



Evaluation of classification algorithms for land use land cover mapping in the snow-fed Alaknanda River Basin of the Northwest Himalayan Region

Gagandeep Singh¹ · Ashish Pandey¹

Received: 29 May 2021 / Accepted: 27 August 2021 / Published online: 5 September 2021
© Società Italiana di Fotogrammetria e Topografia (SIFET) 2021

Abstract

Land Use/Land Cover (LULC) maps are crucial for assessing the status of environmental and natural resources management in any river basin or watershed. LULC is a cross-cutting environmental variable that also finds significant applications in hydrological modeling, watershed management, natural disaster management, climate change studies, and land management. This research study uses three different classification algorithms to investigate the LULC status of the Alaknanda river basin of the northwest Himalayan region in India. The entire area was classified into nine LULC classes using Landsat 8 satellite imagery, initially employing the Maximum Likelihood algorithm. This generated a reasonably good overall accuracy with a high Kappa coefficient of 0.79. However, the producer's accuracies for a few significant classes were not satisfactory. This research attempts to explain the anomaly in the producer's accuracy and improve them using machine learning-based classification algorithms. Furthermore, machine learning-based classification algorithms, namely Random Trees (RT) and Support Vector Machine (SVM) were employed. Both the algorithms generated good overall accuracy with high Kappa values of 0.83 and 0.82, respectively. Interestingly, the qualitative and quantitative comparisons for the classification results revealed that both RT and SVM algorithms resulted in improved and high producer's accuracies. Therefore, this study infers that for mountainous watersheds with high variations in elevation and steep topography, machine learning-based classification algorithms perform better than the conventional statistical classification algorithm.

Keywords LULC classification · Watershed management · Maximum Likelihood · Machine learning · Random Trees · Support Vector Machine

Introduction

Land Use/Land Cover (LULC) maps have emerged as one of the most critical sources of environmental information, employed in assessing the natural resources' status in any river basin or watershed (Nie et al. 2011; Atzberger 2013; Roy et al. 2015). Land use and land cover information also find significant applications in hydrological modeling, watershed management, natural disaster management, and land management (Lillesand et al. 2008; Nie et al. 2011; Sierra-Soler et al. 2016). Satellite remote sensing offers a

unique set of advantages like global coverage, high temporal frequency, synoptic view, and the ability to observe inaccessible areas (Campbell and Wynne 2011). The most important benefit is the availability of optical remote sensing datasets in the open-source domain. Renowned agencies like the United States Geological Survey (USGS) and the European Space Agency (ESA) have been providing optical satellite datasets, viz. Landsat and Sentinel series for free, which has popularized remote sensing for various applications (Saadat et al. 2011; Wondrade et al. 2014; Mishra et al. 2020; Thanh et al. 2020). Numerous studies by different researchers have demonstrated the efficacy of optical remote sensing data products for land cover classification (Gong et al. 2013; Gómez et al. 2016).

Generally, the land cover describes the physical land types viz. area covered by forests, impervious surfaces, agricultural lands, barren lands, and water bodies. On the other hand, as the term suggests, land use describes the utility of

✉ Gagandeep Singh
gsingh@wr.iitr.ac.in

¹ Department of Water Resources Development and Management, Indian Institute of Technology Roorkee, Roorkee 247 667, Uttarakhand, India

the land for various purposes by humans, viz. for development, conservation, or mixed uses. The entire ecosystem is strongly affected by variables responsible for land use and climatic changes (Shrestha et al. 2012). LULC is attributed as one of the most relevant flow contributors in a watershed (Kim et al. 2013; Himanshu et al. 2017). Therefore, for assessing water resources availability and its management in a watershed, LULC change analysis is of utmost importance (Malik and Bhat 2014).

Moreover, LULC changes cause a substantial effect in modifying the rainfall breakup into various hydrological components like surface runoff, infiltration, interception, and evaporation (Costa et al. 2003; Mao and Cherkauer 2009; Sajikumar and Remya 2015). There are two reasons for the variation in the LULC, one being the natural dynamics, and the other is attributed to human activities (Thenkabail et al. 2005; Bontemps et al. 2008; Singh et al. 2014; Zhang et al. 2019). These factors are responsible for deforestation, global warming, loss of biodiversity, increasing natural disasters, and global environmental change (DeFries et al. 2010; Owrangi et al. 2014; Mahmood et al. 2014; Barros et al. 2021).

Remote sensing data and techniques, along with GIS, provide an apt platform to develop and prepare LULC maps. Multi-spectral/temporal satellite data with medium/high spatial resolution have materialized as the most preferred data sources for deriving LULC maps (Güler et al. 2007). Conventionally, maps were prepared using available records and extensive field surveys, rendering them tedious, laborious, time-consuming, and expensive. Moreover, due to the dynamic nature of the environment, the output maps used to become outdated (Dash et al. 2015). In contrast, remote sensing data provides highly vital information in a very cost-effective and less time-consuming manner. High-resolution satellite data products are employed in large cities to estimate LULC changes. However, the too-high cost of such data sets limits their availability (Dwivedi et al. 2005). On the contrary, satellite data products with the medium resolution, specifically from the Landsat series, are among the most popular datasets worldwide for LULC mapping and change detection studies (Kumar et al. 2012; Wang et al. 2009; Odindi et al. 2012).

Numerous methods and techniques have been used for preparing satellite image-based land cover maps in the past (Li et al. 2014). These methods include unsupervised and supervised classification approaches, as well as parametric and non-parametric methods. In very recent times, non-parametric methods based on the machine learning approach have gained tremendous consideration for satellite image-based LULC classification and mapping. Researchers across the globe have been carrying out several studies on LULC mapping and modeling by employing various machine learning algorithms (Civco 1993; Pal 2005; Teluguntla et al. 2018; Talukdar et al.

2020a, b). Also, comparison-based studies have been carried out wherein multiple machine learning-based models, and conventional models were employed for image classification (Rogan et al. 2008; Camargo et al. 2019). The most favored and in-demand algorithms include Support Vector Machine (SVM), Random Forests (RF), and k-Nearest Neighbors (k-NN) (Huang et al. 2002; Franco-Lopez et al. 2001; Kennedy et al. 2015). Object-based classification is another preferred approach that has proved to be very useful for classifying fine resolution satellite images (Machala and Zejdová 2014). The most critical aspect in LULC mapping is classification accuracy, which plays a crucial role as a deciding factor in selecting the best among various classification methods. The significant elements affecting the classification accuracy are the type of satellite sensor, spatial resolution, training data sources, accuracy assessment data sources, the total number of classes, and the classification approach (Manandhar et al. 2009). Consequently, the most critical factor is choosing suitable classification algorithms for achieving acceptable classification accuracy (Lu and Weng 2007).

LULC mapping in mountainous terrain has always been challenging due to hills, valleys, plateaus, and mountains. This complexity in the landscape introduces effects like shadows and illumination issues due to aspect variation causing severe changes in the surface reflectance of various LULC types (Wang et al. 2020). Moreover, complex topography also poses challenges in water body identification. Advanced machine learning-based image classification techniques have been producing promising results, and therefore, their use to generate LULC maps in complex terrain must be explored.

Keeping the aforementioned into consideration, the following are the specific objectives of the present study:

1. To prepare Land Use/Land Cover maps for the Alaknanda River Basin (ARB) using three approaches.
2. To compare the classification results obtained from MLC, SVM, and RT algorithms based on accuracy assessment.

The inferences drawn from this study will provide practical insights into using machine learning techniques for performing LULC classification, especially in snow-covered mountainous regions. Also, the methodology presented can be replicated to classify complex topographic settings elsewhere.

