INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

OPEN ACCESS

Received: 16/01/2025
Accepted: 23/04/2025
Published: 14/05/2025

Citation: Mohanty V, Behera DK, Panda AR, Swetanisha S (2025) Comparative Analysis of Machine Learning and Deep Learning Models for LULC Classification using Remote Sensing Data. Indian Journal of Science and Technology 18(18): 1397-1409. https://doi.org/10.17485/IJST/v18i18.104

* **Corresponding author**.

2081040@kiit.ac.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Comparative Analysis of Machine Learning and Deep Learning Models for LULC Classification using Remote Sensing Data

Vidya Mohanty[1]*, Dayal Kumar Behera[1], Amiya Ranjan Panda[1], Subhra Swetanisha[2]

1 School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India
2 School of Computing, Trident Academy of Technology, Bhubaneswar, Odisha, India

Abstract

Objectives: The primary objective of this study is to evaluate and compare the performance of machine learning and deep learning models for Land Use and Land Cover (LULC) classification using remote sensing data. Specifically, it assesses Support Vector Machine (SVM), XGBoost, an ensemble model (SVM + XGBoost), and a Deep Neural Network (DNN) on pre-processed hyperspectral datasets (Pavia University, Indian Pines) and raw satellite imagery from the Twin Cities of Odisha, India. **Method:** The study follows a systematic workflow, including data acquisition, preprocessing using Principal Component Analysis (PCA), and classification using machine learning and deep learning models. The models are evaluated based on Overall Accuracy (OA), Kappa Index, Precision, Recall, and F1-score. Statistical significance is tested using the Nemenyi post hoc test to determine the superiority of the proposed ensemble model over individual classifiers. **Findings:** The ensemble model (SVM + XGBoost) achieved the highest classification accuracy across all datasets: 95.4% for Pavia University, 88.7% for Indian Pines, and 93.6% for the Twin Cities of Odisha. The statistical analysis confirms that the ensemble model performs significantly better than XGBoost with a 95% confidence interval. The DNN model also demonstrated competitive performance but was slightly outperformed by the ensemble approach. **Novelty:** This study presents a comparative performance analysis of machine learning and deep learning models for LULC classification, demonstrating the effectiveness of an ensemble-based approach in improving classification accuracy. Unlike previous works, the study validates findings using real-world satellite data, applies statistical significance testing, and offers quantitative performance comparisons with existing literature, contributing to more accurate and scalable LULC classification methodologies.

Keywords: Remote Sensing; LULC; Machine Learning; Deep Neural Network; XGBoost

1 Introduction

The land on the earth surfaces can be observed in many ways, either by ground survey: which is more time consuming and expensive or by airborne platform. But the most effective way to do this is from the space. The Earth can be observed from the space using satellite. The images captured by the satellites are of very high spatial resolution and the images are consisting of multiple bands. Land Use and Land Cover (LULC) classification plays a vital role in environmental monitoring, urban planning, and sustainable land management. There are mainly two techniques are adopted for classifying the land: pixel-based and object-based approach[(1)]. In a pixel-based approach, the spectral information contained in individual image pixels is analysed. Using an image segmentation algorithm, the object-based technique collects picture pixels into spectrally homogeneous image objects and then classifies the individual objects. Machine learning models can be applied both in pixel-based and object-based approaches.

The increasing availability of high-resolution remote sensing data has led to significant advancements in classification techniques, ranging from traditional machine learning (ML) models to deep learning (DL) approaches[(2)]. Conventional methods such as Decision Trees, Maximum Likelihood Classification (MLC), and Support Vector Machines (SVM) have been widely used for LULC mapping[(3)]. However, their reliance on handcrafted features often limits their adaptability to diverse landscapes and high-dimensional datasets. Recent studies have explored the integration of ML algorithms for LULC classification, demonstrating improvements in predictive performance[(4)]. For instance, Aryal et al. (2023) conducted a comprehensive study on LULC classification in Melbourne, Australia, using various machine learning models. Their research demonstrated that ensemble learning techniques, particularly Random Forest and Gradient Boosting, significantly improved classification accuracy compared to standalone classifiers[(4)]. The study highlighted the importance of feature selection and hyperparameter tuning in optimizing ML performance for urban land classification. However, their approach lacked a comparative analysis with deep learning models, leaving open questions about the relative strengths of ML versus DL techniques in large-scale urban mapping.

Despite these advancements, traditional ML models often struggle with complex spatial patterns and spectral variability. To overcome these limitations, deep learning techniques, particularly Convolutional Neural Networks (CNNs)[(5)] and Deep Neural Networks (DNNs),[(6)] have gained attraction for LULC classification . Studies have shown that CNNs outperform ML classifiers by effectively capturing spectral and spatial correlations in remote sensing data[(6)],[(7)]. The goal of work[(8)] modeled to assess the efficacy of CNN algorithms for land categorization and LU change detection. Using three pre-trained CNN models AlexNet, GoogLeNet, and VGGNet, eight transferred CNN-based models for LU scene categorization were comprehensively assessed using remote sensing data. However, the application of deep learning is hindered by high computational costs and the need for large, labeled datasets.

Martínez et al. (2025) investigated the use of Synthetic Aperture Radar (SAR) backscatter and InSAR coherence for LULC classification in tropical high-mountain ecosystems[(9)]. Their findings indicated that incorporating SAR data into ML models enhances classification accuracy, particularly in regions with high vegetation variability. The study demonstrated the benefits of fusing multi-source remote sensing data but did not explore deep learning techniques, which have shown potential in automatically learning complex spatial and spectral relationships in LULC classification.

Similarly, Nigar et al. (2024) analyzed ML and DL algorithms for land classification using Google Earth Engine and Python[(8)]. Their study demonstrated that deep learning models, particularly CNNs, achieved higher classification accuracy than traditional ML approaches. However, ensemble ML techniques, such as XGBoost, provided better generalization and required fewer labeled samples. This work highlighted the trade-off between model complexity and training data availability but lacked statistical validation to confirm the superiority of one approach over another. However, these studies lack a direct statistical comparison between ML, ensemble learning[(10)], and DL models[(11)], making it difficult to determine the optimal approach for different geographic settings.

