


# Enhancing Land Use/Land Cover Analysis with Sentinel-2 Bands: Comparative Evaluation of Classification Algorithms and Dimensionality Reduction for Improved Accuracy Assessment

Akil V. Memon<sup>1†</sup>, Nirav V. Shah<sup>1,2</sup>, Yogesh S. Patel<sup>1</sup> and Tarun Parangi<sup>3,4</sup>

<sup>1</sup>Department of Civil Engineering, Sankalchand Patel College of Engineering, Sankalchand Patel University, Visnagar, Gujarat, India

<sup>2</sup>Department of Food Engineering, College of Food Processing Technology and Bio Energy, Anand Agricultural University, Anand, Gujarat, India

<sup>3</sup>Chemistry Department, Parul Institute of Applied Science, Parul University, Vadodara, Gujarat, India

<sup>4</sup>Research & Development Cell (RDC), Parul University, Vadodara, Gujarat, India

†Corresponding author: A. V. Memon: [akilpochi@gmail.com](mailto:akilpochi@gmail.com)

**Abbreviation:** Nat. Env. & Poll. Technol.  
**Website:** [www.neptjournal.com](http://www.neptjournal.com)

*Received:* 02-09-2024

*Revised:* 07-10-2024

*Accepted:* 18-10-2024

## Key Words:

LULC classification  
 Sentinel-2  
 Accuracy evaluation  
 GIS analysis  
 Kappa coefficient

## Citation for the Paper:

Memon, A. V., Shah, N. V., Patel, Y. S. and Parangi, T. 2025. Enhancing land use/land cover analysis with sentinel-2 bands: comparative evaluation of classification algorithms and dimensionality reduction for improved accuracy assessment. *Nature Environment and Pollution Technology*, 24(2), p. B4264. <https://doi.org/10.46488/NEPT.2025.v24i02.B4264>

*Note: From year 2025, the journal uses Article ID instead of page numbers in citation of the published articles.*



**Copyright:** © 2025 by the authors

**Licensee:** Technoscience Publications

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## ABSTRACT

Accurately classifying land use and land cover (LULC) is crucial for understanding Earth's dynamics under human influence. This study proposes a novel approach to assess LULC classification accuracy using Sentinel-2 data. Authors have compared traditional and Principal Component Analysis (PCA)-based approaches for Maximum Likelihood Classification, Random Forest, and Support Vector Machine (SVM) classifiers. Four key classes (agricultural land, water bodies, built-up areas, and wastelands) are classified using supervised learning. Accuracy is evaluated using producer, user, overall accuracy, and kappa coefficient. Our findings reveal superior accuracy with PCA-SVM compared to other methods. PCA effectively reduces data redundancy, extracting essential spectral information. This study highlights the value of combining PCA with SVM for LULC classification, empowering policymakers with enhanced decision-making tools and fostering informed policy development.

## INTRODUCTION

The alteration of land use and land cover (LULC) has emerged as a pivotal element in contemporary approaches to the stewardship of natural resources and the surveillance of environmental transformations. The terms 'Land Use' and 'Land Cover,' initially identified as distinct, have been observed to be used interchangeably across diverse literature (Shrestha et al. 2021). Remote sensing data obtained from satellites are widely employed in the delineation of the Earth's LULC. The global repercussions of changes in LULC are evident, particularly in the contrasting impacts on urban and rural regions. Mapping LULC stands out as a crucial application of remote sensing (Lakhera & Rahi 2021, Tiwari et al. 2024). Land cover serves as a foundational factor influencing and connecting various aspects of both the human and physical environment. It is widely acknowledged that alterations in land cover have substantial implications for essential processes, such as biogeochemical cycling, consequently affecting global warming, soil erosion, and sustainable land use. Over the next century, land cover is anticipated to be the foremost influential variable impacting biodiversity (Cheruto et al. 2016). Remote sensing technologies offer a unique advantage in this context, allowing for repetitive, long-term observations of the same geographic regions. The ability to monitor changes over time provides critical insights into the dynamics of land cover transitions, enabling more accurate predictions of future environmental shifts.

Platforms like Google Earth and the Earth Observation (EO) satellites have revolutionized the way researchers access and analyze spatial data, making it easier to conduct LULC assessments even in remote or poorly monitored regions. This is particularly vital for developing countries, where other forms of high-resolution data might be unavailable due to resource constraints.

Remote sensing fills this gap, providing comprehensive data that can be used for effective land management, urban planning, and environmental conservation (Tilahun & Teferie et al. 2015, Fakeye et al. 2015). Moreover, the rapid development of machine learning (ML) (Loukika et al. 2021) and Deep Learning (DL) techniques has transformed LULC classification. Traditionally, methods like the Bayesian Maximum Likelihood classifier were used, but recent advancements in ML have led to the development of more efficient and accurate classification algorithms. Support Vector Machines (SVM), Random Forest (RF), K-nearest neighbors (KNN), and more recently, deep learning models have enabled researchers to analyze LULC data with unprecedented accuracy (Asif et al. 2023, Avci et al. 2023). These techniques are particularly beneficial in handling the vast amounts of complex data generated by hyperspectral imagery (HSI), which contains hundreds of spectral bands for each pixel. Hyperspectral data provide detailed information about the composition of land surfaces, making them ideal for fine-scale LULC classification. However, this data richness also presents challenges, such as the high dimensionality and computational demands of processing HSI.

Deep learning models, including Convolutional Neural Networks (CNNs), have proven particularly adept at handling hyperspectral data, outperforming traditional methods by capturing intricate spatial and spectral features that other algorithms may miss. The ability of these models to automatically learn feature representations from raw data has significantly improved the performance of LULC classification, leading to more accurate and reliable maps (Tao et al. 2023). This is especially important in heterogeneous landscapes—areas where diverse land use patterns, such as mixed urban and agricultural zones, create complex decision boundaries that can be challenging for conventional methods to classify correctly.

### Predictive Modeling Approaches

Machine learning classifiers are noted for achieving increased accuracy, even when dealing with intricate data and a higher number of input features (Parracciani et al. 2024, Huang et al. 2011). Some well-known classifiers include Artificial Neural Networks (ANN), CART, k-nearest Neighbor (k-NN), RF, and SVM (Jayabaskaran & Das 2023). While

certain classifiers, like ANN, adhere to a neural network structure with multiple layers of nodes that exchange input observations iteratively throughout the learning process (specifically, the Multi-Layer Perceptron), reaching a termination condition, CART constructs a straightforward decision tree based on the provided training data (Sun et al. 2024). RF, on the other hand, employs random subsets of training data to create numerous decision trees (Chowdhury 2024). Other classifiers, such as k-NN, utilize information about neighboring pixels to discern the inherent patterns within the training dataset (Van Groenigen & Stein 1998). In contrast, classifiers like SVM identify a subset of training data known as support vectors by fitting a hyperplane that optimally separates two classes. Across various literature, it is widely suggested that in most classification scenarios, RF and SVM stand out as superior performers compared to other machine classifiers (Huang et al. 2002, Mountrakis et al. 2011, Pal & Foody 2012, Belgiu & Dragut 2016).