Materials and methods

Study area

The study focusses on the Alaknanda River Basin (ARB), which lies between 78°33' E to 80°15' E longitude and 29°59'24" N to 31°04'51" N latitude in the Northwestern

Himalayan region of Uttarakhand in India and encompasses a catchment area of 11,035.3 km² (Fig. 1) with a total length of 183.5 km. Alaknanda originates from the Satopanth, and the Bhagirath Kharak glaciers flow downstream and meet the Bhagirathi River at Devprayag to finally become river Ganga or the Ganges. The River Alaknanda is joined by the Dhauliganga, Nandakini, Mandakini, Pindar and some other tributaries as shown in Fig. 1. The catchment area features a variety of climates, viz. subtropical, temperate, sub-alpine, and alpine, primarily due to the substantial variation in the altitude ranging from 446 to 7801 m. ARB has a unique topographic setting featuring high mountain peaks and glaciated valleys, especially in the northern part (Sharma and Mohanty 2018). The area is characterized by high relief, steep slopes, and high drainage density (Panwar et al. 2017). Furthermore, the topography is generally represented by north–south trending ridges and incised river valleys (Ghosh et al. 2019). Figure 1 shows clearly that the river valleys are very narrow in the upper part catchment, relatively narrow in the middle, and relatively wider in the lower reach.

Data and sources

The satellite data-based LULC classification was carried out using freely available 30 m (spatial resolution) Landsat 8 OLI Level 1 images. The dataset was downloaded from the ‘EarthExplorer’ official website of the United States Geological Survey (USGS) (<https://earthexplorer.usgs.gov/>). A total of 2 multi-band images (Table 1) were downloaded for this purpose. Each L8/OLI image has data in 11 bands

Table 1 Details of satellite data downloaded for LULC classification

S. No	Date of acquisition	Row ID	Path ID	Cloud cover %
1	18-04-2018	39	145	6
2	11-05-2018	39	146	1

but, for classification purposes, data for bands 2 to 7 and 9 (details listed in Table 2) were downloaded.

Tools and techniques used

The Semi-Automatic Classification Plugin (SCP) version 6.4.0.2 in QGIS 3.14 software interface was used for pre-processing the downloaded L8/OLI images, and after that, ArcGIS 10.4 was used for classification and accuracy assessment. ENVI software was used for the estimation of the Normalized Difference Snow Index. For the classification in the ArcGIS interface, three different training algorithms (classifiers) from the spatial analyst toolbox were put into use, namely, MLC, SVM, and RT.

Methodology

This study focused on classifying the mountainous river basin and preparing LULC maps using three different classification algorithms. The resulting maps were compared on the basis of the accuracy assessment of each classified

Fig. 1 Location map of the study area

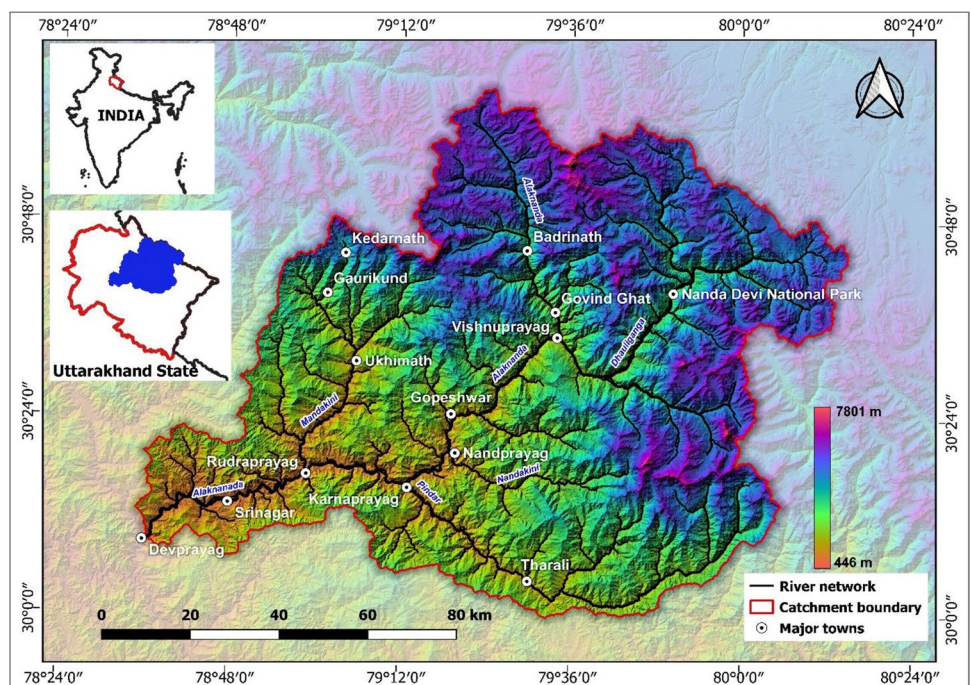


Table 2 Landsat 8 OLI individual bands wavelength range and resolution

Landsat 8 Bands	Wavelength [micro meters]	Resolution [meters]
Band 1—Coastal aerosol	0.43–0.45	30
Band 2—Blue	0.45–0.51	30
Band 3—Green	0.53–0.59	30
Band 4—Red	0.64–0.67	30
Band 5—Near Infrared (NIR)	0.85–0.88	30
Band 6—SWIR 1	1.57–1.65	30
Band 7—SWIR 2	2.11–2.29	30
Band 8—Panchromatic	0.50–0.68	15
Band 9—Cirrus	1.36–1.38	30
Band 10—Thermal Infrared (TIRS) 1	10.60–11.19	100 (resampled to 30)
Band 11—Thermal Infrared (TIRS) 2	11.50–12.51	100 (resampled to 30)

The table has been imported from page 114 of Semi-Automatic Classification Plugin Documentation (2016) by Luca Congedo (licensed under a Creative Commons Attribution-ShareAlike 4.0 International License)

image. The flowchart in Fig. 2 depicts the sequence of steps followed as part of the methodology adopted in this study.

