SSRG International Journal of Civil Engineering
ISSN: 2348-8352/ https://doi.org/10.14445/23488352/IJCE-V12I6P109

Volume 12 Issue 6, 91-104, June 2025

Original Article

Random Forest-Based LULC Mapping in and around Itanagar using IRS LISS-IV Multispectral Data

Konthoujam James Singh[1], Ajay Bharti[2], Salam Shantikumar Singh[3]

[1,2]*Department of Civil Engineering, North Eastern Regional Institute of Science and Technology, Nirjuli, Arunachal Pradesh, India.*
[3]*Department of Statistics, Manipur University, Imphal, Manipur, India.*

[1]*Corresponding Author : kjs@nerist.ac.in*

Received: 07 April 2025 Revised: 10 May 2025 Accepted: 09 June 2025 Published: 28 June 2025

Abstract *- Land Use Land Cover information is an indispensable asset in sustainable resources management. The current study investigates the yield of Land Use Land Cover (LULC) information in and around Itanagar, capital of Arunachal Pradesh, India. The area under research consists of five administrative circles, namely Itanagar, Naharlagun, Doimukh, Gumto, and Banderdewa circle, which are studied individually. In this study, the Indian Remote Sensing (IRS) Linear Imaging Self-Scanning Sensor (LISS)-IV multispectral image of 2022 is used in producing the LULC information. The Random Forest classifier in QGIS is employed for image classification, producing five classes in the study area: Dense forest, Degraded forest, Fallow land, Builtup and Water body. Accuracy assessment uses an equal stratified random method to produce the confusion matrix. The results obtained are encouraging with high overall accuracy, 85%, 83%, 78%, 85%, and 79% for Itanagar, Naharlagun, Doimukh, Gumto and Banderdewa circle, respectively and kappa coefficient ranging from 0.81, 0.79, 0.73, 0.82 and 0.74 for Itanagar, Naharlagun, Doimukh, Gumto and Banderdewa circle, respectively. Finally, the LULC map is prepared in ArcGIS.*

Keywords *- Random forest, LULC, Machine Learning, Confusion matrix, Kappa coefficient.*

1. Introduction

The increasing exploration of natural resources on Earth's surface, driven by the growing population, has created a sense of necessity to understand its spatial distribution. The LULC extent and spatial distribution comprise the critical inputs for planning strategic sustainable developmental projects [1]. Land use is human activity, while land cover is the biophysical characteristics of land surface[2]. LULC data are widely used in many studies, some of which are forest management [3], hydrological studies [4, 5] town planning [6], climate change [7], disaster management [8], wildlife habitat management [9, 10] and transportation management [11]. LULC data at various scales have been derived from the remotely sensed data [12-14]. It has been derived using a variety of sensor data, like light detection and ranging data (LiDAR) [15, 16], satellite data [1, 12, 13] and aerial data [17, 18]. Based on spatial resolution, remotely sensed data may be grouped into coarse resolution, medium resolution, high resolution and very high resolution [19, 20]. Coarse spatial resolution images such as Landsat images of 30 m and LISS–II having 36.35 m are found to be used extensively in LULC mapping and studies related to LULC as an important parameter [6, 13, 19, 21]. Among the medium spatial resolution images IRS LISS-III and Sentinental-2A have also been widely used in similar studies [1, 4]. High spatial resolution images, like those of LISS-IV, have been used recently for LULC mapping and other RS-based applications [12, 22]. Very high spatial resolution images are used in studies of road networks [23], urban sprawl [24], and studies where details are more important [20, 25, 26]. Numerous researches have demonstrated that the desired result of LULC mapping is influenced by spatial resolution selected for the remotely sensed image [16, 27, 28]. Some of the commonly adopted classification methods within the remote sensing community are pixel-based unsupervised [29], supervised[30], hybrid classification[3] and random forest [31-33]. Apart from these methods, advanced methods such as object-based image classification [15, 34, 35] and semantic segmentation [36] have also been used. In recent studies, deep learning-based classifications have also been proposed in many research works [36-38].

It has been found that various studies have been conducted within Arunachal Pradesh that involve LULC information [9, 10, 39, 40-44]. In all of these studies, satellite imageries used were of coarse resolution, such as IRS-LISS-I and LISS-III, Landsat 7, Landsat 8, Landsat 9, Landsat-MSS, Landsat 5 TM, Landsat TM and Landsat 1 MSS. Among these studies, [9, 10] studied wildlife sanctuaries in East Siang district and Lower Dibang Valley district, respectively, [41]

studied LULC impact on climate in Lower Dibang Valley district, [43] studies LULC dynamics in Lower Dibang Valley, Lohit and Changlang districts, [44] studied the dynamics of LULC in Pasighat in East Siang district and [40] analysed the landslide possibility of western Arunachal Pradesh in Tawang and West Kameng. Further, [45] studied the land erosion variation in the Mago Basin of Tawang district and Dibang Basin in the Upper and Lower Dibang districts of Arunachal Pradesh using LULC as an important input parameter. However, the raw satellite image from which the LULC information was obtained was not mentioned in the study but resampled to 30m spatial resolution. [39] carried out a study related to urban sprawl for Himalayans, including Itanagar, which required LULC data. Their study used Landsat 1 Multispectral Scanner, Landsat 5 Thematic Mapper and Landsat 8 imageries to obtain the data on LULC.

In another study, [42] determined LULC using Landsat 8 image to map a potential groundwater zone in Arunachal Pradesh's Papum Pare district. [46] emphasised the need for mapping of cities and rural areas, including Itanagar Arunachal Pradesh, in their study of earthquake impacts on Northeastern India. However, their work did not involve such mapping. [47] discussed thoroughly the urbanisation process and changes that occurred in Arunachal Pradesh from 1961 to 2000 based on population size and the growth rate. It was highlighted that urbanisation could be better studied with spatial and temporal mapping, which was missing in the study. [48] studied the exposure of residents of Itanagar to geohazards due to rapid urbanisation.

This study emphasised on certain key issues among which LULC is one of them. However, the study did not map LULC of the study area. The literature assessment indicates that these studies would have yielded significantly different results had they employed higher spatial resolution imagery. Furthermore, it is noted that limited studies have been conducted in the vicinity of Arunachal Pradesh Itanagar, concerning LULC data, and there has been a deficiency of significant study utilizing LULC data obtained from high-resolution imagery. This paper shares the study on LULC in the region in and around Itanagar, capital of Arunachal Pradesh, a Northeastern India state.

The area chosen for study is in Arunachal Pradesh's Papum Pare district with demography of 51 persons per square kilometre against 17 persons for each square kilometre as per Census of India, 2011. The region of study is located in between the latitudes 27°15'10''N to 26°58'55''N and longitudes 93°29'15''E to 93°54'10''E. The chosen study area consists of five administrative units each called as a circle. These five administrative circles are Itanagar circle, Naharlagun circle, Doimukh circle, Gumto circle and Banderdewa circle. The region has become one of the major evolving regions among the Northeastern states of India. This necessitates a comprehensive and strategic utilization of the resources of the region for sustainable development. In this regard, this research task has been initiated up to develop LULC information in possible detailed manner and present the information concisely in the form of a map.

This may help various stakeholders comprising of government authorities, local communities, business establishments, and environmental organizations to take critical decisions among diverse competing and non-conflicting objectives, and hence, help in achieving sustainable development with minimum environmental degradation. The area of interest is highlighted in Figure 1.

The next section describes the methodologies used in this study, followed by a discussion of the results obtained. Then, the last section draws the conclusion of the study.