Random Forest tree employs a bagging technique, randomly selecting a subset of training samples with replacements to build individual trees. This can lead to overlapping samples and some being excluded from certain trees (Kunapuli 2023, Siqueira et al. 2024). The unused samples (out-of-bag samples) are utilized for unbiased performance evaluation, providing an estimate of generalization error (Blain n.d). Additionally, at each node, Random Forest randomly selects variables to determine the best split, reducing the correlation between trees and lowering generalization error. The choice of pruning methods typically affects tree-based classifiers, but Random Forest is resilient to such influences, as it constructs trees without the need for pruning techniques (Breiman et al. 2001, 2004).

The Maximum Likelihood Classification (MLC) assumes a normal distribution of statistics for each class in every band. It computes the likelihood that a particular pixel belongs to a specific class. Unless a probability threshold is chosen, all pixels receive classification. Each pixel is allocated to the class with the highest probability, i.e., the maximum likelihood. If the maximum probability is below a specified threshold, the pixel remains unclassified (Richards et al. 2013).

Support vector machine (SVM) is a supervised machine learning method that is often used in LULC classification (Halder et al. 2023). SVM demonstrates effective accuracy in LULC applications, creating a hyperplane in high-resolution satellite imagery. Notably, it excels in classifying images with a constrained set of training samples. SVM is regarded as more sophisticated than maximum likelihood classification (MLC) and is capable of achieving superior LULC classification compared to other classifiers, especially

when dealing with a limited number of pixels (Fetene et al. 2023). SVM aims to discover the optimal hyperplane that maximizes the margin between different classes of data points.

Principal component analysis (PCA) utilizing satellite imagery has been widely employed across various domains, notably in the detection of changes in land use and land cover (Moharram & Sundaram 2023). Over the years, it has gained considerable popularity due to its simplicity and effectiveness in amplifying change-related information (Schirpke et al. 2023). PCA, rooted in eigenvector analysis of the data correlation matrix, aims to capture maximum variances within a limited number of orthogonal components (Mahmud & Hafsa 2016, Shekar & Mathew 2022).

The fundamental concept of PCA involves the reduction of dimensionality in a dataset comprising numerous interrelated variables. This reduction is typically achieved by transforming the dataset into a new set of variables known as principal components (PCs). These PCs are both uncorrelated and ordered. When applied to data from multiple spectral bands, PCA tends to concentrate the majority of information in the initial two or three PCs, while the subsequent PCs generally contain noise (Somayajula et al. 2021, Shekar et al. 2023).

This study makes a unique contribution to the existing literature by conducting a direct comparative analysis between a conventional method utilizing Sentinel-2 original

bands and a PCA-based approach for land use and land cover (LULC) classification. While much of the previous work has focused on evaluating classification algorithms in isolation or utilizing only traditional methods, this study evaluates the same classifiers (Maximum Likelihood Classification, Random Forest, and Support Vector Machine) across two distinct dimensionality reduction approaches. By doing so, the study sheds light on how PCA, a commonly used dimensionality reduction technique, impacts the performance of LULC classification algorithms in real-world applications.

The comparison of PCA and conventional methods is impactful because it addresses a key challenge in remote sensing—the curse of dimensionality—especially when handling multi-spectral data. Reducing dimensionality can lead to more efficient classification while maintaining or even improving accuracy. This study provides new insights into how PCA, when combined with machine learning classifiers like SVM, can outperform traditional classification approaches. This adds to the current understanding of LULC classification by highlighting the effectiveness of PCA-SVM, particularly in improving classification accuracy and computational efficiency.

## Study Area

This study delves into the Panam watershed, a left-bank tributary of the Mahi River basin nestled within Gujarat's Mahisagar district. The Panam River originates near Bhadra

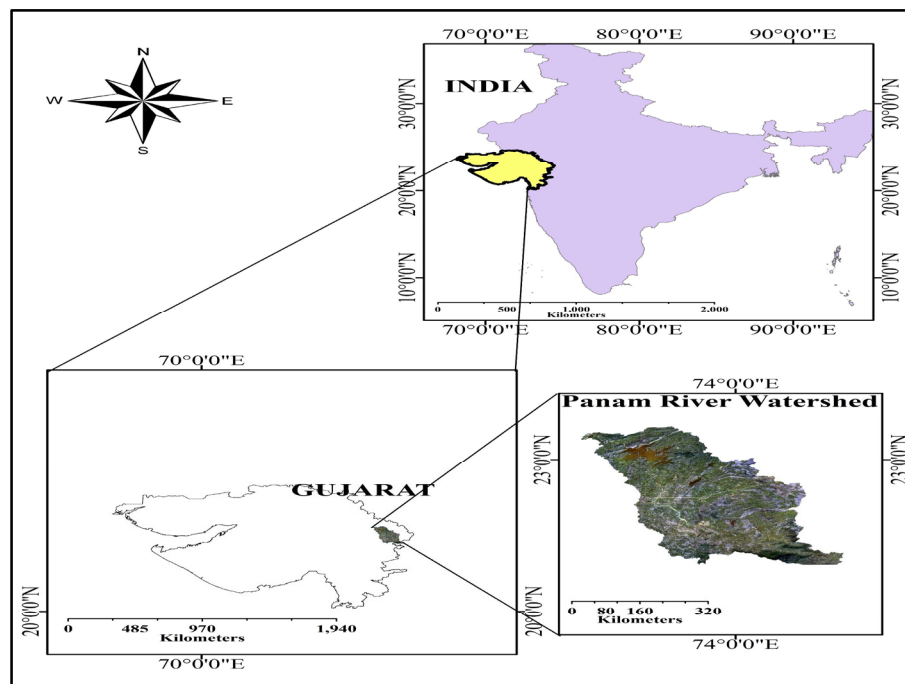


Fig. 1: The study area of the Panam River watershed.

in Madhya Pradesh's Jhabua district, traversing northwest for roughly 125 kilometers before merging with the Mahi in the Panchmahals district of Gujarat state. Encompassing a drainage area of 2400 square kilometers, the region experiences a tropical climate with temperatures fluctuating between 15°C in January and 40°C in May. Rainfall averages 945 mm annually, shaping the watershed's characteristics. This unique confluence of geographical and climatic factors positions the Panam watershed as an ideal canvas for our research endeavors. Fig. 1 provides a visual representation of the study area's location.