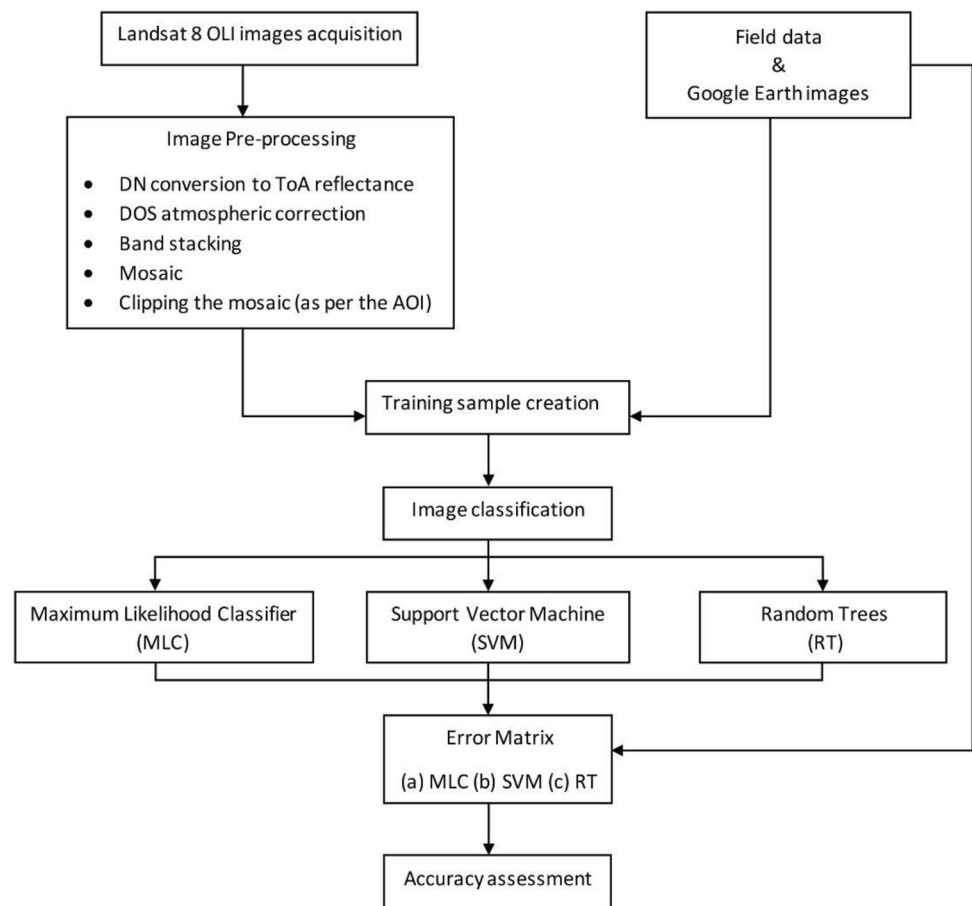
The downloaded satellite images were pre-processed, where the Digital Number (DN) values in the images were converted to the Top of Atmosphere (ToA) Reflectance. During this process, the Dark Object Subtraction (DOS) atmospheric corrections were also applied using the SCP tool in QGIS. All the pre-processed image bands were then stacked together and clipped using the vector boundary of the study area. Both the clipped band stacks were then mosaiced together to create a single image of the entire river basin. The image was classified into nine different classes, namely, Snow, Forest, Sparse vegetation (or barren areas), Built-up, River bed, Water, Agriculture, Cloud, and Shadow. Concepts of visual image interpretation were used to identify each class and consequently assigned to the image pixels for image classification.

For the image classification training, sample polygons were digitized using field data, Google Earth Imagery, and different band combinations of the satellite image. A total of seven band combinations (R-G-B) were used to select training samples, namely, Natural Color Composite (4–3–2), False Color Composite (FCC)-Urban (7–6–4), Color Infrared Composite-Vegetation (5–4–3), Agricultural Area Composite (6–5–2), Healthy Vegetation Composite (5–6–2) and Land Water Composite (5–6–4). The FCC of the two images (Fig. 3a) with a band combination of 7–5–3 is depicting the presence of snow/ice (cyan color) in the high elevation areas. Figure 3b shows the Natural Color Composite with band combination 5–4–3, wherein the snow-covered mountains are prominently visible. The same composite was used for the selection of training samples to classify snow. Also, to verify the classification of snow, the Normalized Difference Snow Index (NDSI) as shown in Fig. 3c was calculated using the following general equation: $NDSI = (Green - SWIR) /$

$(Green + SWIR)$. Based on the previous studies (Hall et al. 1998; Nijhawan et al. 2016) a threshold of 0.4 was selected to distinguish between snow and no snow areas. The process was carried out in ENVI software, using the Band Math tool. Even though the satellite images used in the analysis were acquired in April and May, the snow in the region is justified by the presence of glaciers in the high elevation zones of the Alaknanda River Basin. The higher elevation areas in the northern part of the study area remain snow-covered all round the year. Therefore, to appreciate and associate snow/ice in the study area, an elevation map of the study area using SRTM DEM having a spatial resolution of 30 m was prepared, as shown in Fig. 3d. This representation correlates with the classification of snow.

A signature file with samples for all nine classes was developed to classify the satellite image. The classification was carried out using the Maximum Likelihood Classifier (MLC), Support Vector Machine Classifier (SVM), and Random Trees Classifier (RT). The spectral signature file containing samples for each class was used to generate classified images in each of the classification approaches. Finally, an accuracy assessment was carried out for all the three classified images separately. For the accuracy assessment, reference points shapefile was created using ground truth field data and points identified using Google Earth. A total of 946 points were identified as ground truth references to be compared with the corresponding pixels in the classified image. The reference points or the test pixels were chosen through random sampling, but it was made sure that they were distinct from the training pixels. The number of reference points for ‘Snow’, ‘Forest’, ‘Sparse vegetation’, ‘Built up’, ‘River bed’, ‘Agriculture’, ‘Clouds’, ‘Shadows’ and ‘Water’ were 131, 158, 46, 186, 18, 170, 41, 33 and 163, respectively. A pivot table of reference pixels class and classified pixels class was created in ArcGIS interface

Fig. 2 Methodology flowchart



and then exported to Microsoft excel for the generation of error matrix and calculation of kappa coefficient for all three classified images. Additionally, the Producer’s and User’s accuracies for each class were also computed to access the accuracy of classification. The following Eqs. (1 and 2) were adopted for calculating producer’s accuracy and user’s accuracy:

$$Producer's\ accuracy\ (\%) = \left(\frac{x_{ii}}{x_{+i}} \right) * 100 \tag{1}$$

$$User's\ accuracy\ (\%) = \left(\frac{x_{ii}}{x_{i+}} \right) * 100 \tag{2}$$

Equations (3 and 4) adopted for the overall accuracy and kappa coefficient calculation are:

$$Overall\ accuracy = \frac{1}{N} \sum_{i=1}^c n_i \tag{3}$$

$$Kappa\ coefficient = \left\{ N \sum_{i=1}^c x_{ii} - \sum_{i=1}^c (x_{i+} * x_{+i}) \right\} \div \left\{ N^2 - \sum_{i=1}^c (x_{i+} * x_{+i}) \right\} \tag{4}$$

where N represents total pixels; c represents the total number of classes; x_{ii} = total number of pixels in row ‘i’ and column ‘i’; x_{i+} = total number of samples in a row ‘i’; x_{+i} = total number of samples in column ‘i’ in the error matrix.

Since cloud detection is a significant challenge in any LULC map, the majority of LULC mapping is carried out using cloud-free satellite imageries, or the clouds are masked, and the masked portion of satellite imagery is obtained from the imagery of another year, preferably of the same month of the year. However, in this study, attempt has been made to classify the clouds and their shadows using the pixel values. The exact cloud and shadow pixel values were identified using the Landsat Quality Assessment Toolbox extension for ArcGIS. The Quality Assessment (QA) band was downloaded to use this tool effectively, wherein each pixel contains an integer value that represents bit packed combinations of surface, atmospheric, and sensor conditions. The integer values for cloud and shadow pixels were identified and extracted using the tool mentioned above in the ArcGIS interface, and then it was compared with the area classified as clouds by the