2. Materials and Methods

To achieve the aforementioned objective of the current investigation, the IRS multispectral LISS-IV image of 2022 is purchased from the National Remote Sensing Centre, India. Apart from this image, other ancillary data is used to assist the study. Remote sensing data forms the vital backbone to achieve the objective of developing a LULC thematic map in this work. The IRS LISS-IV multispectral image needs pre-processing before carrying out LULC analysis. The band composite and raster clipping are performed as pre-processing requisites. These pre-processing operations are worked on ArcGIS. The ancillary data is used to generate the study area's shape file.

The generated shape files clip the study area, defined as administrative circles, from the pre-processed dataset. The clipping produces the False Color Composite (FCC) image for each administrative circle. The training set for each LULC class is created manually using the polygon tool available in the SCP plugin in QGIS. Further, for each of the LULC classes, sub-classes are created to represent the variation within a LULC class. Using this training set, the FCC image is classified using the random forest classifier available in the SCP plugin in QGIS. This yields the classified LULC thematic image. After performing random forest classification, the accuracy of the classified image is assessed using Google Earth through the HCMGIS plugin available in QGIS. The accuracy assessment produces the kappa coefficient, overall accuracy, and the user's and producer's accuracy in the confusion matrix. Further, the post-classification processing is performed in ArcGIS to produce the LULC thematic map. A brief description of the random forest classifier and accuracy assessment adopted in the study is presented in the following paragraphs. Table 1 shows the information on the LISS-IV image used in the research.

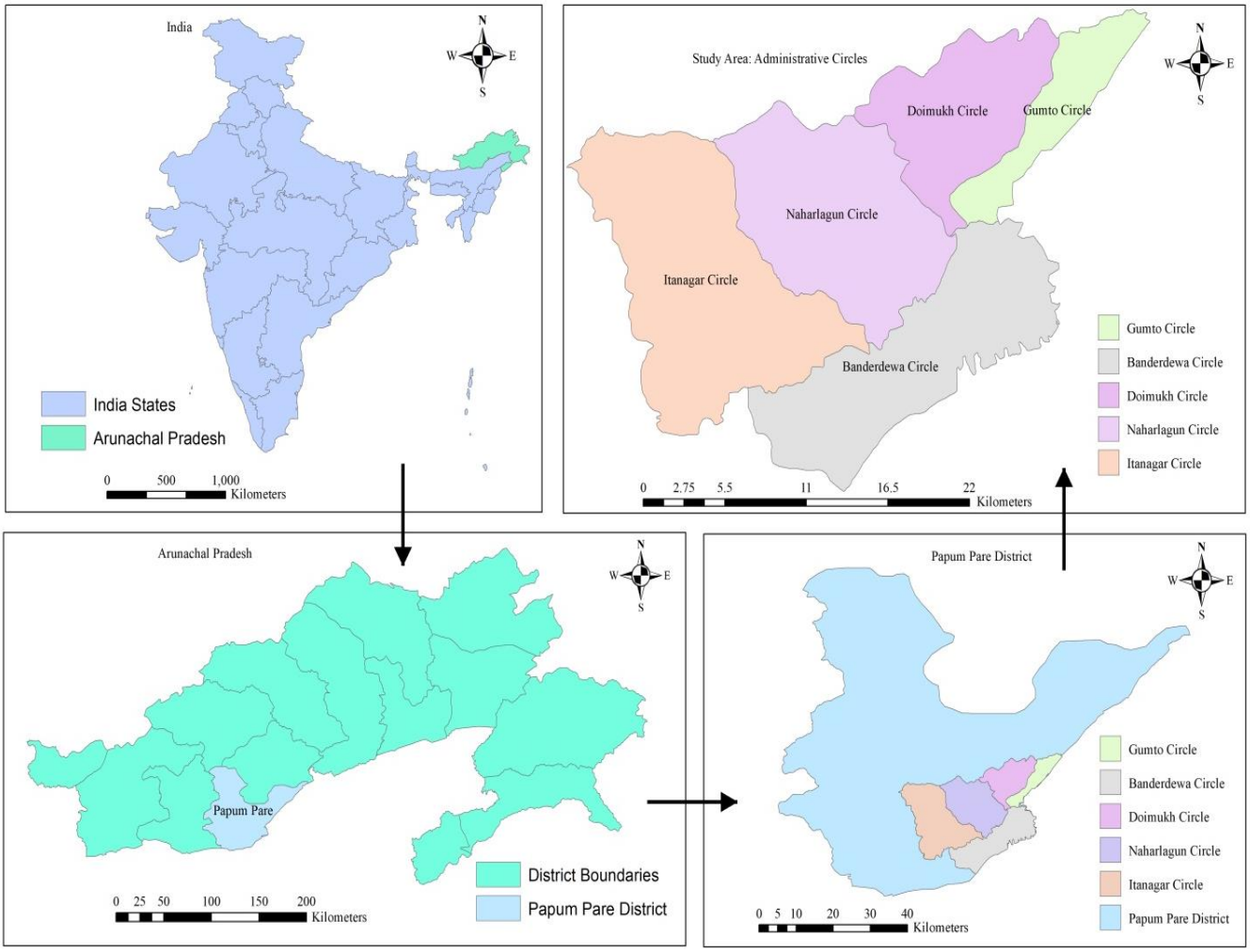


Fig. 1 Study area

Table 1. IRS LISS-IV image specifics

Satellite	Sensor	Bands	Wavelength (μm)	Spatial Resolution	Path/Row	Date of pass	Map Projection/ Ellipsoid/Datum
IRS-R2	L4FX	Green (B2) Red (B3) NIR (B4)	0.52–0.59 0.62–0.68 0.77–0.86	5 m	112/52	8th Jan,2022	UTM/WGS_84/ WGS84

2.1. Random Forest Classifier

Random Forest is a popular and adaptable machine learning algorithm within ensemble learning techniques. This technique was developed by Breiman in 2001 [49], and the technique of random selection of features was developed by [50-53]. Literature on the successful application of random forest in remote sensing image classification is found abundant [33, 54-57]. The technique has many advantages, such as ease of use, speed and accuracy, and has found widespread applications in many disciplines [53]. Details of random forest for further insight may refer to [31]. Keeping this in view, this technique has been adopted in the present study.

2.2. Accuracy Assessment

Accuracy assessment evaluates the acceptance of classified images produced by the image classification algorithm. This assessment is done through the analysis of the confusion matrix. It is a contingency table comprising classified classes against the reference dataset. This matrix quantifies accuracy in terms of kappa coefficient, overall accuracy, user's accuracy and producer's accuracy.

In order to have an overall idea of the size of reference data for use in the validation of the classified image, the reference data size is calculated using a stratified random sampling technique [58, 59] as

$$N = \left(\frac{\sum W_i S_i}{S\left(\hat{O}\right)} \right)^2 \quad (1)$$

Where W_i is the class I mapped area proportion and $S\left(\hat{O}\right)$ is the estimated overall accuracy's standard error that would like to be achieved.

S_o, standard deviation of the stratum I estimated as

$$S_i = \sqrt{U_i\left(1 - U_i\right)} \quad (2)$$

Where U_i is the user's accuracy of class i.