Furthermore, the practical application of these models remains an open challenge. Many studies rely on benchmark datasets without validation on real-world satellite imagery. Rengma and Yadav (2024) conducted a comparative study on patch-based LULC classification in diverse Indian landscapes, evaluating both ML and DL models[(12)]. Their results showed that CNNs outperformed traditional ML classifiers in capturing spatial dependencies in heterogeneous landscapes. However, the study acknowledged the challenge of obtaining large, labeled datasets required for deep learning training, emphasizing the need for hybrid models that combine ML interpretability with DL accuracy.

Tahir et al. (2025) explored the integration of CA-Markov modeling with GIS-based techniques for predicting LULC changes in dynamic landscapes[(13)]. Their study emphasized the importance of temporal analysis in sustainable land management and demonstrated that hybrid modeling approaches improve predictive accuracy. However, while the research effectively captured long-term LULC trends, it relied on traditional ML models and lacked an evaluation of deep learning methods, which are known for their superior feature extraction capabilities in high-resolution satellite imagery. While **Zhao et al. (2023)** provided an extensive review of deep learning applications in LULC classification, covering advancements in CNNs, transformers, and generative models [14]. The study noted that while DL models achieve state-of-the-art accuracy, their high computational cost and black-box nature limit their usability in real-world scenarios. Additionally, the review identified a gap in research related to statistical validation techniques for comparing ML and DL approaches, reinforcing the need for systematic performance benchmarking.

Addressing these gaps, this study conducts a comprehensive evaluation of ML and DL models for LULC classification, using both benchmark hyperspectral datasets (Pavia University, Indian Pines) and real-world remote sensing data (Twin Cities of Odisha, India). The key contributions of this work include:

1. A comparative performance analysis of traditional ML classifiers (SVM, XGBoost), an ensemble model (SVM + XGBoost), and a deep learning model (DNN).
2. Statistical validation using the Nemenyi post hoc test to assess the significance of ensemble learning over individual models.
3. Application of the models to real-world satellite data, ensuring the generalizability of findings beyond controlled datasets.

2 Methodology

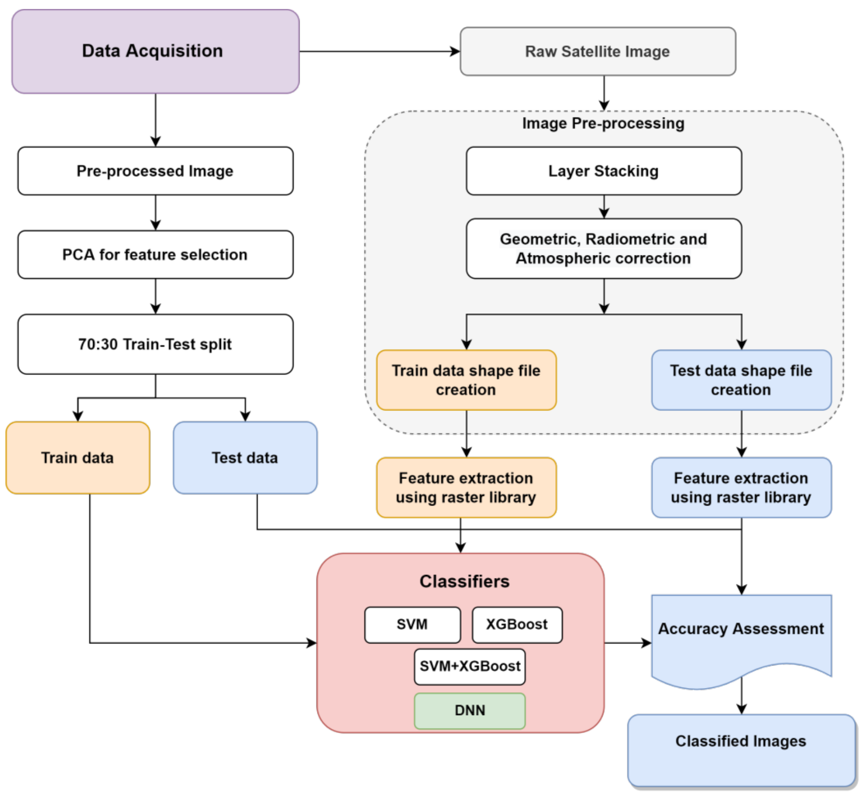
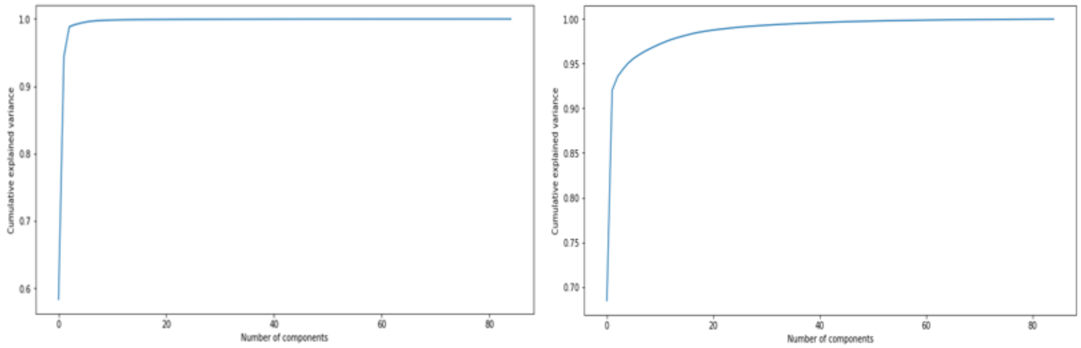


Fig 1. The Proposed Classification Model

The proposed technique is depicted in Figure 1. The benchmark dataset and raw satellite images are collected. The raw satellite image is undergone for data preprocessing whereas both the benchmark datasets are already pre-processed. The Principal Component Analysis (PCA) is applied on hyper spectral images: Pavia University and Indian Pines datasets. The significance of the PCA on Pavia is shown in Figure 2. Using the component's significance, 20 and 40 features are picked from the Pavia and Indian Pines datasets, respectively, as shown in Figure 2 (a), (b). Then, the training and testing datasets are considered, with 70% train data and 30% test data. Following that, classification is carried out utilising machine learning models and deep neural networks for the LULC classification of all the three datasets. The detailed explanation are given in the following sections.