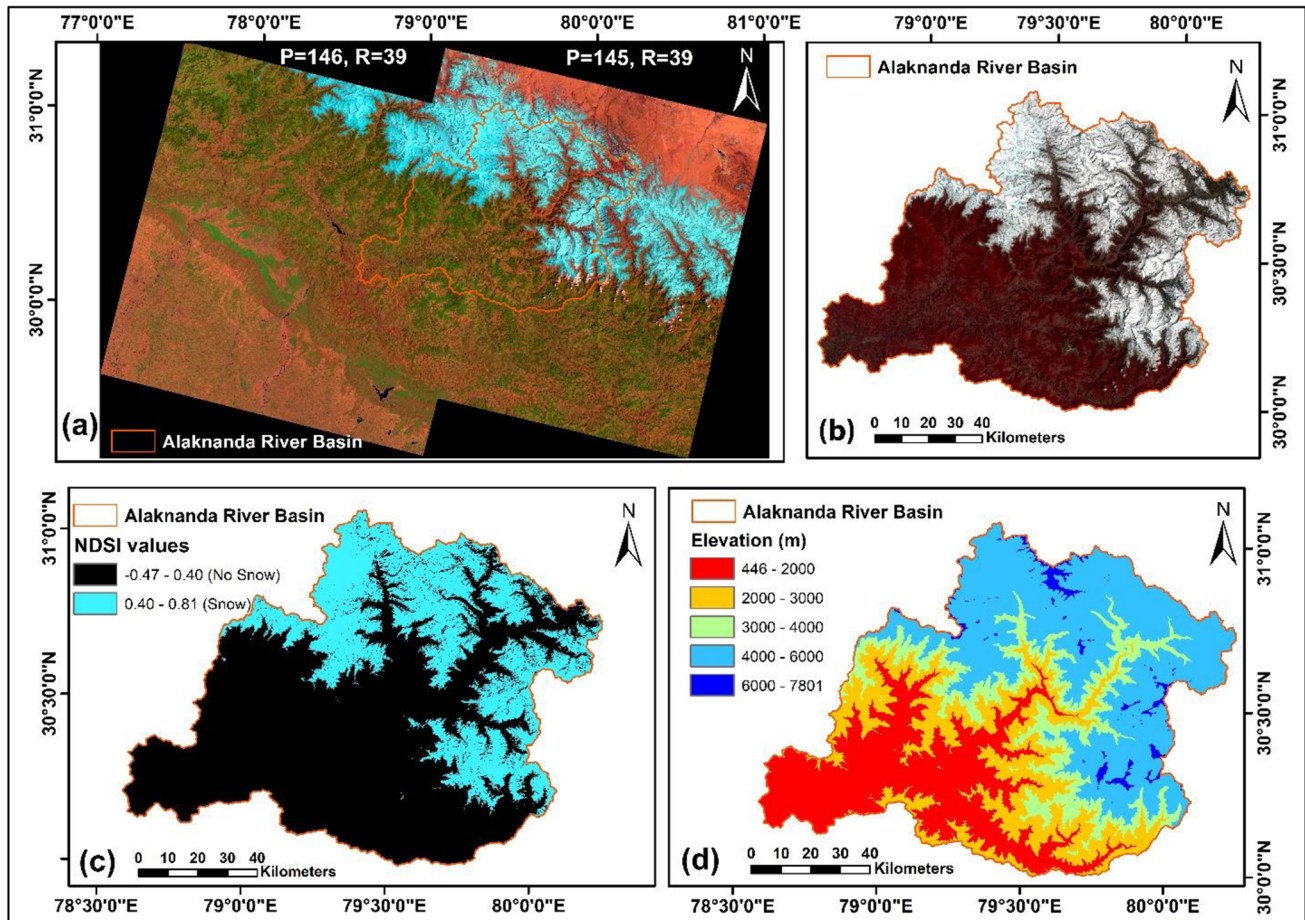


Fig. 3 a FCC of the satellite images downloaded for classification b Clipped Natural Color Composite of ARB c Normalized Difference Snow Index map of the study area d Elevation maps of the study area

classifiers. Using this approach, it was possible to classify the clouds with error-free accuracy using all three classification approaches.

Results and discussion

The LULC maps for the snow-fed Alaknanda River Basin of the Northwest Himalayan region were prepared using three different classification algorithms; namely, MLC, SVM, and RT are presented in Fig. 4. Since the river network in the area is relatively narrow, it is not prominently visible in the maps presented. Therefore, a zoomed-in portion of the southwestern part of the catchment has been shown in Fig. 5 to visually appreciate the difference in classified maps obtained by the three approaches. Furthermore, the accuracy assessment was conducted for each of these maps to compare and evaluate the efficiency of these algorithms. The accuracy assessment results for the LULC maps generated using ML,

SVM, and RT classifiers are presented in Tables 3, 4, and 5, respectively.

Table 3 presents the accuracy assessment of the LULC map obtained using MLC. It shows that with a kappa coefficient of 0.79, the overall accuracy of the classification is 82%. The producer's accuracy of 'Forest', 'Snow' and 'Clouds' is above 90%; 'Agriculture' and 'Sparse vegetation' were classified with a producer's accuracy of above 80%; 'Built up' area with over 75%; 'River bed' and 'shadows' between 60 to 70% and 'Water' having least and unsatisfactory producer's accuracy of 57.93%. As far as the user's accuracy is concerned, except the 'Built up' class, which shows a very low accuracy of 40.22%, all the remaining classes show a user's accuracy of greater than 70%.

Table 4 presents the accuracy assessment of the LULC map obtained using SVM. It shows a slightly higher kappa coefficient of 0.82 and an improved overall accuracy of 84%. The producer's accuracies of 'Snow', 'Forest', 'Sparse vegetation', 'River bed', 'Clouds' and 'Shadows' are above 90%. Furthermore, SVM shows a remarkable improvement in the producer's accuracy of 'water' as high as 86.59%.