## MATERIALS AND METHODS

This research investigates LULC in the Panam River watershed using freely available Sentinel-2 satellite imagery from January 2024, acquired from the Copernicus Open Access Hub with minimal cloud cover. QGIS, free and open-source GIS software, was employed for data visualization, editing, and analysis. The study focused on 12 Sentinel-2 bands (bands 2 to 12) that were mosaicked and clipped to the specific watershed area. Details regarding the utilized Sentinel-2 multi-spectral instrument (MSI) Level 1C bands are provided in Table 1.

As shown in Fig. 2, the methodology commences by importing Sentinel-2-MSI L1 C imagery into the QGIS software, followed by the creation of a seamless image of the study area through mosaicking. A band composite image is then generated using all bands. To streamline data and identify crucial bands for classification, a Principal Component Analysis (PCA) is performed on the band composite image. After data preprocessing, the subsequent step involves collecting training samples for each land use class to be classified. High-resolution Google Earth images

were employed to extract training samples for each LULC class. These images, with their fine spatial detail, allowed for accurate identification of homogeneous areas corresponding to Agricultural land, Water bodies, Built-up areas, and Barren land. Each sample class was carefully delineated through visual interpretation, ensuring that only representative and pure pixels were included.

The size of each sample class was determined based on the area and spatial distribution of the LULC classes, ensuring sufficient representation across the study area. For instance, larger classes like Agricultural land had a higher number of sample pixels compared to smaller classes like Water bodies. On average, around 200–300 pixels were collected per class to train the classifiers. Validation of the sample set was achieved through a stratified random sampling technique, where ground-truth points were cross-referenced with both high-resolution imagery and field data (where available). This process ensured that the samples represented the true variability within each LULC class, leading to robust training datasets for model development.

The classification stage employs three distinct machine learning classifiers: maximum likelihood classifier, random forest tree classifier, and support vector machine classifier. Each classifier undergoes training using the previously collected training samples, distinguishing among the specified land use classes. Various accuracy metrics, including user's accuracy (UA), producer's accuracy (PA), overall accuracy (OA), and Kappa's coefficient ( $k$ ), are calculated for each classifier within the context of the designated land use classes. The kappa coefficient is calculated using the following equation.

$$k = (P_o - P_e) / (1 - P_e) \quad \dots(1)$$

where  $k$  is the kappa coefficient (ranges from 0 to 1),  $P_o$  is the observed agreement probability (sum of diagonal elements of confusion matrix divided by the total number of pairs),  $P_e$  is the expected agreement probability (sum of products of individual agreement probabilities for each category).

A kappa value of 0 indicates absolutely no agreement between raters beyond what could be expected by chance alone. Essentially, their ratings are no better than random guessing. If the kappa falls between 0.01 and 0.20, there's some slight agreement, meaning the raters are occasionally aligned but more often differ. Moving to the 0.21-0.40 range suggests fair agreement. While not perfect, the raters demonstrate some consistency in their assessments. A kappa value between 0.41 and 0.60 signifies moderate agreement, indicating the raters are often in agreement, though occasional discrepancies still exist. Substantial agreement is achieved with a kappa of 0.61 to 0.80. Here,

Table 1: Sentinel-2 MSI Level1 C bands and their bandwidth.

Band No.	Band Name	Central Wavelength (nm)	Bandwidth (nm)
2	Blue	496.6	98.0
3	Green	560.0	45.0
4	Red	664.5	38.0
5	Vegetation Red Edge	705.0	19.0
6	Vegetation Red Edge	740.0	18.0
7	Vegetation Red Edge	783.0	28.0
8	Near Infrared	835.1	145.0
8A	Narrow NIR	865.0	33.0
9	Water vapour	945.0	26.0
10	SWIR-Cirrus	1380.0	75.0
11	SWIR	1610.0	143.0
12	SWIR	2190.0	2420



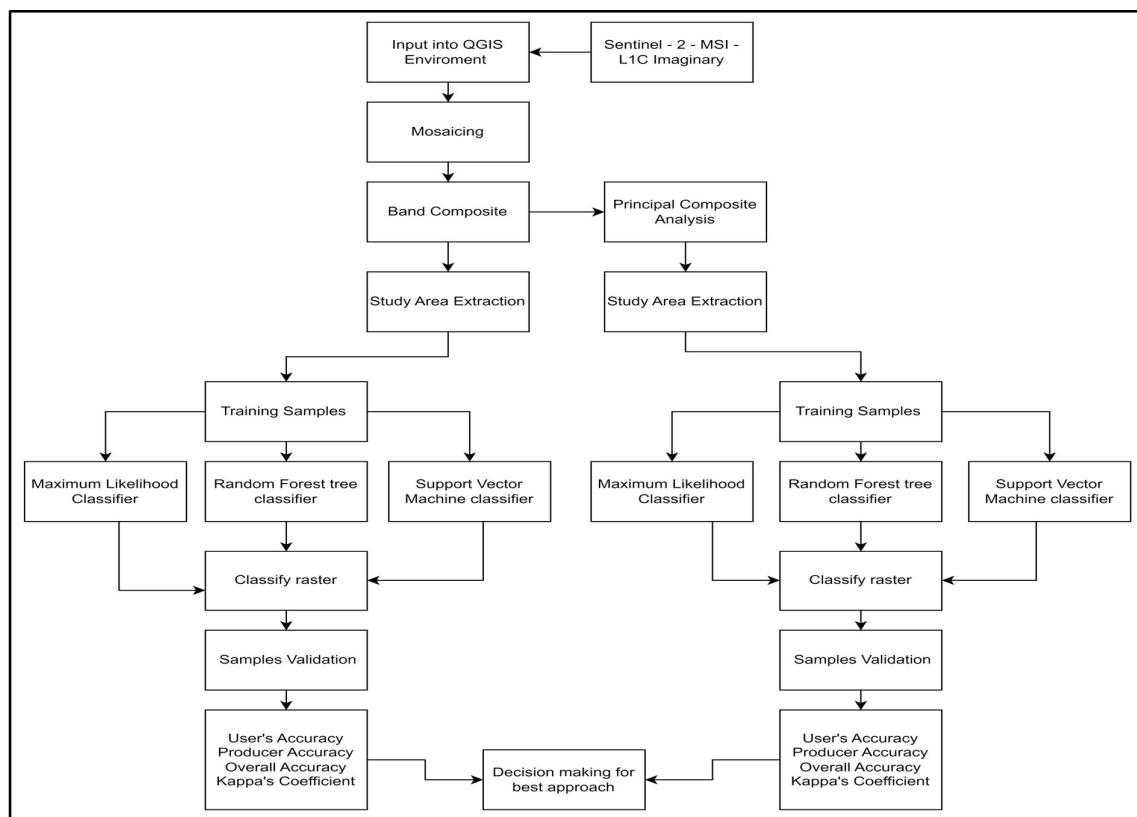


Fig. 2: The methodology adopted for LULC.

the raters demonstrate a strong level of consistency in their evaluations. Finally, a kappa value between 0.81 and 1.00 represents almost perfect agreement. In this case, the raters are nearly always in sync, providing highly reliable and consistent assessments.