The $S(\hat{O})$ Value is assumed to be 0.01, as in [59]. User's accuracy for each class throughout the classification is kept at 0.90 in the present study with the perception of achieving higher accuracy. After the sample size is determined, which is equivalent to the aggregate of pixels in the classified image, it is allocated equally to the different classes present in the classified image. As the higher resolution image for the study area is unavailable to the authors, Google Earth is considered a reference for assessment. The user's accuracy U_i, the producer's accuracy P_i, and overall accuracy O_i can be estimated using Equation (3), Equation (4), and Equation (5), respectively. Another metric that checks accuracy is the kappa coefficient, estimated from the confusion matrix using Equation (6).

$$U_i = \frac{correctly\ identified\ pixel\ of\ the\ class\ i}{total\ number\ of\ pixel\ of\ class\ i\ in\ the\ classified\ map} \quad (3)$$

$$P_i = \frac{correctly\ identified\ pixel\ of\ the\ class\ i}{total\ number\ of\ pixel\ of\ class\ i\ in\ the\ reference\ map} \quad (4)$$

$$O_i = \frac{number\ of\ pixels\ correctly\ classified\ in\ the\ reference\ map}{total\ number\ of\ reference\ pixel} \quad (5)$$

$$k = \frac{(total\ no.of\ correctly\ classified\ pixels)\times(sample\ size) - sum\ of\ the\ product\ of\ U_i\ and\ P_i}{square\ of\ total\ sample\ size - sum\ of\ the\ product\ of\ Ui\ and\ Pi} \quad (6)$$

Table 2. Hypothetical error matrix of three classes

		Reference				
		Class1	**Class2**	**Class3**	**Row Total**	**U_i**
Map	**Class1**	p11	p12	p13	pm1	U_1
	Class2	p21	p22	p23	pm2	U_2
	Class3	p31	p32	p33	pm3	U_3
	Column Total	pr1	pr2	pr3	Total sample size, N	
	P_i	P_1	P_2	P_3		

Let us consider a hypothetical confusion matrix, as shown in Table 2. Referring to Table 2, the user's accuracy U_i, producer's accuracy P_i, overall accuracy O, and the kappa coefficient k are estimated as follows:

$$U_i = \frac{p_{ii}}{\sum_{j=i}^{j=n} p_{ij}} \quad (7a)$$

$$\text{Then, } U_1 = \frac{p11}{pm1} \quad (7b)$$

and so on for U_2 and U_3.

$$P_i = \frac{p_{ii}}{\sum_{i=1}^{i=n} p_{ij,j=i}} \quad (8a)$$

$$\text{Then, } P_1 = \frac{p11}{pr1} \quad (8b)$$

and so on for P_2 and P_3.

$$O = \frac{\sum_{i=1}^{i=n} p_{ij.i=j}}{\sum_{i=1}^{i=n} p_{ij}} \quad (9a)$$

$$O = \frac{\sum_{i=1}^{i=n} p_{ij,i=j}}{N} = \frac{p11+p22+p33}{N} \quad (9b)$$

$$k = \frac{N\sum_{i=1}^{i=n} p_{ij,i=j} - \sum_{i=1}^{i=n} U_i P_i}{N^2 - \sum_{i=1}^{i=n} U_i P_i} \quad (10a)$$

$$k = \frac{N\times(p11+p22+p33) - (U_1\times P_1 + U_2\times P_2 + U_3\times P_3)}{N^2 - (U_1\times P_1 + U_2\times P_2 + U_3\times P_3)} \quad (10b)$$

The flowchart of the methodology adopted is highlighted in Figure 2.

3. Results and Discussion

Following the outlined steps in the methodology, the LULC thematic map is obtained for the study area. As stated in one of the previous sections, the study area consists of five administrative circles-Itanagar Circle, Naharlagun Circle, Doimukh Circle, Gumto Circle and Banderdewa Circle under Papum Pare district, Arunachal Pradesh, India, the false colour composite images of these administrative circles is presented in Figure 3. After a reconnaissance survey of the motorable area within the research region of five administrative circles, five different LULC classes are identified. These five classes consist of Builtup, Dense Forest, Degraded Forest, Waterbody and the Fallow Land. The definition of each LULC class that appeared as a theme in the output classified image is presented in Table 3. The classified images of the area under research, having different classes as themes, are shown in Figure 4. Among all these five administrative circles, Itanagar Circle has the most extensive area coverage of 179.66 sq. km. Gumto has the smallest area coverage of 45.18 sq. km. Doimukh, Banderdewa, and Naharlagun circles have an area of 84.45, 133.30 and 135.17 sq.km., respectively. Each administrative circle consists of all the LULC classes identified in this study. These classes have different proportions of area coverage in each of the circles. The overall coverage of each LULC class in different circles is tabulated in Table 4.

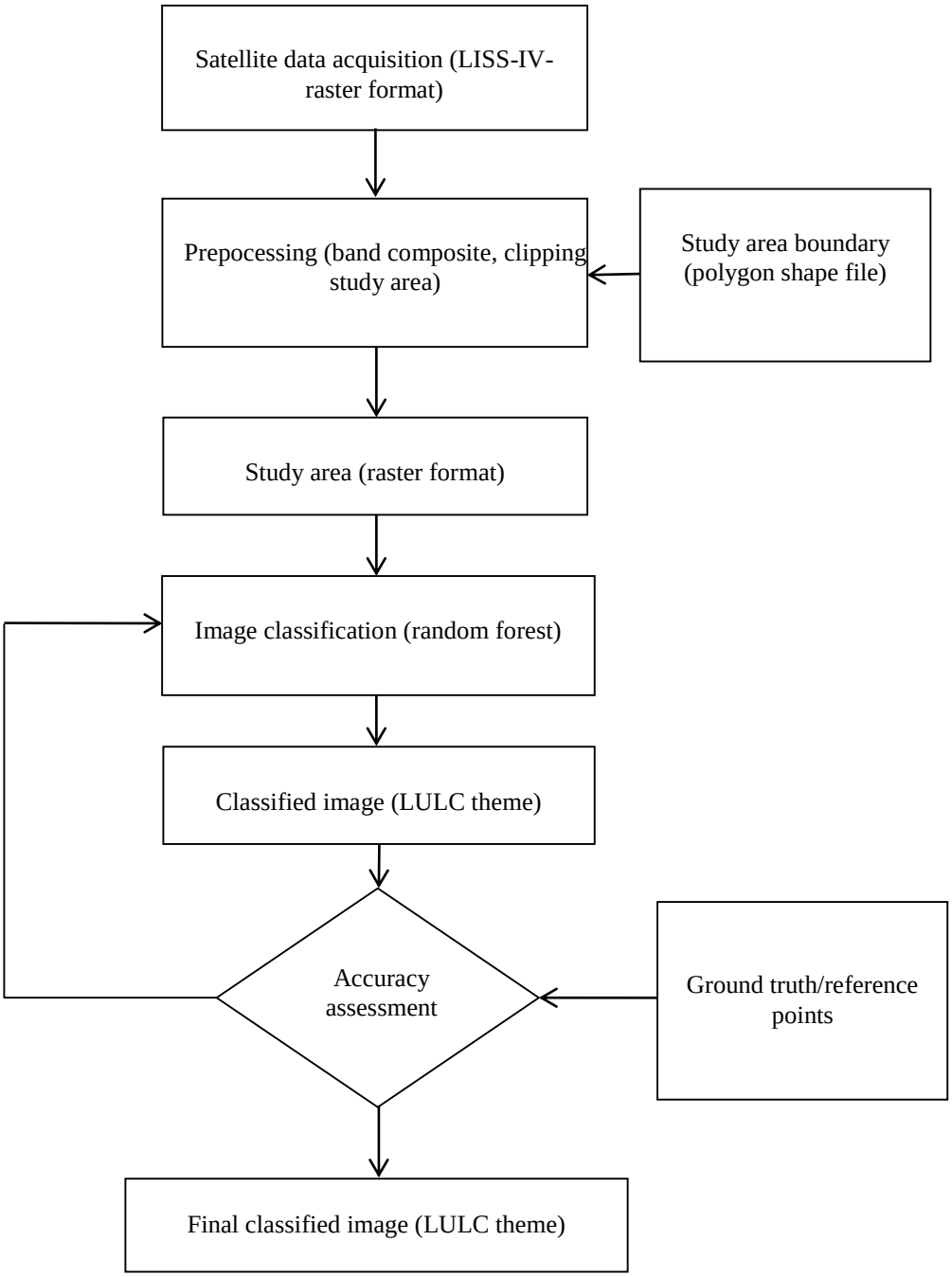
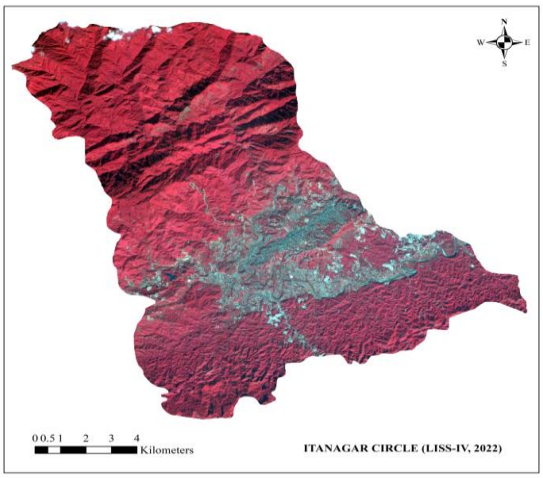


