* **2018**, *13*, e00366. [[CrossRef](#)]
40. Jamei, Y.; Rajagopalan, P.; Sun, Q.C. Time-Series Dataset on Land Surface Temperature, Vegetation, Built up Areas and Other Climatic Factors in Top 20 Global Cities (2000–2018). *Data Br.* **2019**, *23*, 103803. [[CrossRef](#)]
41. Wang, C.; Jia, M.; Chen, N.; Wang, W. Long-Term Surface Water Dynamics Analysis Based on Landsat Imagery and the Google Earth Engine Platform: A Case Study in the Middle Yangtze River Basin. *Remote Sens.* **2018**, *10*, 1635. [[CrossRef](#)]
42. Simonetti, D.; Simonetti, E.; Szantoi, Z.; Lupi, A.; Eva, H.D. First Results from the Phenology-Based Synthesis Classifier Using Landsat 8 Imagery. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1496–1500. [[CrossRef](#)]
43. Zurqani, H.A.; Post, C.J.; Mikhailova, E.A.; Schlautman, M.A.; Sharp, J.L. Geospatial Analysis of Land Use Change in the Savannah River Basin Using Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *69*, 175–185. [[CrossRef](#)]
44. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
45. McFeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [[CrossRef](#)]
46. Brieman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Routledge: London, UK, 1984; Volume 45, pp. 5–32. [[CrossRef](#)]
47. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
48. Belgiu, M.; Drăgu, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
49. Hsu, C.W.; Chang, C.C.; Lin, C.J. *A Practical Guide to Support Vector Classification*; Technical Report; Department of Computer Science and Information Engineering, University of National Taiwan: Taipei, Taiwan, 2003; pp. 1–12. Available online: <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> (accessed on 8 December 2021).
50. Thomas, L.; Ralph, W.; Kiefer, J.C. Remote Sensing and Image Interpretation (Fifth Edition). *Geogr. J.* **2004**, *146*, 448–449.

51. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Int. Jt. Conf. Artif. Intell.* **1995**, *14*, 1137–1145.
52. Stehman, S.V. Sampling Designs for Accuracy Assessment of Land Cover. *Int. J. Remote Sens.* **2009**, *30*, 5243–5272. [[CrossRef](#)]
53. Pal, M.; Foody, G.M. Evaluation of SVM, RVM and SMLR for Accurate Image Classification with Limited Ground Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1344–1355. [[CrossRef](#)]
54. Shetty, S. Analysis of Machine Learning Classifiers for LULC Classification on Google Earth Engine Analysis of Machine Learning Classifiers for LULC Classification on Google Earth Engine. Master's thesis, University of Twente, Enschede, The Netherlands, 2019; pp. 1–65.
55. Chang, K.T.; Merghadi, A.; Yunus, A.P.; Pham, B.T.; Dou, J. Evaluating Scale Effects of Topographic Variables in Landslide Susceptibility Models Using GIS-Based Machine Learning Techniques. *Sci. Rep.* **2019**, *9*, 12296. [[CrossRef](#)] [[PubMed](#)]
56. Shao, Y.; Lunetta, R.S. Comparison of Support Vector Machine, Neural Network, and CART Algorithms for the Land-Cover Classification Using Limited Training Data Points. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 78–87. [[CrossRef](#)]
57. Goldblatt, R.; You, W.; Hanson, G.; Khandelwal, A.K. Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine. *Remote Sens.* **2016**, *8*, 634. [[CrossRef](#)]
58. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2019. [[CrossRef](#)]