(a) **(b)**

Fig 2. PCA on (a) Pavia University (b) Indian Pines

2.1 Study Area

Three study area includes: Pavia University and Indian Pines and Twin cities of Odisha. Data Source: [*https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes]*. Pavia University is situated in northern Italy. Prof. Paolo Gamba of Pavia University's Telecommunication and Remote Sensing Laboratory contributed this benchmark dataset. Another study area is Indian Pines. This data is collected by AVRIS sensor in Northwestern Indiana. Both these datasets are already pre-processed images. The third dataset is a raw satellite image of twin cities of Odisha, India.

2.2 Data Acquisition

Data is collected at Pavia University using a ROSIS (Reflective Optics System Imaging Spectrometer) sensor. Pavia University makes use of 103 spectral bands. Pavia University is an image with a resolution of 610*610 pixels. The geometric resolution is 1.3 meters. The ML and DNN models are used to classify into 9 separate classes such as 6631 samples of Asphalt, 18649 samples of Meadows, 2099 samples of Gravel, 3064 samples of Trees, 1345 samples of Painted metal sheets, 5029 samples of Bare Soil, 1330 samples of Bitumen, 3682 samples of Self-Blocking Bricks, and 947 samples of Shadows.

Indian Pines are collected from Pursue's University Multispec site. It makes use of 224 spectral bands. Indian Pines is an image with a resolution of 145*145 pixels. The Indian Pines mainly covered with variety of crops. The ground truth is available with 16 classes like 46 samples of Alfalfa, 1428 samples of Corn-notill, 830 samples of Corn-mintill, 237 samples of Corn, 483 samples of Grass-pasture, 730 samples of Grass-trees, 28 samples of Grass-pasture-mowed, 478 samples of Hay-windrowed, 20 samples of Oats, 972 samples of Soybean-notill, 2455 samples of Soybean-mintill, 593 samples of Soybean-clean, 205 samples of Wheat, 1265 samples of Woods, 386 samples of Buildings-Grass-Trees-Drives and 93 samples of Stone-Steel-Towers.

In the eastern section of the state are the twin cities of Odisha[1], Bhubaneswar and Cuttack, which are located between latitudes 20° 15' N and 20° 28' N and longitudes 85° 52' E and 85° 54' E. The twin cities' urban administrative region is examined in this study. Landsat satellite-8 ETM+ data for 2020 were used to determine the area's land use and land cover (LU/LC) and classified into seven classes such as a river, canal, pond, forest, urban, agricultural area and sand.

2.3 Visualization of Ground Truth

Figure 3 displays the Pavia University and Indian Pines and twin cities of Odisha's ground truths, respectively. The black colour denotes pixels that have no data in Figure 3 (a) and (b) and will be eliminated during the classification process.

2.4 Training and Testing

2.4.1. Using Deep Neural Network

Twelve fully connected/completely linked or dense layers, batch normalization, and dropout layers are employed in a deep neural network. The data is stratified in a 70:30 ratio between train and test. Keras is used to construct a deep neural network for classifying the land cover of the Pavia University, Indian Pines. The activation function for the aforementioned deep neural network is a rectified linear unit (ReLU), and for the final layer, a softmax activation function is used. Table 1 Depicts the model

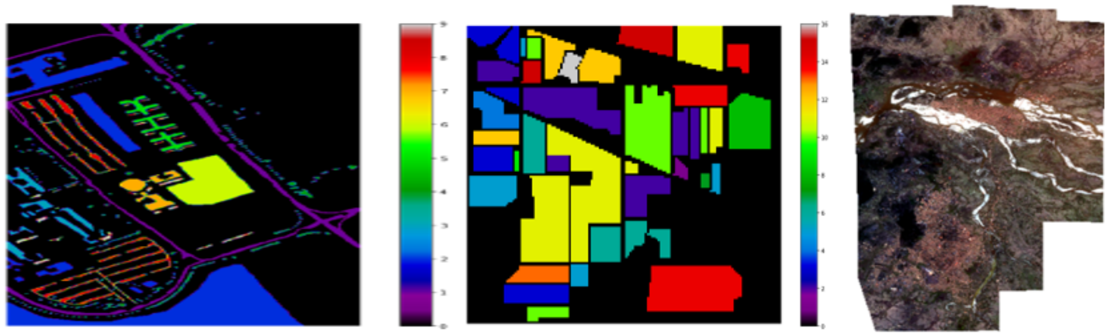


Fig 3. Ground Truth of (a) Pavia University (b) Indian Pines (c) Twin Cities of Odisha

details for Pavia University. Out of 89,574 parameters, 89368 are trainable parameters and 206 are non-trainable parameters. Table 2 shows the details of DNN used for Indian Pines dataset. Out of 102,609 parameters, 102,209 are trainable parameters and 400 are non-trainable parameters.

Table 1. Model Details for Pavia University

Layer(type)	Output Shape	Param #
Layer1 (Dense)	(none, 128)	13312
Layer2 (Dense)	(none, 128)	16512
Layer3 (Dense)	(none, 128)	16512
Layer4 (Dense)	(none, 128)	16512
Dropout1 (Dropout)	(none, 128)	0
Layer5 (Dense)	(none, 64)	8256
Layer6 (Dense)	(none, 64)	4160
Layer7 (Dense)	(none, 64)	4160
Layer8 (Dense)	(none, 64)	4160
Dropout2 (Dropout)	(none, 64)	0
Layer9 (Dense)	(none, 32)	2080
Layer10 (Dense)	(none, 32)	1056
Laye11 (Dense)	(none, 32)	1056
Layer12 (Dense)	(none, 32)	1056
Output Layer (Dense)	(none, 10)	330

The defined DNN, Adam optimizer, categorical cross-entropy, accuracy as a metric, and callbacks are used in the training process. The accuracy and loss graphs for Pavia University and Indian Pines, respectively, are shown in Figure 4 (a), (b), together with the output. The X-axis indicates the number of epochs, while the Y-axis indicates the percentage.

3 Results and Discussion

The performance analysis using DNN, SVM+XGBoost are also represented in the form of confusion matrix depicted in Figure 5 and Figure 6. The results obtained are compared with the state of the art models like SVM, XGBoost. The ensembling model using SVM+XGBoost is also used for performance analysis. XGBoost is an ensemble learning methodology that combines many decision trees in series. Apart from these machine learning models, DNN is also used for performance analysis.