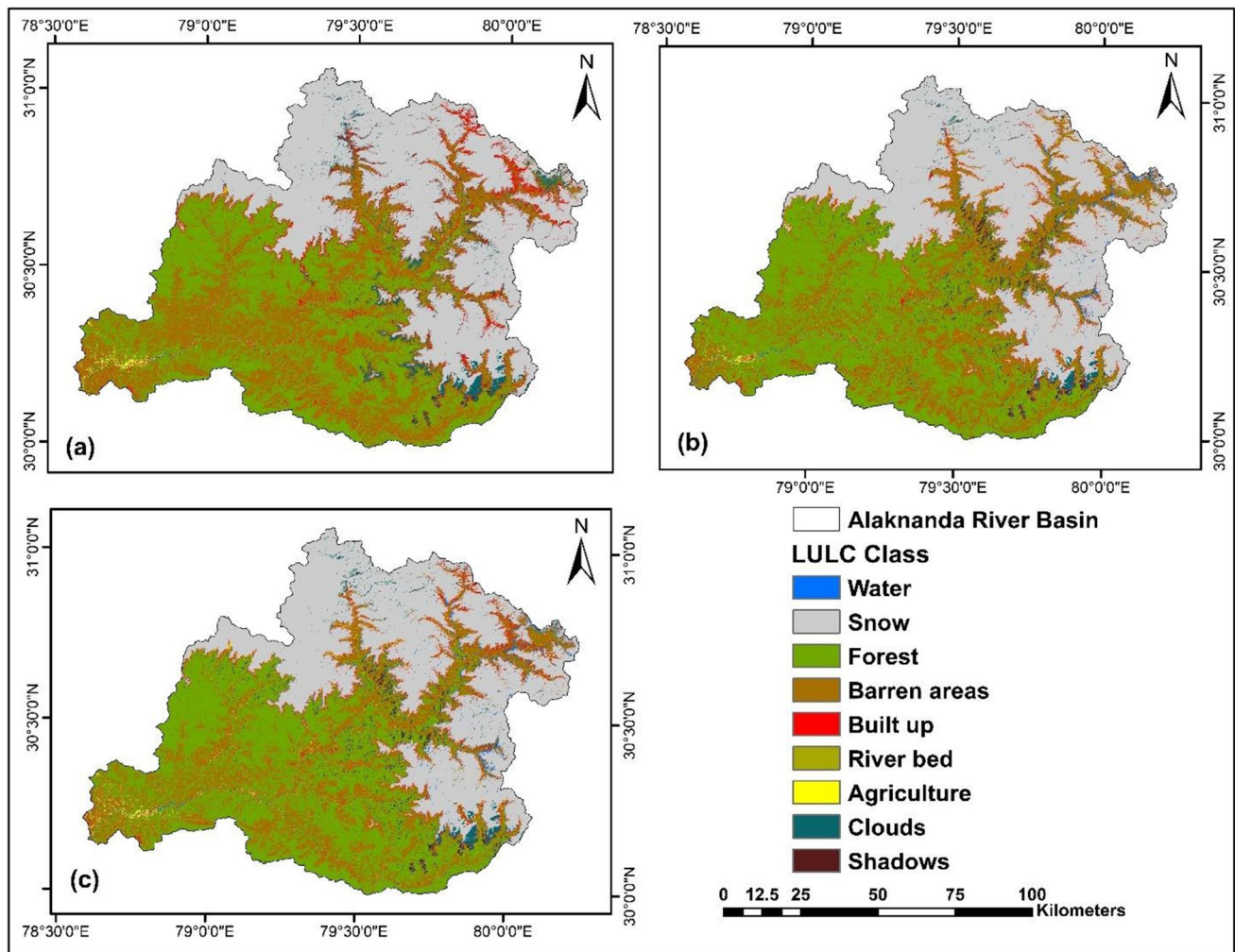


Fig. 4 LULC maps for the Alaknanda River Basin generated using (a) Maximum Likelihood Classifier, (b) Support Vector Machine, and (c) Random Trees

However, the producer’s accuracy of ‘Built up’ class is as low as 59.68%, which is not satisfactory. The SVM approach obtained a user’s accuracy of 42% and 68% for ‘Sparse vegetation’ and ‘River bed’, respectively. The user’s accuracy for the remaining classes is above 80%. Thus, SVM classification yielded much better results than the ML classification algorithm. The objective of SVM classification is to identify the optimal boundary between various classes or samples, also referred to as the support vectors. It is a binary classifier capable of identifying a single boundary between two separate classes (Maxwell et al. 2018).

Finally, Table 5 presents the accuracy assessment of the LULC map obtained using an RT classifier. This classification approach shows the highest values for the kappa coefficient (0.83) and overall accuracy (85%). The producer’s accuracies for ‘Snow’, ‘Forest’, ‘Sparse vegetation’, ‘Water’, ‘Cloud’ and ‘Shadow’ are above 90%. ‘Agriculture’ class shows a good producer’s accuracy of 82.35%, while

for ‘Built up’ and ‘River bed’, it is 54.30% and 61.11%, respectively.

Visual comparison of the three LULC maps presented in Fig. 4 shows a similar pattern for the two machine learning approaches. However, the map classified using MLC looks different. Figure 4a shows that the Maximum Likelihood Classifier has classified a considerable area under as built up in the northeastern part of the basin. On the contrary, Fig. 4b and c shows that the same area has minimal built-up areas as classified by the machine learning algorithms. Another interesting visual comparison is for the sparse vegetation or barren areas class. MLC has classified a considerably large number of pixels in this class as compared to SVM and RT algorithms.

Similarly, comparing the share of area under each class using all three classification approaches reveals a mix of minor and noteworthy differences. To appreciate the percentage of the area classified under each LULC class, a

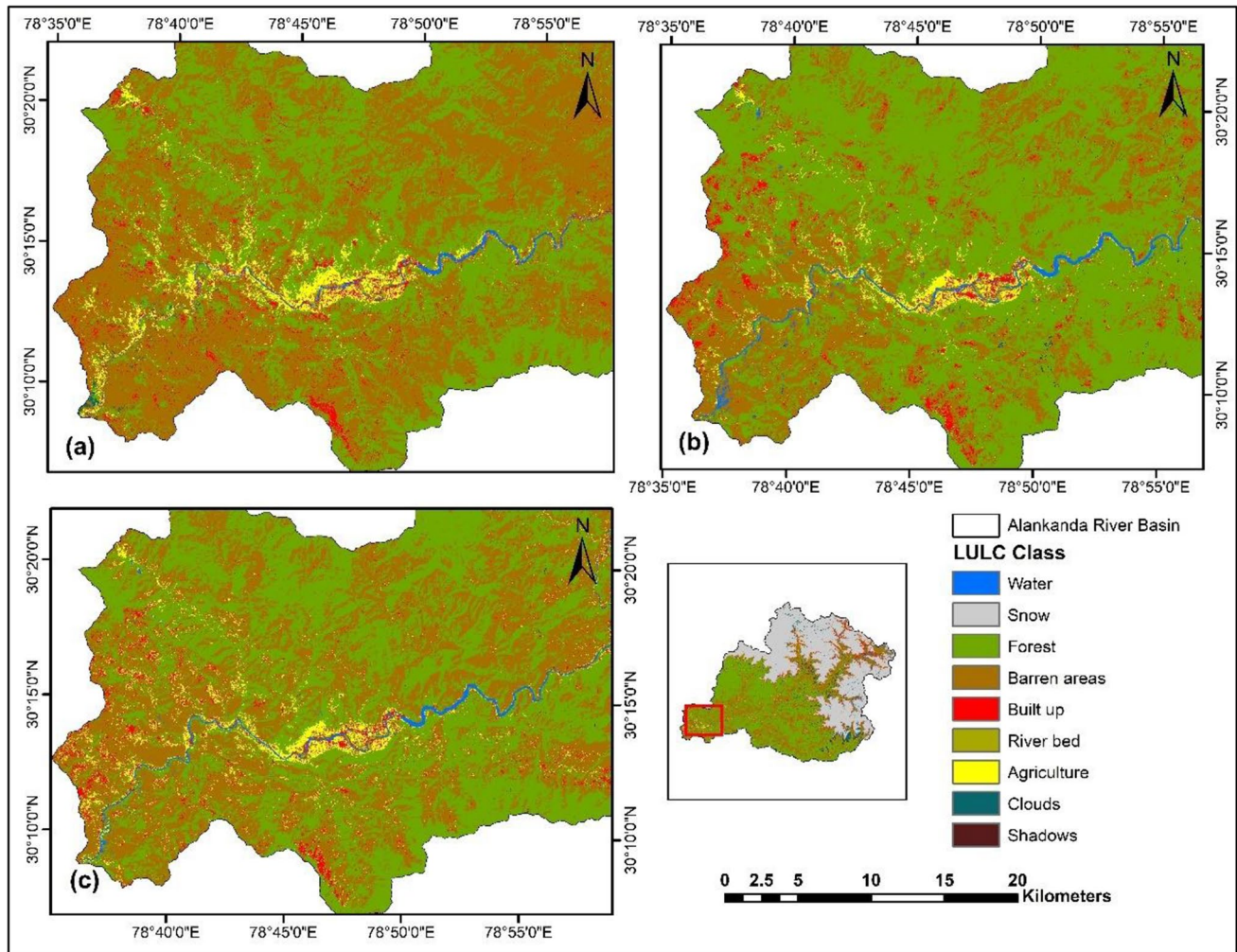


Fig. 5 LULC status of the southwestern portion of ARB (a) Maximum Likelihood Classifier (b) Support Vector Machine and (c) Random Trees

quantitative comparison of the classification results was performed and is presented in the form of bar charts in Fig. 6. The bar charts clearly indicate the variation of classification output generated using MLC, SVM, and RT algorithms.

In this study, particular emphasis is directed towards classifying water using medium resolution satellite data in complex mountainous terrain. Since the Alankanda River Basin features narrow river stretches and steep valleys, it becomes challenging to identify water pixels in optical imagery. Therefore, multiple locations along the river course were carefully selected to prepare the training samples, and classification results indicate that MLC failed to capture water bodies, but both the machine learning approaches performed well. For a better visual representation of water body classification, a subset of the study area from the southwestern part of the catchment is presented in Fig. 5. It is visually evident that MLC has not been able to capture the river, especially in the southwest part of the map (Fig. 5a) while, SVM and RT have detected the river quite well (Fig. 5b, c).