## RESULTS AND DISCUSSION

### Principal Component Analysis of all Bands

Principal component analysis (PCA) was employed to compress the Sentinel-2 multispectral data. This technique aims to statistically capture the most significant evidence from the original bands (bands 2-12) into a reduced set of uncorrelated components termed principal components (PCs). The first few PCs inherently capture the majority of the data's variability (Rana et al. 2020). Notably, the 1<sup>st</sup> principal component (PC1), resulting from the 1<sup>st</sup> eigenvector, captured the greatest portion of the overall alteration within the Sentinel-2 dataset. Furthermore, the first three PCs collectively explained 98.85% of the eigenvalues, highlighting their effectiveness in representing the data. The remaining PCs displayed a decreasing trend in explained variance, corresponding to their respective eigenvalues.

Tables 2 and 3 give the redundancy of information among several bands, indicating that if this redundancy can be reduced through techniques like PCA, then the amount of information can be compressed without significant loss of valuable data. The author concentrated on the crucial data and excluded the later components (bands 4 to 12) because they appeared noisy and lacked useful information. Table 4 presents the eigenvalues and their corresponding cumulative percentage for principal components derived from Sentinel-2 bands. PCA reduced the associated Sentinel-2 dataset to a significantly smaller set of non-related variables that retain most of the original dataset's information. Fig. 3(a-c) displays the PCA bands obtained from the Sentinel-2 information, while Fig. 3(d-f) illustrates the frequency supply of these PC bands. The highest variance is found in the 1<sup>st</sup> PC, followed by the 2<sup>nd</sup> and 3<sup>rd</sup> components, according to the frequency distribution. The considered variances for PCA bands 1, 2, and 3 are 515,498.4, 263,079.1, and 8,843.772, respectively. Because of its high variance, the image produced from PCA band 1 data resembles the original image and contains the majority of the pertinent scene information. In multispectral remote sensing imagery, adjacent bands are often highly correlated and tend to provide similar information about an

Table 2: The covariance matrix for Sentinel-2 Bands.

Covariance Matrix												
Bands	2	3	4	5	6	7	8	8A	9	10	11	12
2	2534	3142	6366	5086	1155	74	246	-2	14	13356	12664	278
3	3142	4302	8274	7382	4774	4420	4968	787	19	19648	17347	5308
4	6366	8274	19571	16612	3777	130	1069	1127	58	48332	43630	2041
5	5086	7382	16612	17449	13716	13701	15026	3646	67	52130	42330	17894
6	1155	4774	3777	13716	60727	82391	82966	13782	53	45670	21589	93635
7	74	4420	130	13701	82391	114641	114838	18556	50	48490	17398	129299
8	246	4968	1069	15026	82966	114838	120480	18465	55	53913	21237	130521
8A	-2	787	1127	3646	13782	18556	18465	4161	17	13710	7207	21660
9	14	19	58	67	53	50	55	17	2	225	159	72
10	13356	19648	48332	52130	45670	48490	53913	13710	225	182512	143458	65389
11	12664	17347	43630	42330	21589	17398	21237	7207	159	143458	122336	27102
12	278	5308	2041	17894	93635	129299	130521	21660	72	65389	27102	147830

Table 3: Correlation matrix for Sentinel-2 Bands.

Correlation matrix												
Bands	2	3	4	5	6	7	8	8A	9	10	11	12
2	1.00	0.95	0.90	0.76	0.09	0.00	0.01	0.00	0.21	0.62	0.72	0.01
3	0.95	1.00	0.90	0.85	0.30	0.20	0.22	0.19	0.22	0.70	0.76	0.21
4	0.90	0.90	1.00	0.90	0.11	0.00	0.02	0.12	0.31	0.81	0.89	0.04
5	0.76	0.85	0.90	1.00	0.42	0.31	0.33	0.43	0.38	0.92	0.92	0.35
6	0.09	0.30	0.11	0.42	1.00	0.99	0.97	0.87	0.16	0.43	0.25	0.99
7	0.00	0.20	0.00	0.31	0.99	1.00	0.98	0.85	0.11	0.34	0.15	0.99
8	0.01	0.22	0.02	0.33	0.97	0.98	1.00	0.82	0.12	0.36	0.17	0.98
8A	0.00	0.19	0.12	0.43	0.87	0.85	0.82	1.00	0.19	0.50	0.32	0.87
9	0.21	0.22	0.31	0.38	0.16	0.11	0.12	0.19	1.00	0.39	0.34	0.14
10	0.62	0.70	0.81	0.92	0.43	0.34	0.36	0.50	0.39	1.00	0.96	0.40
11	0.72	0.76	0.89	0.92	0.25	0.15	0.17	0.32	0.34	0.96	1.00	0.20
12	0.01	0.21	0.04	0.35	0.99	0.99	0.98	0.87	0.14	0.40	0.20	1.00

object. The correlation between PCs 1 and 3, 1 and 2, and 2 and 3 was all found to be precisely zero. The random scatter observed in Fig. 4(a–c) and minimal correlation values indicate a complete absence of a relationship between the PCs. Consequently, classification tasks can often benefit from using the first few PCs instead of the entire original dataset. In our study, visual inspection revealed that PCA band 1 generally exhibited brighter pixel values and higher contrast compared to PCA band 2. This suggests that PCA band 1 may capture information related to high-variance features in the data, potentially making it more suitable for specific classification tasks depending on the target features of interest.

LULC classes were chosen based on a thorough understanding of the specific study area. The study identified four primary LULC classes: Agricultural land

Table 4: Total variance explained for Sentinel-2 Bands.