The performance analysis of LULC of Pavia University and Indian Pines using DNN, machine learning models in terms of Precision, recall, F1-Score are represented in Table 3, Table 4 respectively in form of classification report. Figure 7 also shows the graphical performance analysis representation on Pavia. The other class level accuracy such as Producer accuracy (PA) and

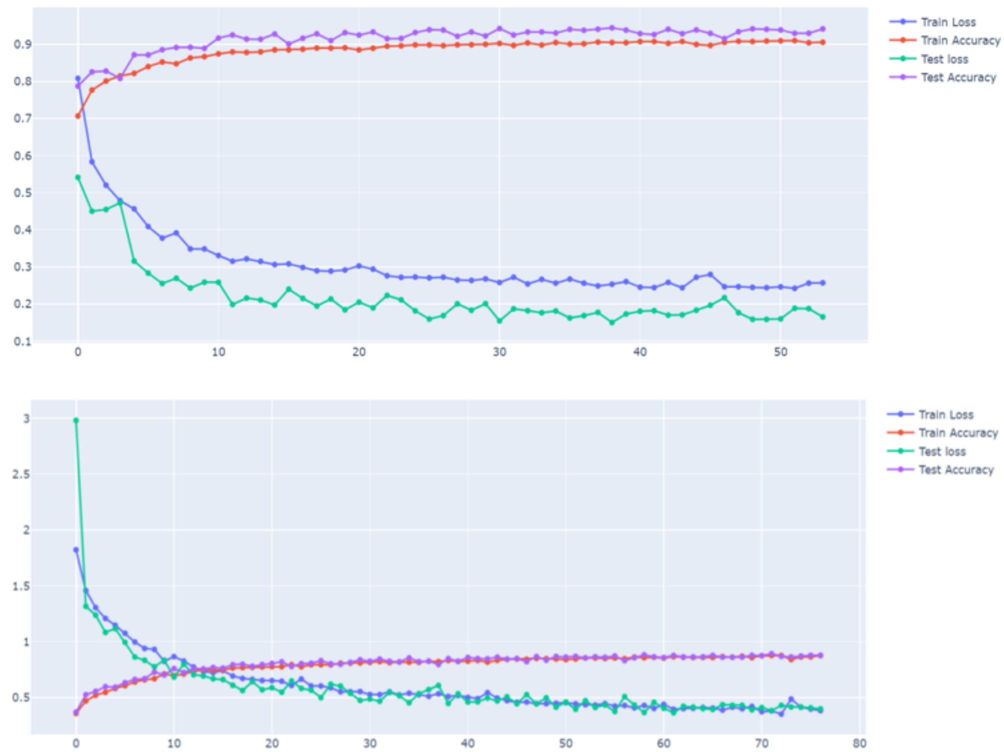


Fig 4. (a) - Test Loss and Accuracy Graph of (a) Pavia University (b) - Train –Test Loss and Accuracy Graph of Indian Pines

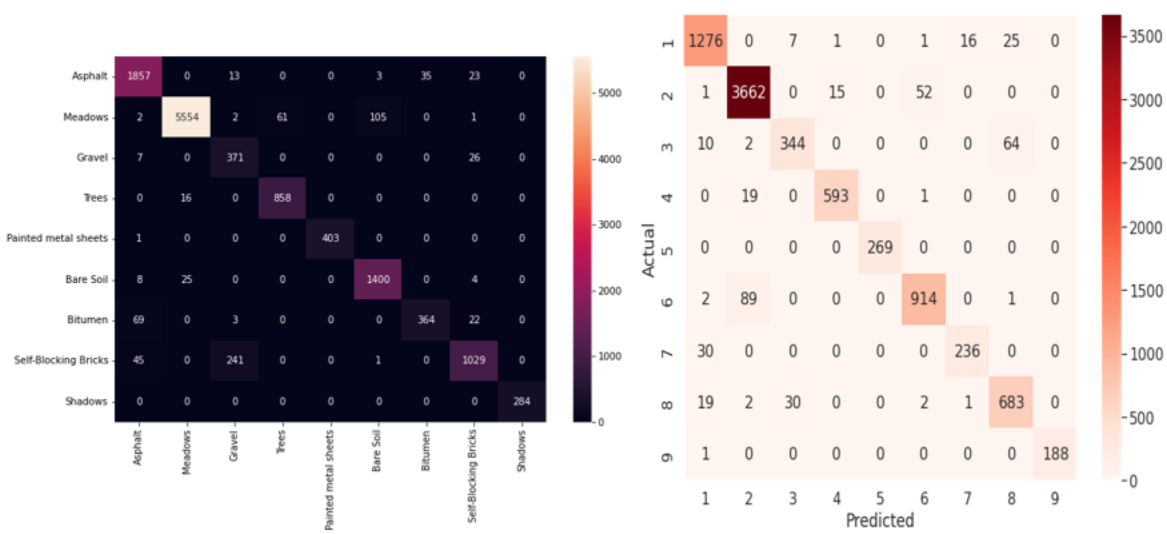


Fig 5. Confusion Matrix of Pavia University using (a) DNN (b) SVM +XGBoost

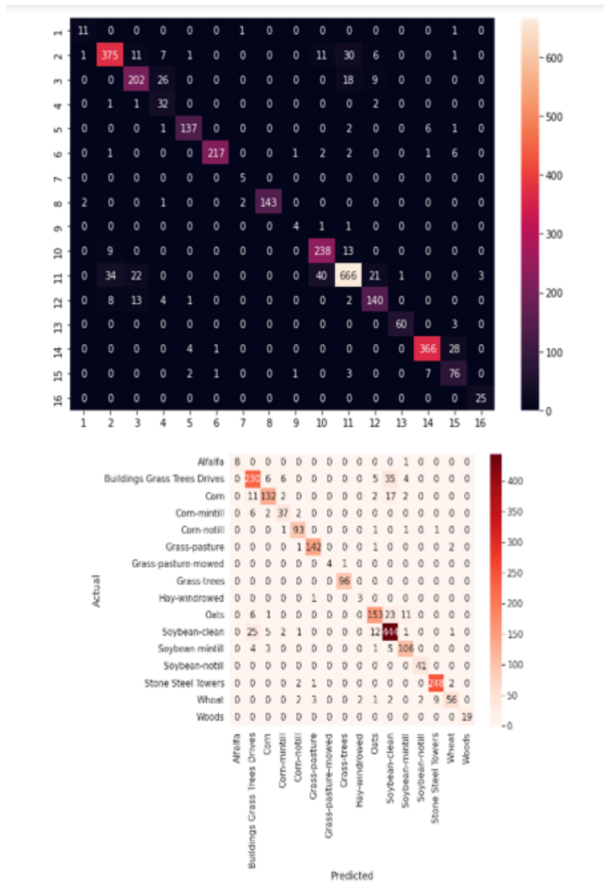


Fig 6. Confusion Matrix of Indian Pines using (a) DNN (b) SVM + XGBoost

User accuracy (UA) for Pavia University, Indian Pines and Twin Cities of Odisha are depicted in Table 5, Table 6 and Figure 8 respectively.