In light of the classification results, the above discussion strongly suggests that machine learning algorithms (SVM and RT) have outperformed the conventional MLC technique to generate LULC maps for the snow-covered mountainous Alankanda River Basin. The results are in line with other studies wherein several machine learning models like SVM, random forest (RF), radial basis function (RBF), decision tree (DT), and naïve bayes (NB) have performed better in comparison to the conventional classification approaches (Ma et al. 2019; Shih et al. 2019; Talukdar et al. 2020a, b). Researchers have concluded explicitly that SVM and RF (Mountrakis et al. 2011; Ma et al. 2017; Carranza et al. 2019) are the best ML-based image classification techniques.

Summary and conclusions

It has been reported that Machine Learning is capable of generating a classification with higher accuracy for the remotely sensed satellite data in comparison to the

Table 3 Error matrix showing the Kappa coefficient and overall accuracy for the LULC map generated using the Maximum Likelihood Classifier

Class	Snow	Forest	Sparse vegetation	Built up	River bed	Agriculture	Cloud	Shadow	Water	Predictions	User's Accuracy
Snow	130	0	0	1	0	0	0	6	42	179	72.63
Forest	0	144	0	1	0	6	0	3	0	154	93.51
Sparse vegetation	0	11	37	29	2	7	0	1	5	92	40.22
Built-up	0	0	5	141	5	7	0	0	5	163	86.50
River bed	0	0	0	0	11	0	0	0	3	14	78.57
Agriculture	0	2	0	14	0	150	0	0	0	166	90.36
Clouds	1	0	4	0	0	0	41	0	10	56	73.21
Shadows	0	0	0	0	0	0	0	23	4	27	85.19
Water	0	0	0	0	0	0	0	0	95	95	100
Total (truth count)	131	157	46	186	18	170	41	33	164	946 points	
Producer's accuracy (%)	99.24	91.72	80.43	75.81	61.11	88.24	100	69.70	57.93		
Overall accuracy	82										
Kappa	0.79										

Table 4 Error matrix showing the Kappa coefficient and overall accuracy for the LULC map generated using the SVM classifier

Class	Snow	Forest	Sparse vegetation	Built up	River bed	Agriculture	Cloud	Shadow	Water	Predictions	User's Accuracy
Snow	130	0	0	0	0	0	0	0	7	137	94.89
Forest	0	150	0	2	0	11	0	0	2	165	90.91
Sparse vegetation	0	6	42	40	0	9	0	0	3	100	42
Built-up	0	0	1	111	1	22	1	0	1	137	81.02
River bed	0	0	3	2	17	0	0	0	3	25	68
Agriculture	0	0	0	31	0	128	0	0	0	159	80.50
Clouds	1	0	0	0	0	0	40	0	0	41	97.56
Shadows	0	0	0	0	0	0	0	33	6	39	84.62
Water	0	1	0	0	0	0	0	0	142	143	99.30
Total (truth count)	131	157	46	186	18	170	41	33	164	946 points	
Producer's accuracy (%)	99.24	95.54	91.30	59.68	94.44	75.29	97.56	100	86.59		
Overall accuracy	84										
Kappa	0.82										

Table 5 Error matrix showing the Kappa coefficient and overall accuracy for the LULC map generated using the RT classifier

Class	Snow	Forest	Sparse vegetation	Built up	River bed	Agriculture	Cloud	Shadow	Water	Predictions	User's Accuracy
Snow	130	0	0	0	0	0	0	0	0	130	100
Forest	0	151	0	1	0	6	0	0	2	160	94.38
Sparse vegetation	0	6	45	36	2	15	0	0	3	107	42.06
Built up	0	0	1	101	5	9	0	0	1	117	86.32
River bed	0	0	0	3	11	0	0	0	0	14	78.57
Agriculture	0	0	0	44	0	140	0	0	1	185	75.68
Clouds	1	0	0	0	0	0	41	0	0	42	97.62
Shadows	0	0	0	0	0	0	0	33	7	40	82.5
Water	0	0	0	1	0	0	0	0	150	151	99.34
Total (truth count)	131	157	46	186	18	170	41	33	164	946 points	
Producer's accuracy (%)	99.24	96.18	97.83	54.30	61.11	82.35	100	100	91.46		
Overall accuracy	85										
Kappa	0.83										

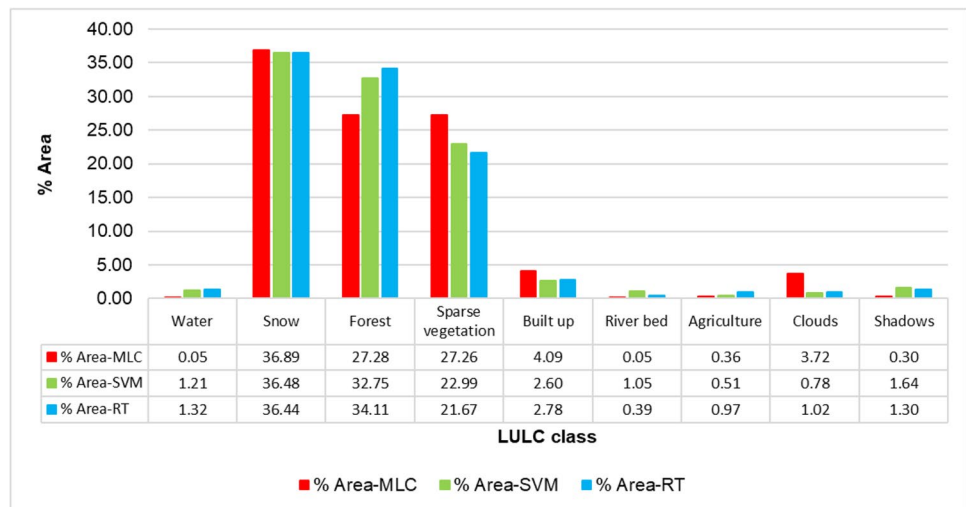
parametric approaches like Maximum Likelihood (Maxwell et al. 2018). In the current study, an attempt has been made to compare three different classification approaches, namely ML, SVM, and RT, to prepare LULC maps for the snow-fed Alaknanda River Basin of the Northwest Himalayan region. The following conclusions are drawn from this study:

- The highest values of the Kappa coefficient and overall accuracy were obtained using the RT approach, followed by SVM and ML. This shows that advanced machine learning-based classifiers performed better than the parametric Maximum Likelihood Classifier in the Alaknanda River Basin.
- Snow has been classified with a very high accuracy of 99.24% in all the classification approaches. Forest has also been classified with high accuracy in all three methods, but SVM and RT are better than ML. Similarly, Sparse vegetation has been most accurately classified by RT, followed by SVM and ML.
- Classification of Built-up shows flipped results where ML obtained the highest accuracy of 75.81. On the other hand, both SVM and RT performed poorly, with accuracies of less than 60%. Also, Agriculture was best classified by the ML classifier, followed by RT and SVM.
- The SVM classifier outperformed the 'River bed' classification. In all three approaches, the clouds were very accurately classified. ML and RT showed a cent percent accuracy.
- The shadow pixels were classified with an accuracy of 70% by ML classifier, but the machine learning-based approaches classified these with 100% accuracy.
- Finally, the water pixels in the satellite images were most accurately classified by the RT classifier followed by SVM and were poorly classified by the ML classifier.