Percent and Accumulative Eigenvalues			
PC Layer	Eigen Value	% of Eigen Values	Accumulative of Eigen Values
1	515498.4	64.7168	64.7168
2	263079.1	33.0275	97.7444
3	8843.772	1.1103	98.8547
4	3807.653	0.478	99.3327
5	2426.928	0.3047	99.6374
6	1031.848	0.1295	99.7669
7	770.6416	0.0967	99.8636
8	462.6402	0.0581	99.9217
9	324.2669	0.0407	99.9624
10	230.819	0.029	99.9914
11	67.13062	0.0084	99.9998
12	1.29266	0.0002	100

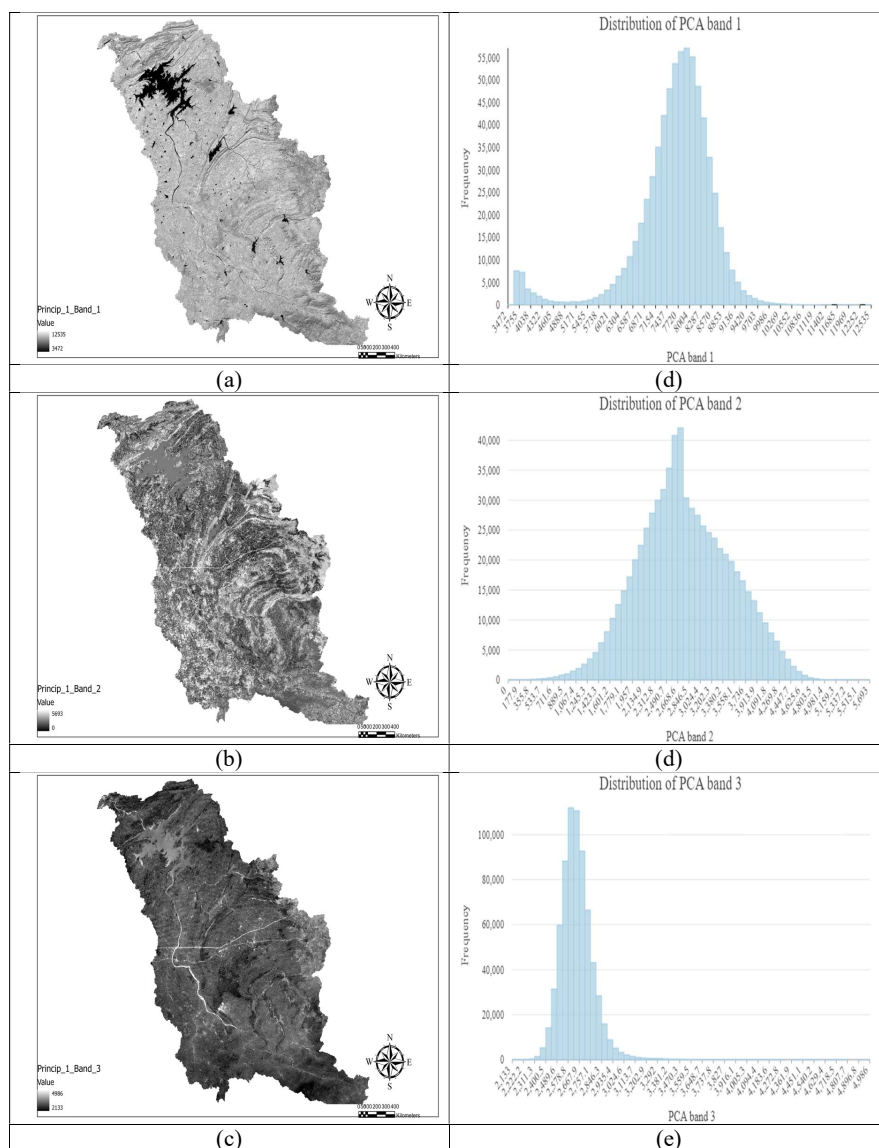


Fig. 3: Sentinel-2 data in three principal component bands (a-c) alongside their respective frequency distributions (d-f).

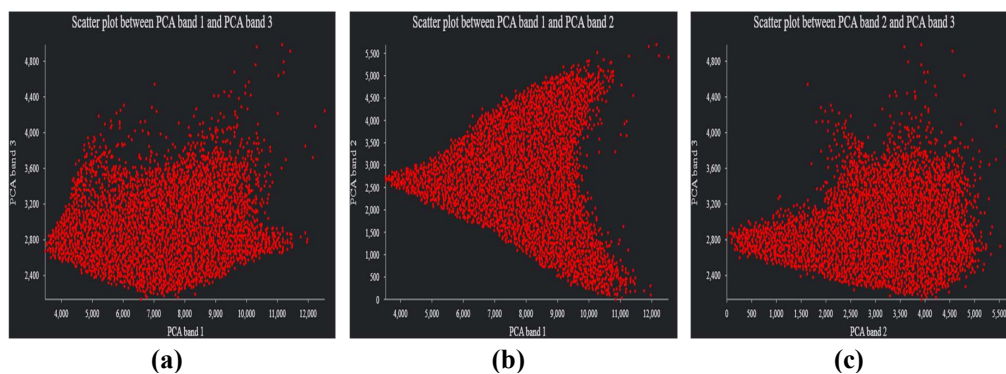


Fig. 4(a-c): Scatter Plots Reveal Weak Linkages Between Principal Components.

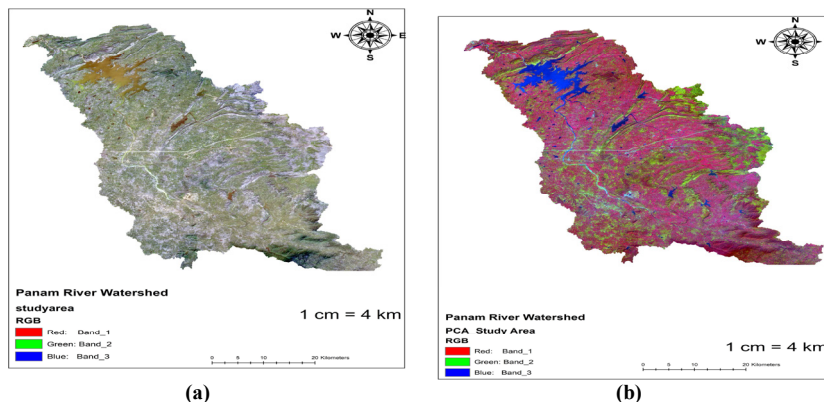


Fig. 5: Original band composite (a) and PCA - band composite (b).