Table 2. Performance Analysis of LULC of Pavia University using Different Classifier

Class	Precision				Recall				F1-Score			
	DNN	SVM+XG Boost	XGBoost	SVM	DNN	SVM+XG Boost	XGBoost	SVM	DNN	SVM+XG Boost	XGBoost	SVM
Asphalt	0.93	0.95	0.93	0.95	0.96	0.96	0.92	0.96	0.95	0.96	0.93	0.96
Meadows	0.99	0.97	0.93	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.95	0.97
Gravel	0.59	0.90	0.85	0.91	0.92	0.82	0.71	0.80	0.72	0.86	0.78	0.85
Trees	0.93	0.97	0.97	0.98	0.98	0.97	0.93	0.97	0.96	0.97	0.95	0.97
Painted metal sheets	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Bare Soil	0.93	0.94	0.91	0.94	0.97	0.91	0.77	0.90	0.95	0.93	0.84	0.92
Bitumen	0.91	0.93	0.87	0.94	0.79	0.89	0.82	0.88	0.85	0.91	0.84	0.91
Self-Blocking Bricks	0.93	0.88	0.82	0.87	0.78	0.93	0.89	0.93	0.85	0.90	0.85	0.90
Shadows	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	0.99

Table 3. Performance Analysis of LULC of Indian Pines using Different Classifier

Class	Precision				Recall				F1-Score			
	DNN	SVM+XG Boost	XGBoostSVM		DNN	SVM+XG Boost	XGBoostSVM		DNN	SVM+XG Boost	XGBoostSVM	
Alfalfa	0.79	1.00	0.89	1.00	0.85	0.89	0.89	0.89	0.81	0.94	0.89	0.94
Corn-notill	0.88	0.82	0.68	0.82	0.85	0.80	0.68	0.79	0.86	0.81	0.68	0.81
Corn-mintill	0.81	0.89	0.80	0.91	0.79	0.80	0.55	0.77	0.80	0.84	0.65	0.83
Corn	0.45	0.77	0.61	0.83	0.89	0.79	0.43	0.72	0.60	0.78	0.50	0.77
Grass-pasture	0.94	0.92	0.94	0.94	0.93	0.96	0.88	0.95	0.94	0.94	0.91	0.94
Grass-trees	0.99	0.97	0.89	0.95	0.94	0.97	0.97	0.97	0.97	0.97	0.93	0.96
Grass-pasture-mowed	0.62	1.00	1.00	1.00	1.00	0.80	0.80	0.80	0.77	0.89	0.89	0.89
Hay-windrowed	1.00	0.99	0.98	0.99	0.97	1.00	0.99	1.00	0.98	0.99	0.98	0.99
Oats	0.67	0.60	0.33	1.00	0.67	0.75	0.25	0.50	0.67	0.67	0.29	0.67
Soybean-notill	0.82	0.87	0.78	0.88	0.92	0.79	0.65	0.77	0.86	0.83	0.71	0.82
Soybean-mintill	0.90	0.84	0.72	0.82	0.85	0.90	0.86	0.90	0.87	0.87	0.79	0.86
Soybean-clean	0.79	0.84	0.72	0.80	0.83	0.89	0.71	0.90	0.81	0.87	0.71	0.85
Wheat	0.98	0.95	0.97	0.98	0.95	1.00	0.93	1.00	0.97	0.98	0.95	0.99
Woods	0.96	0.96	0.91	0.95	0.92	0.98	0.99	0.98	0.94	0.97	0.95	0.97
Buildings-Grass-Trees-Drives	0.66	0.92	0.84	0.88	0.84	0.73	0.60	0.75	0.74	0.81	0.70	0.81
Stone-Steel-Towers	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00

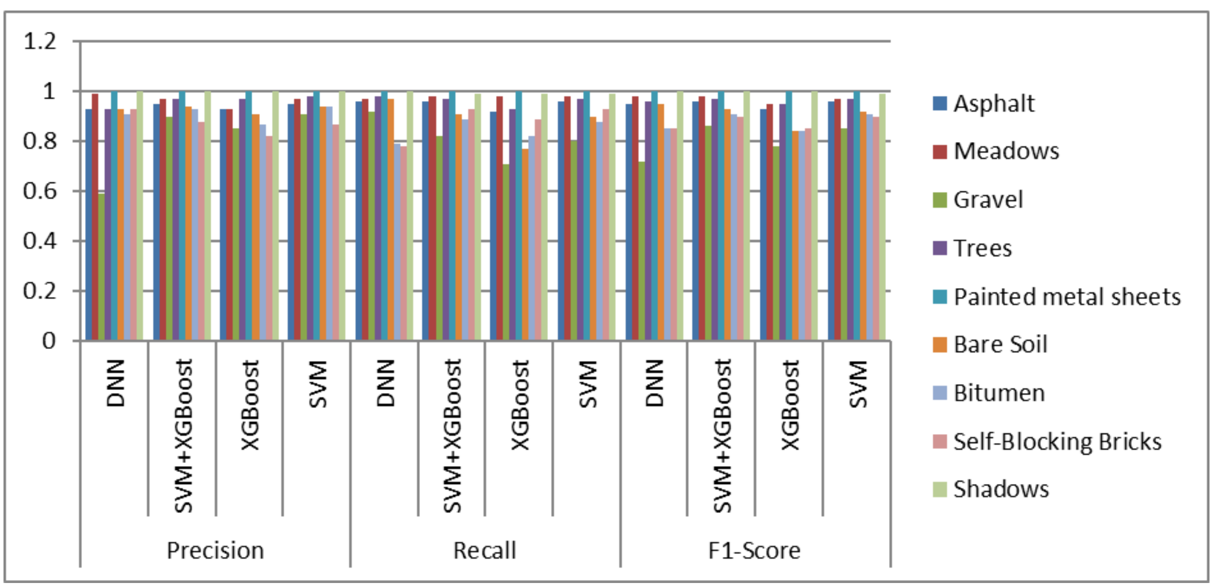


Fig 7. Performance Analysis of LULC of Pavia University using Different Classifier