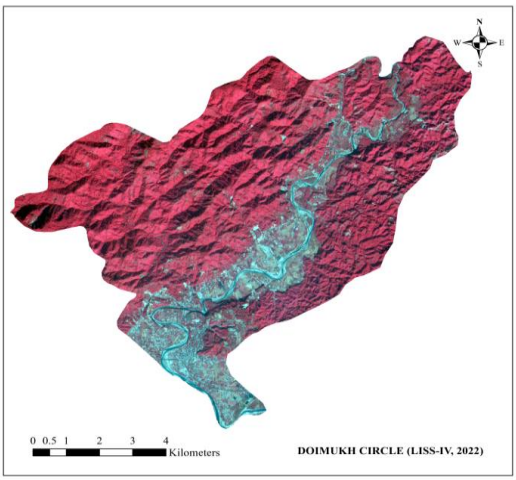
Fig. 2 Methodology flowchart

Table 3. Definition of LULC class adopted

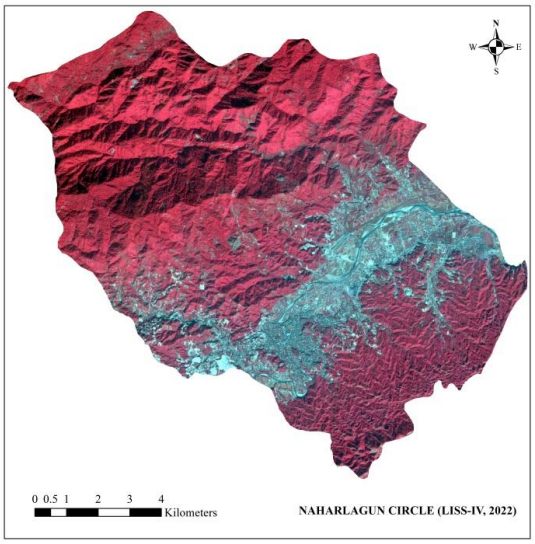
LULC class	Definition
Builtup	Land portion used for construction for human settlement or industrial or any other recreational activities
Dense Forest	The land portion where trees/plantations are grown densely
Degraded Forest	The land portion where trees/plantations are less or have been cut down
Waterbody	The land portion which retains water, such as ponds, lakes and rivers
Fallow Land	The land portion that is used for certain agricultural practices but, at times, left unused or large open ground



(a)

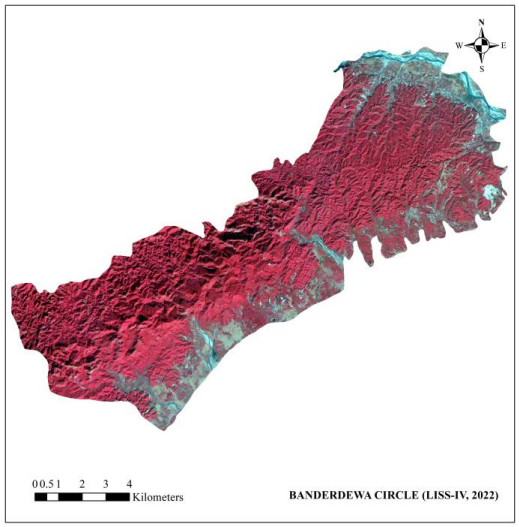


(c)



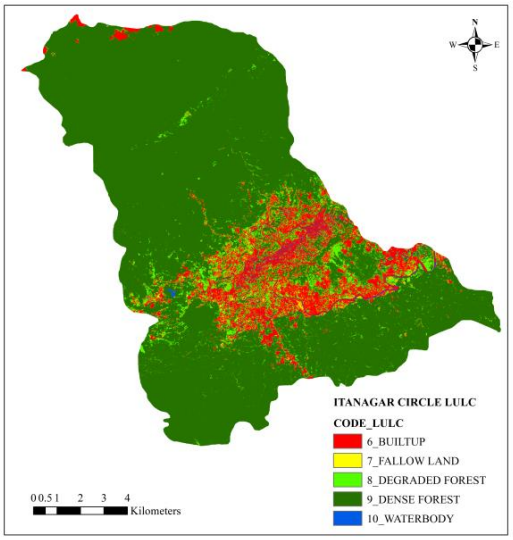
(b)

(d)

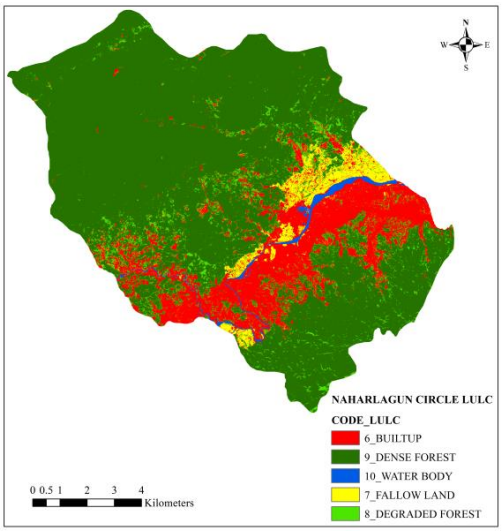


(e)

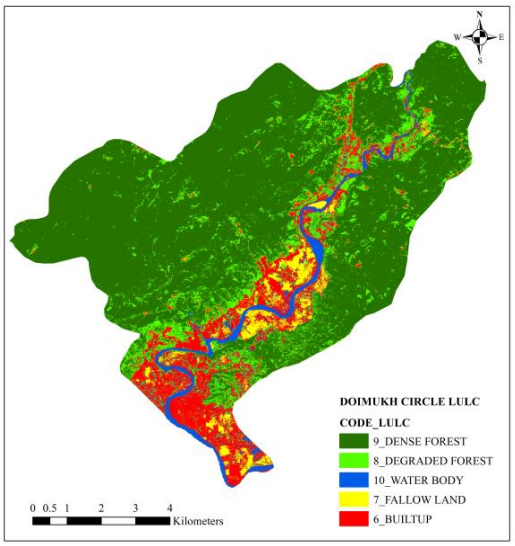
Fig. 3 The FCC images of LISS-IV (2022) for the study area: (a) Itanagar circle, (b) Naharlagun circle, (c) Doimukh circle, (d) Gumto circle, and (e) Banderdewa circle.