Finally, it can be concluded that for the snow-fed Alaknanda River Basin, the advanced machine learning-based parametric classifiers have performed better than the Maximum Likelihood Classifier. The results cannot be generalized, but the accuracy assessment shows that Machine Learning based classifiers have outperformed in accurately classifying the majority of the classes. Especially for the watershed LULC mapping, accurate classification of water bodies is of utmost importance. This study suggests that the RT classifier is the best of three for accurately classifying water bodies. This particular inference signifies a robust utility of advanced machine learning algorithms for performing LULC classification.

Furthermore, the methodology can be replicated in snow-fed mountainous regions elsewhere. Change

Fig. 6 Percentage of the area under different classes for each classification approach



detection studies can also be conducted using RT, SVM, or even other machine learning-based algorithms using freely available medium resolution satellite data products. However, it can be argued that microwave remote sensing data is most preferred to classify water, but in highly complex and undulated terrain, the occurrence of foreshortening and shadow effect introduces errors in microwave data. Therefore, the study shows that the Machine learning-based classification approaches improve water detection capability and LULC mapping functionality using satellite-based remote sensing data. Moreover, researchers in the remote sensing domain have increasingly interested in exploring and employing advanced machine learning algorithms for image classification (Rodriguez-Galiano et al. 2012; Yeom et al. 2013; Jamali 2020). The field is developing rapidly, and new algorithms and implementations are becoming available continuously. The application of machine learning algorithms in LULC classification can result in high-quality results, as the classification results of this research shows.

Acknowledgements We wish to express a deep sense of gratitude and sincere thanks to the Department of Water Resources Development and Management (WRD&M), IIT Roorkee, for providing a conducive environment and resources to conduct the research work.

Data availability The authors confirm that the data supporting the findings of this study are available within the manuscript in the form of tables.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Atzberger C (2013) Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens* 5(2):949–981. <https://doi.org/10.3390/rs5084124>
- Barros JL, Tavares AO, Santos PP (2021) Land use and land cover dynamics in Leiria City: relation between peri-urbanization processes and hydro-geomorphologic disasters. *Nat Hazards* 106(1):757–784. <https://doi.org/10.1007/s11069-020-04490-y>
- Bontemps S, Bogaert P, Titeux N, Defourny P (2008) An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution. *Remote Sens Environ* 112(6):3181–3191. <https://doi.org/10.1016/j.rse.2008.03.013>
- Camargo FF, Sano EE, Almeida CM, Mura JC, Almeida TA (2019) Comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. *Remote Sens* 11:1600. <https://doi.org/10.3390/rs11131600>
- Campbell JB, Wynne RH (2011) Introduction to remote sensing. Guilford Press. <https://doi.org/10.1111/phor.12021>
- Carranza-García M, García-Gutiérrez J, Riquelme JC (2019) A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sens* 2019(11):274. <https://doi.org/10.3390/rs11030274>
- Civco DL (1993) Artificial neural networks for land-cover classification and mapping. *Int J Geogr Inf Sci* 7:173–186. <https://doi.org/10.1080/02693799308901949>
- Congedo L (2016). Semi-Automatic Classification Plugin Documentation. <https://doi.org/10.13140/RG.2.2.29474.02242/1>
- Costa MH, Botta A, Cardille JA (2003) Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. *J Hydrol* 283(1–4):206–217. [https://doi.org/10.1016/S0022-1694\(03\)00267-1](https://doi.org/10.1016/S0022-1694(03)00267-1)
- Dash PP, Kakkur R, Shreenivas V, Prakash PJ, Mythri DJ, Kumar KV, ... & Sahai RM N. (2015). “Quantification of urban expansion using geospatial technology—a case study in Bangalore.” *Adv Remote Sens*, 4(04), 330. <https://doi.org/10.4236/ars.2015.44027>

- DeFries RS, Rudel T, Uriarte M, Hansen M (2010) Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nat Geosci* 3(3):178–181. <https://doi.org/10.1038/ngeo756>
- Dwivedi RS, Sreenivas K, Ramana KV (2005) Land-use/land-cover change analysis in part of Ethiopia using Landsat Thematic Mapper data. *Int J Remote Sens* 26(7):1285–1287. <https://doi.org/10.1080/01431160512331337763>
- Franco-Lopez H, Ek AR, Bauer ME (2001) Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens Environ* 77(3):251–274. [https://doi.org/10.1016/S0034-4257\(01\)00209-7](https://doi.org/10.1016/S0034-4257(01)00209-7)
- Ghosh TK, Jakobsen F, Joshi M, Pareta K (2019) Extreme rainfall and vulnerability assessment: case study of Uttarakhand rivers. *Nat Hazards* 99(2):665–687. <https://doi.org/10.1007/s11069-019-03765-3>
- 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>
- Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, ... & Li C (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int J Remote Sens*, 34(7), 2607–2654. <https://doi.org/10.1080/01431161.2012.748992>
- Hall DK, Foster JL, Verbyla DL, Klein AG, Benson CS (1998) Assessment of snow-cover mapping accuracy in a variety of vegetation-cover densities in central Alaska. *Remote Sens Environ* 66(2):129–137. <https://doi.org/10.1080/01431160110040323>
- Himanshu SK, Pandey A, Shrestha P (2017) Application of SWAT in an Indian river basin for modeling runoff, sediment and water balance. *Environ Earth Sci* 76(1):1–18. <https://doi.org/10.1007/s12665-016-6316-8>
- Huang C, Davis LS, Townshend JRG (2002) An assessment of support vector machines for land cover classification. *Int J Remote Sens* 23(4):725–749. <https://doi.org/10.1080/01431160110040323>
- Jamali A (2020) Land use land cover mapping using advanced machine learning classifiers: a case study of Shiraz city, Iran. *Earth Sci Inf* 13(4):1015–1030. <https://doi.org/10.1007/s12145-020-00475-4>
- Kennedy RE, Yang Z, Braaten J, Copass C, Antonova N, Jordan C, Nelson P (2015) Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. *Remote Sens Environ* 166:271–285. <https://doi.org/10.1016/j.rse.2015.05.005>
- Kim J, Choi J, Choi C, Park S (2013) Impacts of changes in climate and land use/land cover under IPCC RCP scenarios on streamflow in the Hoeya River Basin, Korea. *Sci Total Environ* 452:181–195. <https://doi.org/10.1016/j.scitotenv.2013.02.005>
- Li C, Wang J, Wang L, Hu L, Gong P (2014) Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sens* 6(2):964–983. <https://doi.org/10.3390/rs6020964>
- Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. *Int J Remote Sens* 28(5):823–870. <https://doi.org/10.1080/01431160600746456>
- Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson BA (2019) Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS J Photogramm Remote Sens* 2019(152):166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Ma L, Li M, Ma X, Cheng L, Du P, Liu Y (2017) A review of supervised object-based land-cover image classification. *ISPRS J Photogramm Remote Sens* 130:277–293. <https://doi.org/10.1016/j.isprsjprs.2017.06.001>
- Machala M, Zejdová L (2014) Forest mapping through object-based image analysis of multispectral and LiDAR aerial data. *Eur J Remote Sens* 47(1):117–131. <https://doi.org/10.5721/eujrs20144708>
- Mahmood R, Pielke RA Sr, Hubbard KG, Niyogi D, Dirmeyer PA, McAlpine C, Hale R, Gameda S, Beltrán-Przekurat A, Baker B (2014) Land cover changes and their biogeophysical effects on climate. *Int J Climatol* 34(4):929–953. <https://doi.org/10.1002/joc.3736>
- Malik MI, Bhat MS (2014) Integrated approach for prioritizing watersheds for management: a study of Lidder catchment of Kashmir Himalayas. *Environ Manage* 54(6):1267–1287. <https://doi.org/10.1007/s00267-014-0361-4>
- Manandhar R, Odeh IO, Ancev T (2009) Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. *Remote Sens* 1(3):330–344. <https://doi.org/10.3390/rs1030330>
- Mao D, Cherkauer KA (2009) Impacts of land-use change on hydrologic responses in the Great Lakes region. *J Hydrol* 374(1–2):71–82. <https://doi.org/10.1016/j.jhydrol.2009.06.016>
- Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. *Int J Remote Sens* 39(9):2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Mishra PK, Rai A, Rai SC (2020) Land use and land cover change detection using geospatial techniques in the Sikkim Himalaya, India. *Egypt J Remote Sens Space Sci* 23(2):133–143. <https://doi.org/10.1016/j.ejrs.2019.02.001>
- Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* 2011, 66, 247–259.
- Nie W, Yuan Y, Kepner W, Nash MS, Jackson M, Erickson C (2011) Assessing impacts of Landuse and Landcover changes on hydrology for the upper San Pedro watershed. *J Hydrol* 407(1–4):105–114. <https://doi.org/10.1016/j.jhydrol.2011.07.012>
- Nijhawan R, Garg PK, Thakur PK (2016) Monitoring of glacier in Alaknanda basin using remote sensing data. *Perspect Sci* 8:381–383. <https://doi.org/10.1016/j.pisc.2016.04.081>
- Güler M, Yomralıoğlu T, Reis S (2007) Using landsat data to determine land use/land cover changes in Samsun, Turkey. *Environ Monit Assess* 127(1–3):155–167. <https://doi.org/10.1007/s10661-006-9270-1>
- Odindi J, Mhangara P, Kakembo V (2012) Remote sensing land-cover change in Port Elizabeth during South Africa's democratic transition. *S Afr J Sci* 108(5–6):60–66. <https://doi.org/10.4102/sajs.v108i5/6.886>
- Owringi AM, Lannigan R, Simonovic SP (2014) Interaction between land-use change, flooding and human health in Metro Vancouver, Canada. *Nat Hazards* 72(2):1219–1230. <https://doi.org/10.1007/s11069-014-1064-0>
- Pal M (2005) Random forest classifier for remote sensing classification. *Int J Remote Sens* 26:217–222. <https://doi.org/10.1080/01431160412331269698>
- Panwar S, Agarwal V, Chakrapani GJ (2017) Morphometric and sediment source characterization of the Alaknanda river basin, headwaters of river Ganga, India. *Nat Hazards* 87(3):1649–1671. <https://doi.org/10.1007/s11069-017-2838-y>
- Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogramm Remote Sens* 67:93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Rogan J, Franklin J, Stow D, Miller J, Woodcock C, Roberts D (2008) Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. *Remote Sens Environ* 112:2272–2283. <https://doi.org/10.1016/j.rse.2007.10.004>
- Roy PS, Roy A, Joshi PK, Kale MP, Srivastava VK, Srivastava SK, Dwevidi RS, Joshi C, Behera MD, Meiyappan P, Sharma Y (2015)