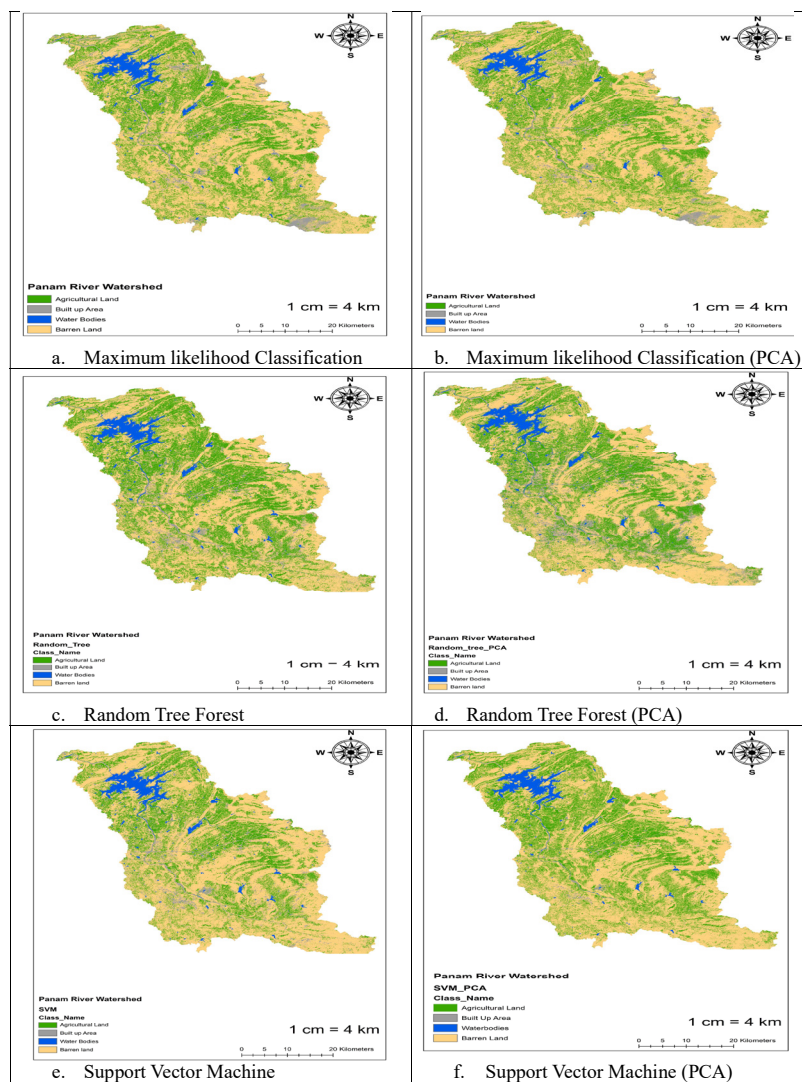


Fig. 6: Comparative analysis of classification methods: (a) MLE, (b) MLE with PCA, (c) Random Tree Forest, (d) Random Tree Forest with PCA, (e) Support vector machine, (f) Support vector machine with PCA.



(Agricultural zones, forest, etc.), water bodies (Reservoirs, rivers, streams, swamps, lakes), built-up areas (buildings and other manmade edifices, areas designated as mixed urban, industrial, or built territory), and barren land (Areas perpetually stripped bare, boasting less than 10% vegetation cover. Windswept plains and rocky outcrops reign supreme, their surfaces a tapestry of exposed dirt, sand, and stone). Classification was carried out using two approaches: the conventional method using Sentinel-2 original bands, as shown in Fig. 5(a), and a PCA-based approach, as shown in Fig. 5(b).

The effectiveness of these two LULC classification approaches was evaluated by assessing the predictive performance of three classification algorithms, namely MLC, RF, and SVM, along with the training data. Training data for each LULC category was collected as a set of pixels, and test data was obtained using stratified random sampling. The LULC maps generated using various classifiers for both approaches are depicted in Fig. 6(a-f).

### Performance Evaluation

The authors evaluated the efficiency of each model, RF, and SVM for both approaches by considering the user's accuracy (UA) and producer's accuracy (PA) for each LULC class. The results for each class were individually outlined as the models exhibited varying performance across different types. The specifics of UA and PA for each LULC class using MLC, RF, and SVM were thoroughly examined and described.

#### UA and PA for Agricultural Land

The land cover classification results for Agricultural Land using the Sentinel-2 Conventional Approach and Sentinel-2 PCA Approach with different classifiers reveal varying levels of accuracy. As shown in Table 5 under the Conventional Approach, both MLC and RF classifiers achieve high User's accuracy (UA) for Agricultural Land, with MLC reaching 100.00% UA and a corresponding Producer's accuracy (PA) of 65.63%, while RF achieves a UA of 100.00% and a PA of 62.50%. As shown in Table 6 in the PCA Approach, MLC shows a decrease in accuracy for Agricultural Land, with a UA of 80.95% and a lower PA of 39.29%. RF experiences a significant decline in both UA (62.79%) and PA (3.57%) for Agricultural Land, indicating diminished accuracy. SVM exhibits a perfect UA of 100.00% with a PA of 50.00% under the Conventional Approach, and in the PCA Approach, it maintains a relatively high UA of 82.61% with a lower PA of 32.14%. These results highlight the varied performance of classifiers in accurately classifying Agricultural Land under different approaches.

#### UA and PA for Built-Up Areas

In the land cover classification results for Built-up Areas using the Sentinel-2 Conventional Approach and Sentinel-2 PCA Approach with different classifiers, there are notable differences in the accuracy metrics. As shown in Table 5, under the conventional approach, maximum likelihood classification (MLC) achieves a user's accuracy (UA) of 80.95%, indicating reasonably accurate classification, with a high producer's accuracy (PA) of 94.44%. Random Forest Tree performs exceptionally well, achieving a perfect UA of 100.00% and a high PA of 91.67%, showcasing precise classification for built-up areas. Support vector machine (SVM) exhibits a UA of 78.05% with a corresponding PA of 88.89%. As shown in Table 6, the PCA Approach, MLC maintains accuracy with a UA of 81.58% and a perfect PA of 100.00%. Random Forest Tree achieves a UA of 83.78% and a perfect PA of 100.00%, indicating reliable classification. SVM also demonstrates consistent accuracy, with a UA of 83.78% and a perfect PA of 100.00%. These results highlight the varying performance of classifiers in accurately classifying Built-up Areas under different approaches, with each classifier showcasing strengths in specific accuracy metrics.

#### UA and PA for Water Bodies

In the land cover classification results for Water Bodies using the Sentinel-2 Conventional Approach and Sentinel-2 PCA Approach with different classifiers, distinct patterns in classification accuracy emerge. As shown in Table 5 under the CA, MLC exhibits high accuracy, achieving a UA of 96.43% and a PA of 79.41%. RFT, while displaying a lower UA of 82.93%, achieves a perfect PA of 100.00%, indicating precise classification for Water Bodies. SVM demonstrates a UA of 96.77% with a PA of 88.24%. As shown in Table 6, in the PCA Approach, all classifiers perform exceptionally well for water bodies. MLC achieves a perfect UA of 100.00% and a high PA of 92.11%. Random forest tree and SVM both achieve perfect UA and PA values of 100.00% and 97.37%, respectively, showcasing precise and consistent classification for Water Bodies.

#### UA and PA for Barren Land

In the land cover classification results for Barren Land using the Sentinel-2 CA and Sentinel-2 PCA approach with different classifiers, there are noticeable variations in classification accuracy.