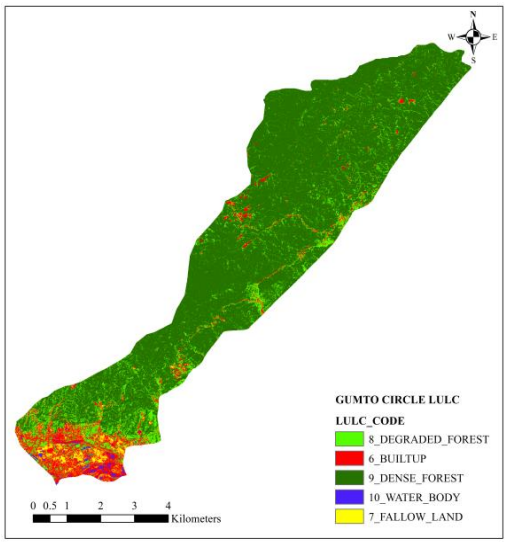
(a)



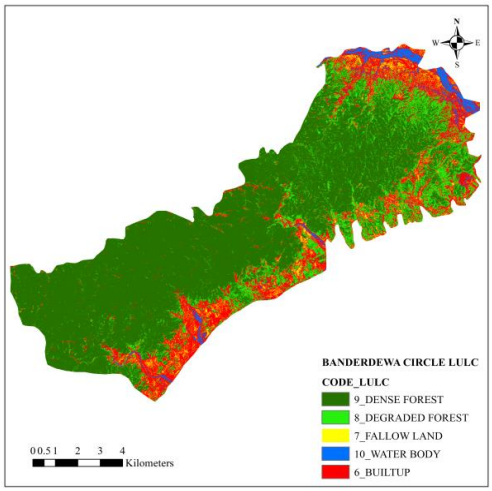
(b)



(c)



(d)



(e)

Fig. 4 Classified images of the area of study: (a) Itanagar circle, (b) Naharlagun circle, (c) Doimukh circle, (d) Gumto circle, and (e) Banderdewa circle.

In all the administrative circles composing the research region, the dense forest dominates area coverage among all the classes and the water body the least, except in Doimukh, where the fallow land has less area coverage than the water body. Further, the built-up, degraded forest and fallow land follow successively in order of area coverage in Itanagar, Naharlagun and Banderdewa circles. However, a contrasting trend emerges in the Doimukh and Gumto circles, where the degraded forest class outnumbers the built-up class in area coverage. The fallow land and water body classifications cover a much smaller percentage of the area than other classes. During the reconnaissance survey, it was also observed that agricultural land classified under the fallow land class was lessseen. This explains the reduced area coverage of the class fallow land in the classification. The fact that the dense forest area coverage is the largest can also be conceived by the visual inspection of raw or unclassified images of the area under research. The area extent percentage of individual LULC classes in different administrative circles under study is tabulated in Table 5. Figure 5 shows the plot of different LULC class coverage in different administrative circles of the area under study. The accuracy assessment of the classified image is conducted using the number of pixels calculated using Equation 3 provided in the methodology section. Table 6 shows the calculation involved in determining a probable sample size to be used in the accuracy assessment for the Gumto circle. The sample size is calculated to be 900, assuming the values assumed in the methodology section. Similarly, the sample size for all the circles is computed to be 900. These reference pixels are distributed equally to all the classes in the classified images, as mentioned in the methodology description. These pixels are checked to see if they are correctly assigned to the class as they are in the field. The confusion matrices for each administrative circle in the study region are presented in Tables 7(a) through 7(e).

Table 4. Area coverage for LULC of each class in the study area

Administrative Circle	LULC coverage (in sq.km)					
	Water Body	Fallow Land	Builtup	Degraded Forest	Dense Forest	Total
Itanagar	1.73	2.98	18.75	11.72	144.48	179.66
Naharlagun	1.39	4.90	23.50	8.54	96.84	135.17
Doimukh	3.55	3.18	9.35	10.48	57.89	84.45
Gumto	0.28	0.83	3.36	5.45	35.27	45.18
Banderdewa	2.95	4.04	22.23	19.14	84.93	133.30

Table 5. Percentage of LULC area coverage of each class in the study area

Administrative Circle	LULC class				
	Water Body	Fallow Land	Builtup	Degraded Forest	Dense Forest
Itanagar	0.96	1.66	10.11	6.52	80.42
Naharlagun	1.03	3.63	17.39	6.32	71.64
Doimukh	4.20	3.77	11.07	12.41	68.55
Gumto	0.63	1.83	7.43	12.06	78.05
Banderdewa	2.22	3.03	16.68	14.36	63.72

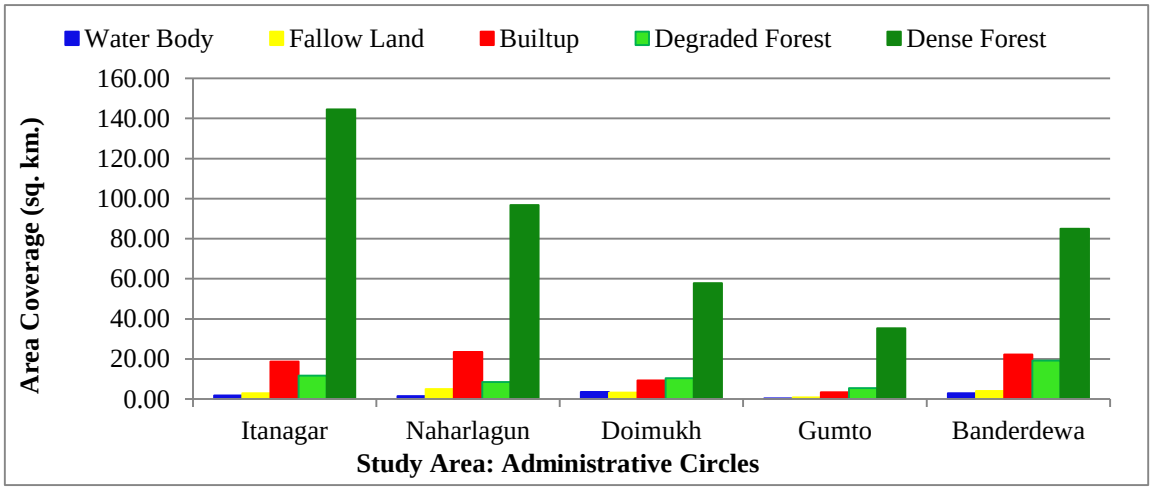


Fig. 5 Plot of different LULC class coverage in different administrative circles

Table 6. Calculation table for sample size for Gumto circle

LULC class	No. of pixel	Pixel size (mm)	Area (m^2)	W_i	U_i	S_i	W_{ii}	Allocation
Degraded Forest	217995	5x5	5449875	0.1206	0.9	0.3	0.0362	180
Builtup	134277	5x5	3356925	0.0743	0.9	0.3	0.0223	180
Fallow Land	33087	5x5	827175	0.0183	0.9	0.3	0.0055	180
Dense Forest	1410625	5x5	35265625	0.7805	0.9	0.3	0.2341	180
Water Body	11353	5x5	283825	0.0063	0.9	0.3	0.0019	180