Table 4. Producer and User Accuracy of Different Classifiers on Pavia University

Class	Producer Accuracy (PA)				User Accuracy (UA)			
	DNN	SVM+XG Boost	XGBoost	SVM	DNN	SVM+XG Boost	XGBoost	SVM
Asphalt	93.36350	95.29500	92.59819	95.14200	96.16779	96.22926	92.45852	96.00302
Meadows	99.26720	97.03233	92.84441	96.62714	97.01310	98.17694	97.74799	98.31099
Gravel	58.88889	90.28871	85.38682	90.76087	91.83168	81.90476	70.95238	79.52381
Trees	93.36235	97.37274	96.93878	97.52883	98.16934	96.73736	92.98532	96.57423
Painted metal sheets	100.0000	100.0000	100.0000	100.0000	99.75248	100.000	100.000	100.000
Bare Soil	92.77667	94.22680	90.78180	94.05010	97.42519	90.85487	77.33598	89.56262
Bitumen	91.22807	93.28063	86.50794	94.35484	79.47598	88.72180	81.95489	87.96992
Self-Blocking Bricks	93.12217	88.35705	81.92020	87.02290	78.19149	92.67300	89.14518	92.80868
Shadows	100.0000	100.000	100.0000	100.000	78.19149	99.47090	99.47090	98.94180

Table 5. Producer and User Accuracy of Different Classifiers on Indian Pines

Class	Producer Accuracy (PA)				User Accuracy (UA)			
	DNN	SVM+XG Boost	XGBoost	SVM	DNN	SVM+XG Boost	XGBoost	SVM
Alfalfa	78.57143	100.000	88.88889	100.000	78.57143	88.88889	88.88889	88.88889
Corn-notill	87.61682	82.50000	67.59582	81.94946	84.65011	80.76923	67.83217	79.37063
Corn-mintill	81.12450	88.66667	80.00000	90.71429	79.21569	80.12048	55.42169	76.50602
Corn	45.07042	78.72340	60.60606	82.92683	88.88889	78.72340	42.55319	72.34043
Grass-pasture	94.48276	92.07921	94.44444	93.87755	93.19728	95.87629	87.62887	94.84536
Grass-trees	99.08676	96.62162	89.24051	95.30201	94.34783	97.94521	96.57534	97.26027
Grass-pasture-mowed	62.50000	100.000	100.000	100.000	100.000	80.00000	80.00000	80.00000
Hay-windrowed	100.000	98.96907	97.93814	98.96907	96.62162	100.000	98.95833	100.000
Oats	66.66667	60.00000	33.33333	100.000	66.66667	75.00000	25.00000	50.0000
Soybean-notill	81.50685	88.06818	77.91411	87.71930	91.53846	79.89691	65.46392	77.31959
Soybean-mintill	90.36635	84.62998	72.35495	81.73432	84.62516	90.83503	86.35438	90.22403
Soybean-clean	78.21229	84.12698	71.79487	80.45113	83.33333	89.07563	70.58824	89.91597
Wheat	98.36066	95.34884	97.43590	97.61905	95.23810	100.000	92.68293	100.000
Woods	96.31579	96.12403	90.90909	95.40230	91.72932	98.02372	98.81423	98.41897
Buildings-Grass-Trees-Drives	65.51724	91.80328	83.63636	87.87879	84.44444	72.72727	59.74026	75.32468
Stone-Steel-Towers	89.28571	100.000	100.000	100.000	100.000	100.000	100.000	100.000

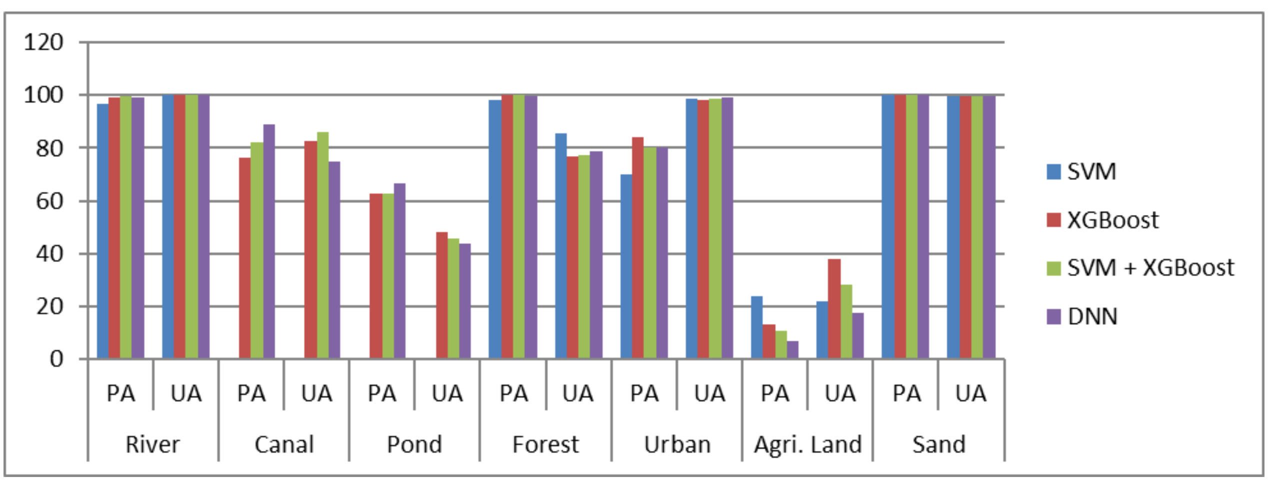


Fig 8. Producer and User Accuracy of Different Classifiers on Twin Cities of Odisha

Table 7 shows a comparison between different Machine learning models and DNN based on Overall accuracy and Kappa Index. This comparison is also represented in Figure 9. As illustrated in Table 7, the ensemble model outperforms other models.