As shown in Table 5 under the CA, MLC achieves a UA of 70.67%, indicating moderate accuracy, with a high PA of 98.15%. RTF displays slightly higher UA at 71.83%,

Table 5: Accuracy of different Classifiers for LULC using conventional approach for Sentinel-2 data set.

Sr. No	Conventional approach					
	Classifiers	Classes	UA	PA	OA	Kappa
1.	Maximum Likelihood Classification	Agricultural Land	100.00%	65.63%	80.13%	<b>0.8013</b>
		Built up Area	80.95%	94.44%		
		Water bodies	96.43%	79.41%		
		Barren Land	70.67%	98.15%		
2.	Random Forest Tree	Agricultural Land	100.00%	62.50%	82.69%	<b>0.8269</b>
		Built up Area	100.00%	91.67%		
		Water bodies	82.93%	100.00%		
		Barren Land	71.83%	94.44%		
3.	Support Vector Machine	Agricultural Land	100.00%	50.00%	81.41%	<b>0.8141</b>
		Built up Area	78.05%	88.89%		
		Water bodies	96.77%	88.24%		
		Barren Land	73.97%	100.00%		

Table 6: Accuracy of different classifiers for LULC using PCA approach for Sentinel-2 dataset.

Sr. No	PCA approach					
	Classifiers	Classes	UA	PA	OA	Kappa
1.	Maximum Likelihood Classification	Agricultural Land	80.95%	39.29%	83.22%	0.8322
		Built up Area	81.58%	100.00%		
		Water bodies	100.00%	92.11%		
		Barren Land	73.47%	78.26%		
2.	Random Forest Tree	Agricultural Land	62.79%	3.57%	83.92%	0.8392
		Built up Area	83.78%	100.00%		
		Water bodies	100.00%	97.37%		
		Barren Land	96.15%	54.35%		
3.	Support Vector Machine	Agricultural Land	82.61%	32.14%	86.01%	0.8601
		Built up Area	83.78%	100.00%		
		Water bodies	100.00%	97.37%		
		Barren Land	78.26%	78.26%		

with a PA of 94.44%. SVM performs well, achieving a UA of 73.97% and a perfect PA of 100.00%, indicating reliable classification for Barren Land.

As shown in Table 6 of the PCA Approach, MLC maintains reasonable accuracy with a UA of 73.47% and a PA of 78.26%. Random Forest Tree demonstrates significantly higher accuracy, with a UA of 96.15% and a PA of 54.35%, indicating precise classification for Barren Land. SVM shows consistent accuracy with a UA of 78.26% and a PA of 78.26%.

### Overall Accuracy and Kappa's Coefficient

In evaluating the land cover classification results using the Sentinel-2 Conventional Approach and Sentinel-2 PCA Approach with different classifiers, Overall Accuracy (OA)

and Kappa coefficients provide insights into the performance of each approach. Under the Conventional Approach, MLC achieves an OA of 80.13% with a Kappa coefficient of 0.8013, indicating reasonably accurate classification. Random Forest Tree demonstrates higher accuracy, with an OA of 82.69 % and a Kappa coefficient of 0.8269. SVM maintains competitive accuracy with an OA of 81.41% and a Kappa coefficient of 0.8141.

In the PCA approach, MLC showcases improved accuracy with an OA of 83.22 % and a Kappa coefficient of 0.8322. Random Forest Tree maintains similar accuracy, achieving an OA of 83.92 % and a Kappa coefficient of 0.8392. Notably, SVM excels with the highest accuracy, presenting an OA of 86.01% and a Kappa coefficient of 0.8601.

The practical advantages of using the PCA-SVM method for land use and land cover (LULC) classification are substantial, particularly in addressing issues like high dimensionality and data complexity. By reducing the number of input features, PCA captures the most important variance in the data, thereby improving computational efficiency and reducing the risk of overfitting. This makes PCA-SVM particularly useful in cases where the training data is limited, as SVM's strong generalization capability allows it to perform well even with fewer samples. Moreover, PCA-SVM proves highly effective in heterogeneous landscapes, where complex land use patterns (e.g., urban and agricultural mixes) require sophisticated classification models that can handle intricate decision boundaries. As demonstrated in the study, PCA-SVM achieved the highest OA and Kappa coefficients, making it a superior approach for LULC classification, especially in regions with limited training samples and diverse landscape features.

## CONCLUSIONS

In conclusion, this study undertook land use and land cover classification in the study area, delineating four primary classes: Agricultural land, water bodies, built-up areas, and barren land. The predictive performance of three classification algorithms (MLE, RF, and SVM) was evaluated using both traditional and PCA-based approaches using the original bands of Sentinel-2. The results demonstrated varying accuracies across land cover classes and classifiers. Particularly noteworthy was the Sentinel-2 PCA Approach, notably with the Support Vector Machine classifier, which exhibited superior accuracy for Agricultural Land (UA: 82.61%, PA: 32.14%), Built-up Area (UA: 83.78%, PA: 100.00%), Water Bodies (UA: 100.00%, PA: 97.37%), and Barren Land (UA: 78.26%, PA: 78.26%) compared to the Conventional Approach.

The detailed assessment of User's Accuracy (UA), Producer's Accuracy (PA), Overall Accuracy (OA), and Kappa coefficients provided comprehensive insights into the strengths and weaknesses of each approach and classifier. With an Overall Accuracy of 86.01% and a Kappa coefficient of 0.8601, the Sentinel-2 PCA Approach with the SVM classifier emerged as the most effective approach for accurate land cover classification in this study. These findings underscore the potential applicability of this approach for land use and land cover mapping and monitoring throughout similar regions, demonstrating its utility for broader applications in land cover studies. The integration of Sentinel-2 data with advanced classification methods can contribute significantly to more accurate and efficient land cover assessments in diverse geographical areas.

## ACKNOWLEDGEMENT

The authors are grateful to the U. S. Geological Survey (USGS) Earth Explorer (<https://earthexplorer.usgs.gov/>) for offering access to Sentinel 2 data, which played a crucial role in our analysis.