Table 7(a). Classification Accuracy Assessment Confusion matrix of Itanagar circle (pixel counts)

LULC class	Builtup	Fallow Land	Degraded Forest	Dense Forest	Water Body	Total	User Accuracy	Kappa
Builtup	160	3	2	8	7	180	0.89	-
Fallow Land	15	144	9	2	10	180	0.80	-
Degraded Forest	10	23	143	3	1	180	0.79	-
Dense Forest	0	0	10	170	0	180	0.94	-
Water Body	26	3	4	0	147	180	0.82	-
Total	211	173	168	183	165	900	-	-
Producer Accuracy	0.76	0.83	0.85	0.93	0.89	-	0.85	-
Kappa coefficient	-	-	-	-	-	-	-	0.81

Table 7(b). Classification Accuracy Assessment Confusion matrix of Naharlagun circle (pixel counts)

LULC class	Builtup	Fallow Land	Degraded Forest	Dense Forest	Water Body	Total	User Accuracy	Kappa
Builtup	32	22	13	3	10	180	0.73	-
Fallow Land	28	129	4	3	16	180	0.72	-
Degraded Forest	4	5	156	10	5	180	0.87	-
Dense Forest	0	0	18	162	0	180	0.90	-
Water Body	1	10	0	0	169	180	0.94	-
Total	165	166	191	178	200	900	-	-
Producer Accuracy	0.80	0.78	0.82	0.91	0.85	-	0.83	-
Kappa coefficient	-	-	-	-	-	-	-	0.79

Table 7(c). Classification Accuracy Assessment Confusion matrix of Doimukh circle (pixel counts)

LULC class	Builtup	Fallow Land	Degraded Forest	Dense Forest	Water Body	Total	User Accuracy	Kappa
Builtup	128	32	10	0	10	180	0.71	-
Fallow Land	12	161	4	0	3	180	0.89	-
Degraded Forest	5	27	127	20	1	180	0.71	-
Dense Forest	0	0	8	172	0	180	0.96	-
Water Body	29	23	12	1	115	180	0.64	-
Total	174	243	161	193	129	900	-	-
Producer Accuracy	0.74	0.66	0.79	0.89	0.89	-	0.78	-
Kappa coefficient	-	-	-	-	-	-	-	0.73

Table 7(d). Classification Accuracy Assessment Confusion matrix of Gumto circle (pixel counts)

LULC class	Builtup	Fallow Land	Degraded Forest	Dense Forest	Water Body	Total	User Accuracy	Kappa
Builtup	147	14	0	6	13	180	0.82	-
Fallow Land	14	154	6	1	5	180	0.86	-
Degraded Forest	11	13	152	1	3	180	0.84	-
Dense Forest	1	0	8	170	1	180	0.94	-
Water Body	21	10	2	1	146	180	0.81	-
Total	194	191	168	179	168	900	-	-
Producer Accuracy	0.76	0.81	0.90	0.95	0.87	-	0.85	-
Kappa coefficient	-	-	-	-	-	-	-	0.82

Table 7(e). Classification Accuracy Assessment Confusion matrix of Banderdewa circle (pixel counts)

LULC class	Builtup	Fallow Land	Degraded Forest	Dense Forest	Water Body	Total	User Accuracy	Kappa
Builtup	115	40	10	3	12	180	0.64	-
Fallow Land	10	158	7	0	5	180	0.88	-
Degraded Forest	8	15	146	11	0	180	0.81	-
Dense Forest	3	0	8	168	1	180	0.93	-
Water Body	45	8	1	0	126	180	0.70	-
Total	181	221	172	182	144	900	-	-
Producer Accuracy	0.64	0.71	0.85	0.92	0.88	-	0.79	-
Kappa coefficient	-	-	-	-	-	-	-	0.74

Table 8. Producer accuracy (%) for each class in the study circle

Study circle / LULC class	Itanagar	Naharlagun	Doimukh	Gumto	Banderdewa
Dense Forest	93	91	89	95	92
Degraded Forest	85	82	79	90	85
Fallow Land	83	78	66	81	71
Builtup	76	80	74	76	64
Water Body	89	85	89	81	88

Table 9. Overall accuracy and kappa coefficients of the study circle

Administrative circle	Overall accuracy (%)	Kappa coefficient
Itanagar	85	0.81
Naharlagun	83	0.79
Doimukh	78	0.73
Gumto	85	0.82
Banderdewa	79	0.74

The Itanagar circle has achieved its classification user accuracy of 94% in the dense forest, the degraded forest has the lowest user accuracy of 79%, and the built-up, fallow land and water body achieved a user accuracy of 89%, 80% and 82%, respectively. In the case of producer accuracy, this circle has achieved 93% by dense forest class and the least of 79% by the built-up class, further 83%, 85% and 89% by fallow land, degraded forest and water body, respectively. The overall accuracy of 85% with a 0.81 kappa coefficient is achieved for this circle. In Naharlagun circle, dense forest has achieved the highest user accuracy of 94%, followed by builtup, water body, fallow land and degraded forest, with user

accuracy of 89%, 82%, 80% and 79%, respectively. For producer accuracy, dense forest, water body, degraded forest, fallow land and built-up have 93%, 89%, 85%, 83% and 76%, respectively, with 85% overall accuracy along with a 0.81 kappa coefficient. For the Doimukh and Gumto study circles, the dense forest has user accuracy of 96% and 94%, respectively, being the highest accuracy of the class in the respective circle. The degraded forest and the built-up classes in Doimukh circle have the same user accuracy of 71%, the fallow land of 89%, and the water body class in Doimukh and the builtup class in Banderdewa have the lowest user accuracy of all the classes with 64%. Further, the Gumto circle has achieved its classification with user accuracies of 86%, 81%, 84% and 82% for fallow land, degraded forest, and built and water bodies, respectively. The Doimukh circle has achieved an overall accuracy of 78%, and the Gumto circle has 85% user accuracy, while their respective kappa coefficients are 0.73 and 0.82. Finally, the Banderdewa research circle has the highest user accuracy in dense forests with 93%, followed by fallow land, degraded forest and water bodies with 88%, 81% and 70%. This circle could achieve 79% overall classification accuracy and 0.74 kappa coefficient. The producer accuracy for each class is tabulated in Table 8, while Table 9 gives the overall accuracy and kappa coefficient information for each circle under study.

4. Conclusion

The study aims to map LULC in the region in and around Itanagar, Arunachal Pradesh, India, using IRS LISS-IV of 2022 with a simple approach, adopting the random forest classifier available in QGIS plugin SCSv7 and accuracy assessment in ArcGIS. The classification identifies five thematic classes and attained satisfactory results, which will facilitate the development of a database for the study area. The classification has resulted in a significantly.