- Development of decadal (1985–1995–2005) land use and land cover database for India. *Remote Sens* 7(3):2401–2430. <https://doi.org/10.3390/rs70302401>
- Saadat H, Adamowski J, Bonnelli R, Sharifi F, Namdar M, Ale-Ebrahim S (2011) Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. *ISPRS J Photogramm Remote Sens* 66(5):608–619. <https://doi.org/10.1016/j.isprsjprs.2011.04.001>
- Sajikumar N, Remya RS (2015) Impact of land cover and land use change on runoff characteristics. *J Environ Manage* 161:460–468. <https://doi.org/10.1016/j.jenvman.2014.12.041>
- Sharma G, Mohanty S (2018) Morphotectonic analysis and GNSS observations for assessment of relative tectonic activity in Alaknanda basin of Garhwal Himalaya, India. *Geomorphology* 301:108–120. <https://doi.org/10.1016/j.geomorph.2017.11.002>
- Shih HC, Stow DA, Tsai YH (2019) Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping. *Int J Remote Sens* 2019(40):1248–1274. <https://doi.org/10.1080/01431161.2018.1524179>
- Shrestha UB, Gautam S, Bawa KS (2012) Widespread climate change in the Himalayas and associated changes in local ecosystems. *PLoS One* 7(5)
- Sierra-Soler A, Adamowski J, Malard J, Qi Z, Saadat H, Pingale S (2016) Assessing agricultural drought at a regional scale using LULC classification, SPI, and vegetation indices: case study in a rainfed agro-ecosystem in Central Mexico. *Geomat Nat Haz Risk* 7(4):1460–1488. <https://doi.org/10.1080/19475705.2015.1073799>
- Singh SK, Srivastava PK, Gupta M, Thakur JK, Mukherjee S (2014) Appraisal of land use/land cover of mangrove forest ecosystem using support vector machine. *Environ Earth Sci* 71(5):2245–2255. <https://doi.org/10.1007/s12665-013-2628-0>
- Sundara Kumar K, Harika M, Begum SA, Yamini S, Balakrishna K (2012) Land use and land cover change detection and urban sprawl analysis of Vijayawada city using multitemporal landsat data. *Int J Eng Sci Technol* 4(01):170–178
- Talukdar S, Singha P, Mahato S, Pal S, Liou YA, Rahman A (2020) Land-use land-cover classification by machine learning classifiers for satellite observations—a review. *Remote Sens* 12(7):1135. <https://doi.org/10.3390/rs12071135>
- Talukdar S, Singha P, Mahato S, Praveen B, Rahman A (2020b) Dynamics of ecosystem services (ESs) in response to land use land cover (LU/LC) changes in the lower Gangetic plain of India. *Ecol Indic* 112:106121. <https://doi.org/10.1016/j.ecolind.2020.106121>
- Teluguntla P, Thenkabail PS, Oliphant A, Xiong J, Gumma MK, Congalton RG, Huete A (2018) A 30-m Landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J Photogramm Remote Sens* 144:325–340. <https://doi.org/10.1016/j.isprsjprs.2018.07.017>
- Thanh HNT, Doan TM, Tomppo E, McRoberts RE (2020) Land use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods—a case study from Dak Nong, Vietnam. *Remote Sens* 12(9):1367. <https://doi.org/10.3390/rs12091367>
- Thenkabail PS, Schull M, Turrall H (2005) Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. *Remote Sens Environ* 95(3):317–341. <https://doi.org/10.1016/j.rse.2004.12.018>
- Wang L, Chen J, Gong P, Shimazaki H, Tamura M (2009) Land cover change detection with a cross-correlogram spectral matching algorithm. *Int J Remote Sens* 30(12):3259–3273. <https://doi.org/10.1080/01431160802562164>
- Wang H, Liu C, Zang F, Yang J, Li N, Rong Z, Zhao C (2020) Impacts of topography on the land cover classification in the Qilian Mountains, Northwest China. *Can J Remote Sens* 46(3):344–359. <https://doi.org/10.1080/07038992.2020.1801401>
- Wondrade N, Dick OB, Tveite H (2014) GIS based mapping of land cover changes utilizing multi-temporal remotely sensed image data in Lake Hawassa Watershed, Ethiopia. *Environ Monit Assess* 186(3):1765–1780. <https://doi.org/10.1007/s10661-013-3491-x>
- Yeom J, Han Y, Kim Y (2013) Separability analysis and classification of rice fields using KOMPSAT-2 High Resolution Satellite Imagery. *Res J Chem Environ* 17:136–144
- Zhang F, Yushanjiang A, Jing Y (2019) Assessing and predicting changes of the ecosystem service values based on land use/cover change in Ebinur Lake Wetland National Nature Reserve, Xinjiang, China. *Sci Total Environ* 656:1133–1144. <https://doi.org/10.1016/j.scitotenv.2018.11.444>