## REFERENCES

- Asif, M., Li-Qun, Z., Zeng, Q., Atiq, M., Ahmad, K., Tariq, A. and Hatamleh, A.A., 2023. Comprehensive genomic analysis of *Bacillus paralicheniformis* strain BP9, the pan-genomic and genetic basis of biocontrol mechanism. *Computational and Structural Biotechnology Journal*, 21, pp.4647-4662.
- Avci, C., Budak, M., Yağmur, N. and Balçık, F., 2023. Comparison between random forest and support vector machine algorithms for LULC classification. *International Journal of Engineering and Geosciences*, 8(1), pp.1-10.
- Belgiu, M. and Dragut, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp.24-31.
- Blain, T., n.d. Out-Of-Bag Error in Random Forests. [online] Available at: <https://tblainuob.github.io/files/OOBErrorInRF.pdf>
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), pp.5-32.
- Breiman, L., 2004. Bagging predictors. *Machine Learning*, 24, pp.123-140.
- Cheruto, M.C., Kauti, M.K., Kisangau, D.P. and Kariuki, P.C., 2016. Assessment of land use and land cover change using GIS and remote sensing techniques: A case study of Makuani County, Kenya. *Journal of Remote Sensing & GIS*, 5(4), p.1000175.
- Chowdhury, M.S., 2024. Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network, and maximum likelihood method in land use/cover classification of urban setting. *Environmental Challenges*, 14, p.100800.
- Fakeye, A.M., Aitsebaomo, F.O., Osadebe, C.C., Lamidi, R.B. and Okonufua, E.O., 2015. Digital modeling of land use changes in some parts of Eastern Nigeria. *American Journal of Remote Sensing*, 3(3), pp.37-42.
- Fetene, D.T., Lohani, T.K. and Mohammed, A.K., 2023. LULC change detection using support vector machines and cellular automata-based ANN models in Guna Tana watershed of Abay basin, Ethiopia. *Environmental Monitoring and Assessment*, 195(11), p.1329.
- Halder, S., Das, S. and Basu, S., 2023. Use of support vector machine and cellular automata methods to evaluate the impact of an irrigation project on LULC. *Environmental Monitoring and Assessment*, 195(1), p.3.
- Huang, C., Davis, L. and Townshend, J., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), pp.725-749.
- Huang, G.-B., Zhou, H., Ding, X. and Zhang, R., 2011. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2), pp.513-529.
- Jayabaskaran, M. and Das, B., 2023. Land Use Land Cover (LULC) Dynamics by CA-ANN and CA-Markov Model Approaches: A Case Study of Ranipet Town, India. *Nature Environment & Pollution Technology*, 22(3), p.32.
- Kunapuli, G., 2023. *Ensemble Methods for Machine Learning*. Simon and Schuster, Manning, UK.
- Lakhera, S. and Rahi, D.C., 2021. Detection of change in land cover in Jabalpur District from 1991–2021 using remote sensing. *Research Square*, 3, p.229. DOI
- Loukika, K.N., Keesara, V.R. and Sridhar, V., 2021. Analysis of land use and land cover using machine learning algorithms on Google Earth engine for Munneru River Basin, India. *Sustainability*, 13(24), p.13758.

- Mahmud, K.H. and Hafsa, B., 2016. PCA as a tool of LULC based change detection in a suburban area of Dhaka city, Bangladesh. *The Jahangirnagar Review, Part II: Social Science*, 36, p.112
- Moharram, M.A. and Sundaram, D.M., 2023. Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges, and future directions. *Neurocomputing*, 536, pp.90-113.
- Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), pp.247-259.
- Pal, M. and Foody, G.M., 2012. Evaluation of SVM, RVM, and SMLR for accurate image classification with limited ground data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(5), pp.1344-1355.
- Parracciani, C., Gigante, D., Mutanga, O., Bonafoni, S. and Vizzari, M., 2024. Land cover changes in grassland landscapes: combining enhanced Landsat data composition, Land Trendr, and machine learning classification in Google Earth Engine with MLP-ANN scenario forecasting. *GIScience & Remote Sensing*, 61(1), p.2302221.
- Rana, V.K., Maruthi, T. and Suryanarayana, V., 2020. Performance evaluation of MLE, RF, and SVM classification algorithms for watershed-scale land use/land cover mapping using Sentinel 2 bands. *Remote Sensing Applications: Society and Environment*, 19, p.100351.
- Richards, J., 2013. Remote Sensing Digital Image Analysis. In: *Remote Sensing Digital Image Analysis*, 5th ed. Springer-Verlag Berlin and Heidelberg GmbH & Co., Switzerland, pp.203-246.
- Schirpke, U., Tasser, E., Borsky, S., Braun, M., Eitzinger, J., Gaube, V., Getzner, M., Glatzel, S., Gschwantner, T. and Kirchner, M., 2023. Past and future impacts of land-use changes on ecosystem services in Austria. *Journal of Environmental Management*, 345, p.118728.
- Shekar, P.R. and Mathew, A., 2022. Prioritizing sub-watersheds using morphometric analysis, principal component analysis, and land use/land cover analysis in the Kinnerasani River basin, India. *H2Open Journal*, 5(3), pp.490-514.
- Shekar, P.R., Mathew, A., PS, A. and Gopi, V.P., 2023. Sub-watershed prioritization using morphometric analysis, principal component analysis, hypsometric analysis, land use/land cover analysis, and machine learning approaches in the Peddavagu River Basin, India. *Journal of Water and Climate Change*, 14(1), pp.1-20.
- Shrestha, R., Bhandari, S. and Twayana, R., 2021. Monitoring land cover/land use change using remote sensing and GIS technique at Dhulikhel and Banepa municipality, Nepal. *International Journal of Current Research in Science Engineering & Technology*, 3(5), p.712.
- Siqueira, R.G., Moquedace, C.M., Fernandes-Filho, E.I., Schaefer, C.E., Francelino, M.R., Sacramento, I.F. and Michel, R.F., 2024. Modeling and prediction of major soil chemical properties with Random Forest: Machine learning as a tool to understand soil–environment relationships in Antarctica. *Catena*, 235, p.107677.
- Somayajula, V.K.A., Ghai, D. and Kumar, S., 2021. Land Use/Land Cover Change Analysis using NDVI, PCA. In: *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE 2021, pp.723-746.
- Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M. and Liang, X., 2024. An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, p.121549.
- Tao, J., Y. Gu, X. Hao, R. Liang, B. Wang, Z. Cheng, B. Yan, and G. Chen. 2023. Combination of Hyperspectral Imaging and Machine Learning Models for Fast Characterization and Classification of Municipal Solid Waste. *Resources, Conservation & Recycling* 188, 106731.
- Tao, J., Gu, Y., Hao, X., Liang, R., Wang, B., Cheng, Z., Yan, B. and Chen, G. 2023. Combination of hyperspectral imaging and machine learning models for fast characterization and classification of municipal solid waste. *Resources, Conservation & Recycling*, 188, p.106731. DOI
- Tilahun, A. and Teferie, B. 2015. Accuracy assessment of land use land cover classification using Google Earth. *American Journal of Environmental Protection*, 4(4), 193–198. DOI
- Tiwari, A.K., Pal, A. and Kanchan, R. 2024. Mapping and monitoring of land use/land cover transformation using geospatial techniques in Varanasi City Development Region, India. *Nature Environment & Pollution Technology*, 23(1), p.31. DOI
- Van Groenigen, J. and Stein, A. 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality*, 17, p.61.