A high accuracy percentage in the dense forest class may be why the forest area coverage is large in ground, and the reflectance energy is distinctive from other classes. However, it is also observed that certain pixels representing the water body, specifically the streams of smaller size uphill, are not correctly represented in the raw image itself. This may be attributed to the 5 m spatial resolution of the LISS-IV image. These pixels are overshadowed by the dense forest class in the raw image itself and hence contribute to the dense forest class, though the proportion of the area may be small. This relatively insignificant area contribution to dense forest has a major impact on the area coverage of the water body class, which has led to the water body being insignificant or non-existent when they are present on the ground. Such a mismatch in classification may be reduced to a certain extent by using higher spatial resolution with extra cost incurred. As in most image classifications, another significant cause of mismatch between the reference and map pixels is the intermixing of spectral representation of different features on the ground [20]. This study is no exception to this issue and has many of such in the theme classes adopted, which ultimately accounts for lesser accuracy values. Further, the classification inaccuracies are exacerbated by the presence of shadow within the image data, which aligns with many [34, 60] research works. The classification accuracy is assessed through a confusion matrix. A certain number of representative pixels from the classified images is required to assess classification. This quantity is estimated using Equation (3). During the estimation of the number of pixels, which is referred to as sample size, the authors of this work assumed the overall accuracy standard deviation value to be 0.01 and user accuracy for all the classes to be 0.90. A variation in these assumptions may also lead to different reasonable sample sizes and, hence, different classification accuracies. Further, the authors felt that the stratified random sampling method could still justify the accuracy assessment from the available literature. In addition, equal allocation of the sample is chosen with the expectation that it would give a better or unbiased sample distribution among the different LULC classes, which have significant differences in area coverage within the study region. The prospect of the study is optimistic, as it adopts a more advanced methodological design to determine if there is any increase in the demanding accuracies of the classification. The accuracy may be further improved if the shadows could be removed and substituted by an appropriate reflectance value representing the actual features on the land surface at the time of imaging. Higher-resolution images may also play a crucial role in achieving higher accuracy of classification. The authors asserted that the work has been documented fully and methodically, that the results will be helpful to certain stakeholders, and that they consent to the strengths and limitations of the methodology used in this study.

Acknowledgments

The authors acknowledge Mr. Jei Dignium, NERIST, for sharing the ancillary data.

References

[1] M. Palaniyandi, and V. Nagarathinam, “Land Use/Land Cover Mapping and Change Detection using Space Borne Data,” *Journal of the Indian Society of Remote Sensing*, vol. 25, pp. 27-33, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[2] D.G. Brown, B.C. Pijanowski, and J.D. Duh, “Modeling the Relationships between Land Use and Land Cover on Private Lands in the Upper Midwest, USA,” *Journal of Environmental Management*, vol. 59, no. 4, pp. 247-263, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[3] T.P. Singh et al., “Vegetation Mapping and Characterization in West Siang District of Arunachal Pradesh, India - A Satellite Remote Sensing-based Approach,” *Current Science*, vol. 83, no. 10, pp. 1221-1230, 2002. [Google Scholar] [Publisher Link]

[4] Santosh Babar, and H. Ramesh, "Streamflow Response to Land Use–Land Cover Change over the Nethravathi River Basin, India," *Journal of Hydrologic Engineering*, vol. 20, no. 10, 2015. [CrossRef] [Google Scholar] [Publisher Link]
[5] S. Setti et al., "Attribution of Hydrologic Changes in a Tropical River Basin to Rainfall Variability and Land-Use Change: Case Study from India," *Journal of Hydrologic Engineering*, vol. 25, no. 8, pp. 1-15, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[6] Ji Han et al., "Evaluating Land-Use Change in Rapidly Urbanizing China: Case Study of Shanghai," *Journal of Urban Planning and Development*, vol. 135, no. 4, 2009. [CrossRef] [Google Scholar] [Publisher Link]
[7] Parth Sarathi Roy et al., "Anthropogenic Land Use and Land Cover Changes—A Review on Its Environmental Consequences and Climate Change," *Journal of the Indian Society of Remote Sensing*, vol. 50, pp. 1615-1640, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[8] Jawan Singh Rawat, and Ramesh Chandra Joshi, "Remote-Sensing and GIS-based Landslide-Susceptibility Zonation using the Landslide Index Method in Igo River Basin, Eastern Himalaya, India," *International Journal of Remote Sensing*, vol. 33, no. 12, pp. 3751-3767, 2012. [CrossRef] [Google Scholar] [Publisher Link]
[9] Gopala Areendran et al., "A Geospatial Study to Assess the Land Use Land Cover of Mehao Wildlife Sanctuary in Arunachal Pradesh, India," *Asian Journal of Geographical Research*, vol. 1, no. 2, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[10] Yomto Mayi, Joanica Delicia Jyrwa, and Santanu K Patnaik, "Monitoring of Land Use/ Land Cover Changes of Daying Ering Wildlife Sanctuary, Arunachal Pradesh, India, Using Remote Sensing and Geographic Information System," *Journal of Applied and Natural Science*, vol. 15, no. 2, pp. 713-719, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[11] Zhongqi Wang, Qi Han, and Bauke de Vries, "Land Use/Land Cover and Accessibility: Implications of the Correlations for Land Use and Transport Planning," *Applied Spatial Analysis and Policy*, vol. 12, pp. 923-940, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[12] Amritpal Digra et al., "Temporal Land-Use/Land-Cover Change Analysis in Kotla Sub-Watershed of Rupnagar District (Punjab) Using Remote Sensing and GIS," *Journal of the Indian Society of Remote Sensing*, vol. 50, pp. 1371-1391, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[13] Nova D. Doyog, "Urban Forest Cover and LST Change Monitoring through Optical and Thermal Remote Sensing Approach," *Journal of the Indian Society of Remote Sensing*, vol. 51, pp. 2467-2480, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[14] Heng Sun, Wayne Forsythe, and Nigel Waters, "Modeling Urban Land Use Change and Urban Sprawl: Calgary, Alberta, Canada," *Networks and Spatial Economics*, vol. 7, pp. 353-376, 2007. [CrossRef] [Google Scholar] [Publisher Link]
[15] A.S. Antonarakis, K.S. Richards, and J. Brasington, "Object-based Land Cover Classification using Airborne LiDAR," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 2988-2998, 2008. [CrossRef] [Google Scholar] [Publisher Link]
[16] Dilong Li et al., "AGFP-Net: Attentive Geometric Feature Pyramid Network for Land Cover Classification using Airborne Multispectral LiDAR Data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 108, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[17] Andrew Clark, Stuart Phinn, and Peter Scarth, "Optimised U-Net for Land Use–Land Cover Classification Using Aerial Photography," *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 91, pp. 125-147, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[18] N.C. Gautam, "Aerial Photo-Interpretation Techniques for Classifying Urban Land Use," *Photogrammetric Engineering & Remote Sensing*, vol. 42, no. 6, pp. 815-822, 1976. [Google Scholar] [Publisher Link]
[19] Jonathan R.B. Fisher et al., "Impact of Satellite Imagery Spatial Resolution on Land Use Classification Accuracy and Modeled Water Quality," *Remote Sensing in Ecology and Conservation*, vol. 4, no. 2, pp. 137-149, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[20] Rongjun Qin, and Tao Liu, "A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability," *Remote Sensing*, vol. 14, no. 3, pp. 1-28, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[21] V. Jayaraman et al., "Derivation of Enhanced Radiometric Information from IRS LISS-2 Data," *Acta Astronautica*, vol. 35, no. 7, pp. 483-487, 1995. [CrossRef] [Google Scholar] [Publisher Link]
[22] A.O. Varghese et al., "Habitat Suitability Assessment of Ardeotis Nigriceps (Vigors) in Great Indian Bustard Sanctuary, Maharashtra (India) Using Remote Sensing and GIS," *Journal of the Indian Society of Remote Sensing*, vol. 44, pp. 49-57, 2016. [CrossRef] [Google Scholar] [Publisher Link]
[23] Mingting Zhou et al., "Large-scale Road Extraction from High-Resolution Remote Sensing Images based on a Weakly-Supervised Structural and Orientational Consistency Constraint Network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 193, pp. 234-251, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[24] Hao Wang et al., "Urban Boundary Extraction and Urban Sprawl Measurement Using High-Resolution Remote Sensing Images: A Case Study of China's Provincial," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 1713-1719, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[25] Mengmeng Li, and Alfred Stein, "Mapping Land Use from High Resolution Satellite Images by Exploiting the Spatial Arrangement of Land Cover Objects," *Remote Sensing*, vol. 12, no. 24, pp. 1-21, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] A. Şekertekin et al., "Analysing the Effects of Different Land Cover Types on Land Surface Temperature Using Satellite Data," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, pp. 665-667, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[27] Markand Oza, "Choosing Optimal Spatial Resolution —Study of Two Agriculture Dominated Areas," *Journal of the Indian Society of Remote Sensing*, vol. 21, pp. 179-183, 1993. [CrossRef] [Google Scholar] [Publisher Link]