Table 6. Performance comparison between different Machine learning models and DNN based on Overall accuracy and Kappa Index

	Models	Overall Accuracy (OA)	Kappa
Pavia University	XGBoost	0.92	0.889
	SVM+XG Boost	0.954	0.939
	SVM	0.952	0.935
	DNN	0.944	0.926
Indian Pines	XGBoost	0.794	0.762
	SVM+ XGBoost	0.887	0.871
	SVM	0.877	0.859
	DNN	0.876	0.858
Twin Cities of Odisha	XGBoost	0.9353	0.8818
	SVM+ XGBoost	0.9356	0.8824
	SVM	0.9332	0.8742
	DNN	0.9349	0.8806

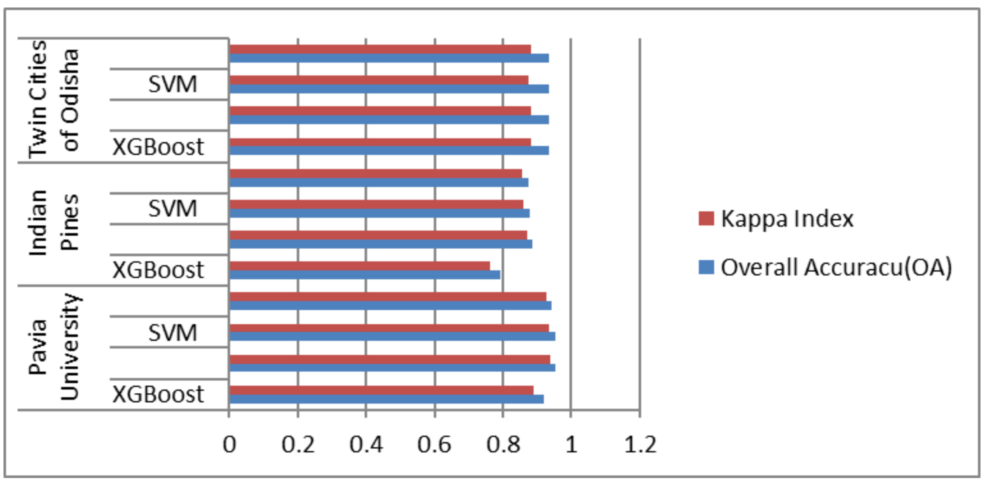


Fig 9. Overall Accuracy and Kappa Index Comparison in Pavia University, Indian Pines and Twin cities of Odisha

3.1 Statistical Analysis

Statistical analysis is done by taking four number of classifiers such as XGBoost, SVM, SVM+ XGBoost, DNN and two performance matrices such as accuracy and Kappa Index. First the rank of each classifier is found out for both the performance metrics on Pavia University dataset, Indian Pines dataset as well as on Twin cities of Odisha. Then the average rank is calculated and finally reranking of average rank is done for different performance metrics. At last, pairwise Nemenyi post hoc test is conducted to compare the ensemble model with other classifier. Rank for accuracy & Kappa index and re ranking of the average ranks for different classifier are depicted in Table 7.

Table 7. Rank for Accuracy and Kappa Index

Different Performance Metrices	XGBoost	SVM	SVM+XG Boost	DNN
Accuracy	4	2	1	3
Kappa Coefficient	4	2	1	3

For pairwise Nemenyi post hoc test [15] we have used critical distance, to compare ensemble model with other model. The formula for critical distance is shown in **equation 1**.

$$CD = Q_{0.05} * \sqrt{\frac{NC(NC+1)}{6ND}} \tag{1}$$

Where,

$Q_{0.05}$: Critical Value = 2.569

NC: No of classifiers = 4

ND: No of datasets = 3

α = 0.05

CD = Critical Distance

Putting the values in equation 1, the critical distance CD = 2.707. After the critical distance is calculated, the pairwise comparison of ensemble model with other classifier using Nemenyi post hoc test is done and is shown in Table 8.

Table 8. Pairwise comparison of ensemble model with other classifier using Nemenyi post hoc test

Classifier Name	Ensemble Model Value					Is ensemble significantly different than A?
	A	**B**	**C=A-B**	**CD**	**C>CD?**	
XGBoost	4	1	3	2.707	Yes	Yes
SVM	2	1	1	2.707	No	No
DNN	3	1	2	2.707	No	No

From the above test, it is clear that ensemble model is significantly different than the XGBoost model. The classified image after the implementation of ensemble model (SVM + XGBoost) on Pavia University and Indian Pines and Twin Cities of Odisha are shown in Figure 10.

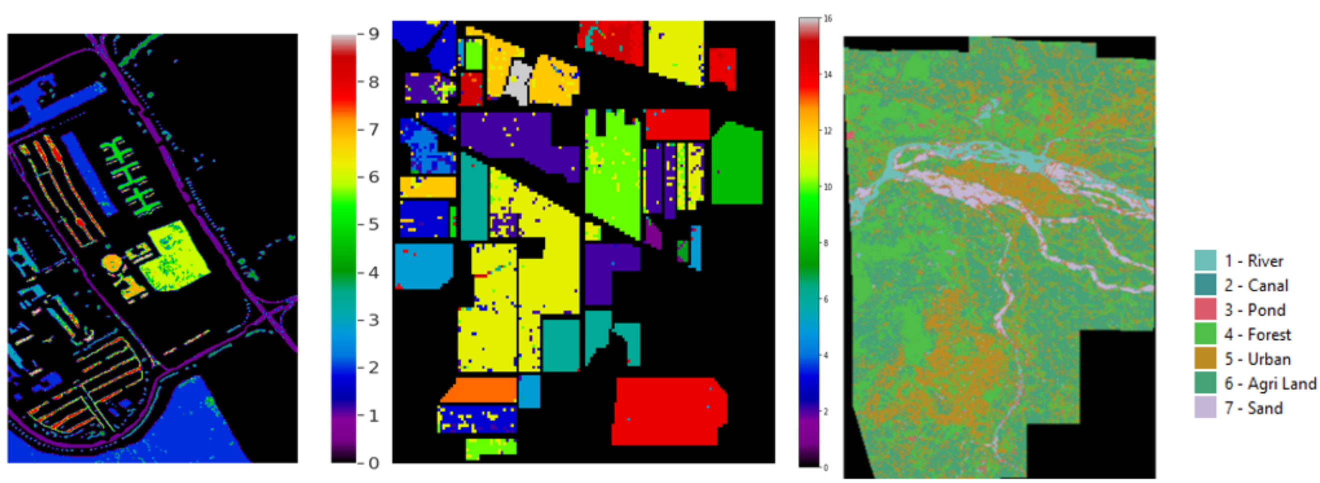


Fig 10. LULC Classified Map Using SVM + XGBoost of (a) Pavia University (b) Indian Pines (c) Twin Cities of Odisha