[28] James Thomas Steven Savage et al., "Quantifying the Importance of Spatial Resolution and other Factors through Global Sensitivity Analysis of a Flood Inundation Model," *Water Resources Research*, vol. 52, no. 11, pp. 9146-9163, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[29] Pooja Sonde, Sanjay Balamwar, and Rohini S. Ochawar, "Urban Sprawl Detection and Analysis Using Unsupervised Classification of High Resolution Image Data of Jawaharlal Nehru Port Trust Area in India," *Remote Sensing Applications: Society and Environment*, vol. 17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[30] Katalin Varga et al., "Improved Land Cover Mapping using Aerial Photographs and Satellite Images," *Open Geosciences*, vol. 7, no. 1, 15-26, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[31] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001. [CrossRef] [Google Scholar] [Publisher Link]

[32] Hamid Ebrahimy et al., "Per-Pixel Land Cover Accuracy Prediction: A Random Forest-based Method with Limited Reference Sample Data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 172, pp. 17-27, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[33] V.F. Rodriguez-Galiano et al., "An Assessment of the Effectiveness of a Random Forest Classifier for Land-cover Classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93-104, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[34] Weiqi Zhou et al., "Object-based Land Cover Classification of Shaded Areas in High Spatial Resolution Imagery of Urban Areas: A Comparison Study," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1769-1777, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[35] S. Rajesh et al., "Land Cover/Land Use Mapping of LISS IV Imagery Using Object-Based Convolutional Neural Network with Deep Features," *Journal of the Indian Society of Remote Sensing*, vol. 48, pp. 145-154, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[36] Foivos I. Diakogiannis et al., "ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94-114, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[37] Victor Alhassan et al., "A Deep Learning Framework for Land-use/Land-Cover Mapping and Analysis using Multispectral Satellite Imagery," *Neural Computing and Applications*, vol. 32, pp. 8529-8544, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[38] Wei Han et al., "A Semi-Supervised Generative Framework with Deep Learning Features for High-Resolution Remote Sensing Image Scene Classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 23-43, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[39] Diksha, and Amit Kumar, "Analysing Urban Sprawl and Land Consumption Patterns in Major Capital Cities in the Himalayan Region using Geoinformatics," *Applied Geography*, vol. 89, pp. 112-123, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[40] Soumik Saha et al., "Modelling and Predicting of Landslide in Western Arunachal Himalaya, India," *Geosystems and Geoenvironment*, vol. 2, no. 2, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[41] Prashant Patil et al., "Monitoring Land Use Land Cover Change and Its Impact on Climatic Parameters Using Remote Sensing and GIS: A Case Study of Lower Dibang Valley, Arunachal Pradesh, India," *Geoinformatica Polonica*, vol. 23, pp. 59-75, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[42] Ranjit Mahato et al., "AHP and GIS-based Delineation of Groundwater Potential of Papum Pare District of Arunachal Pradesh, India," *Journal of the Geological Society of India*, vol. 98, no. 1, pp. 102-112, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[43] Jyotishman Deka et al., "Study on Land-use and Land-cover Change Dynamics in Eastern Arunachal Pradesh, N.E. India using Remote Sensing and GIS," *Tropical Ecology*, vol. 60, pp. 199-208, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[44] Preeti Barsha Borah, "Assessment of Land use Land Cover Change Detection and Future Prediction: A GIS based Study of Pasighat, Arunachal Pradesh, India," *Ecology Environment and Conservation*, vol. 30, no. 3, pp. 1037-1047, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[45] M. Vese et al., "Decadal Variations in Area under Different Soil Erosion Classes using RUSLE and GIS: Case Studies of River Basins from Western and Eastern Arunachal Pradesh," *Journal of the Geological Society of India*, vol. 99, pp. 1725-1737, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[46] Trilochan Singh, and Prakash Tewari, "Impact of Earthquake Disasters on New Urbanization Pattern in North Eastern Region of India," *Journal of Environmental Systems*, vol. 32, no. 1, pp. 69-90, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[47] Ram Krishna Mandal, "Urbanization in Arunachal Pradesh: Trends and Patterns 1961-2000," *Journal of Global Economy*, vol. 5, no. 4, pp. 320-340, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[48] Swapna Acharjee, "Urban Land use and Geohazards in Itanagar, Arunachal Pradesh, India: the Need for Geotechnical Intervention and Geoethical Policies in Urban Disaster Resilience Programmes in a Changing Climate," *Geological Society, London, Special Publications*, vol. 419, no. 1, pp. 63-68, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[49] Leo Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996. [CrossRef] [Google Scholar] [Publisher Link]

[50] Tin Kam Ho, "Random Decision Forests," *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, Canada, pp. 278-282, 1995. [CrossRef] [Google Scholar] [Publisher Link]

[51] Tin Kam Ho, "The Random Subspace Method for Constructing Decision Forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, 1998. [CrossRef] [Google Scholar] [Publisher Link]

[52] Yali Amit, and Donald Geman, "Shape Quantization and Recognition with Randomized Trees," *Neural Computation*, vol. 9, no. 7, pp. 1545-1588, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[53] Khaled Fawagreh, Mohamed Medhat Gaber, and Eyad Elyan, "Random Forests: From Early Developments to Recent Advancements," *Systems Science & Control Engineering*, vol. 2, no. 1, pp. 602-609, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[54] Charlotte Pelletier et al., "Assessing the Robustness of Random Forests to Map Land Cover with High Resolution Satellite Image Time Series Over Large Areas," *Remote Sensing of Environment*, vol. 187, pp. 156-168, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[55] Maruf Billah et al., "Random Forest Classifications for Landuse Mapping to Assess Rapid Flood Damage using Sentinel-1 and Sentinel-2 Data," *Remote Sensing Applications: Society and Environment*, vol. 30, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[56] Di Shi, and Xiaojun Yang, "An Assessment of Algorithmic Parameters Affecting Image Classification Accuracy by Random Forests," *Photogrammetric Engineering & Remote Sensing*, vol. 82, no. 6, pp. 407-417, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[57] Hsiao-Chien Shih et al., "From Land Cover to Land Use: Applying Random Forest Classifier to Landsat Imagery for Urban Land-use Change Mapping," *Geocarto International*, vol. 37, no. 19, pp. 5523-5546, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[58] William Gemmell Cochran, *Sampling Techniques*, 3rd ed., Wiley, pp. 1-428, 1977. [Google Scholar] [Publisher Link]

[59] Pontus Olofsson et al., "Good Practices for Estimating Area and Assessing Accuracy of Land Change," *Remote Sensing of Environment*, vol. 148, pp. 42-57, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[60] Seyed Kazem Alavipanah, "The Shadow Effect on Surface Biophysical Variables Derived from Remote Sensing: A Review," *Land*, vol. 11, no. 11, pp. 1-30, 2022. [CrossRef] [Google Scholar] [Publisher Link]