The results of this study are supported by recent research in LULC classification. Aryal et al. showed that ensemble models like Random Forest and Gradient Boosting improved classification accuracy in urban areas [4]. Similarly, Anam et al. found that Convolutional Neural Networks (CNNs) performed better than traditional machine learning models on Google Earth Engine [8]. Our study aligns with these findings, as the SVM + XGBoost ensemble model gave the best results across all datasets. This confirms that combining models can improve accuracy. Martínez et al. also showed that adding SAR data to ML models helped improve performance in complex landscapes [9]. We extend these findings by using statistical testing (Nemenyi post hoc test) to show that the ensemble model is significantly better than some individual models. In our study, the ensemble model performed well or better than the deep neural network (DNN), especially when training data was limited. Rengma and Yadav also noted that using deep learning in diverse Indian landscapes can be difficult due to variation in data and the need for large training sets[12]. These comparisons show that ensemble models are a strong and practical option and suggest that future work could focus on combining the strengths of both ML and DL approaches.

4 Conclusion

This study presents a comparative evaluation of machine learning (ML), ensemble learning, and deep learning (DL) models for Land Use and Land Cover (LULC) classification using both benchmark hyperspectral datasets (Pavia University, Indian Pines) and real-world satellite imagery (Twin Cities of Odisha, India). The findings highlight that the ensemble model (SVM + XGBoost) achieved the highest classification accuracy, surpassing both individual ML classifiers and the deep neural network (DNN) model. The statistical validation using the Nemenyi post hoc test provides strong evidence that ensemble learning offers a significant performance improvement over standalone models, confirming its robustness in LULC classification tasks. Unlike previous studies that either focus on ML or DL in isolation, this work provides a direct statistical comparison of both approaches, making it one of the first studies to formally validate model effectiveness through statistical significance testing.

The strengths of this research lie in its comprehensive comparative approach, statistical validation, and application to both benchmark and real-world datasets. The study provides clear insights into the practical usability of ML and DL models, demonstrating that ensemble learning methods are particularly effective when labeled training data is limited. Furthermore, the results confirm that deep learning does not always outperform ML-based approaches, especially when ensemble learning is utilized. This finding challenges the assumption that DL is universally superior for remote sensing applications and suggests that hybrid approaches integrating ML and DL could further enhance classification accuracy.

Despite these strengths, the research has certain limitations. The study focused primarily on pixel-based classification, which may not fully capture spatial relationships present in high-resolution imagery. Additionally, computational efficiency and training time were not extensively analyzed, which could be important factors in large-scale remote sensing applications. While the ensemble model performed well, its interpretability remains a challenge compared to traditional ML methods. Another limitation is that the study did not explore transformer-based deep learning models, which have recently gained attention for remote sensing applications.

Future research should focus on integrating object-based classification approaches, incorporating additional spectral indices, and optimizing computational efficiency for large-scale applications. Further exploration of hybrid models that combine ML, DL, and geospatial data fusion techniques could provide new insights into improving LULC classification accuracy. Additionally, domain adaptation techniques could be explored to enhance model generalizability across different geographic regions. Addressing these open questions will further advance the field of LULC classification, making models more adaptable and effective in real-world applications.

References

1) Najjar HA, Kalantar B, Halin AA. Advancements in land cover classification and machine learning techniques for urban areas using remote sensing big data. *Frontiers in Environmental Science*. 2025;13. https://doi.org/10.3389/fenvs.2025.1584485.
2) Paoletti ME, Haut JM, Plaza J, Plaza A. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2019;158:279–317. https://doi.org/10.1016/j.isprsjprs.2019.09.006.
3) Quinn JA, Nyhan MM, Navarro C, et al. Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2018;376. https://doi.org/10.1098/rsta.2017.0363.
4) Aryal J, Sitaula C, Frery AC. Land use and land cover (LULC) performance modeling using machine learning algorithms: a case study of the city of Melbourne, Australia. *Scientific Reports*. 2023;13:13510. https://doi.org/10.1038/s41598-023-40564-0.
5) Li W, Liu H, Wang Y, et al. Deep learning-based classification methods for remote sensing images in urban built-up areas. *IEEE Access*. 2019;7:36274–36284. https://doi.org/10.1109/ACCESS.2019.2903127.
6) Liu B, Yu A, Zuo X, et al. Spatial-spectral feature classification of hyperspectral image using a pretrained deep convolutional neural network. *European Journal of Remote Sensing*. 2021;54:385–397. https://doi.org/10.1080/22797254.2021.1942225.
7) Otukei JR, Blaschke T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*. 2010;12:27–31. https://doi.org/10.1016/j.jag.2009.11.002.
8) Nigar A, Li Y, Baloch MYJ, et al. Comparison of machine and deep learning algorithms using Google Earth Engine and Python for land classifications. *Frontiers in Environmental Science*. 2024;12:1378443. https://doi.org/10.3389/fenvs.2024.1378443.
9) Martinez W, Lizarazo I, Große-Stoltenberg A. Integration of InSAR coherence and SAR backscatter increases accuracy of LULC mapping in tropical high-mountain ecosystems. *Geocarto International*. 2025;40:2451174. https://doi.org/10.1080/10106049.2025.2451174.
10) Kalita I, Roy M. Deep neural network-based heterogeneous domain adaptation using ensemble decision making in land cover classification. *IEEE Transactions on Artificial Intelligence*. 2020;1:167–180. https://doi.org/10.1109/TAI.2020.3043724.

11) Carranza-Garcia M, Garcia-Gutiérrez J, Riquelme JC. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*. 2019;11:274. https://doi.org/10.3390/rs11030274.
12) Rengma NS, Yadav M. Generation and classification of patch-based land use and land cover dataset in diverse Indian landscapes: a comparative study of machine learning and deep learning models. *Environmental Monitoring and Assessment*. 2024;196:568. https://doi.org/10.1007/s10661-024-12719-7.
13) Tahir Z, Haseeb M, Mahmood SA, et al. Predicting land use and land cover changes for sustainable land management using CA-Markov modelling and GIS techniques. *Scientific Reports*. 2025;15:3271. https://doi.org/10.1038/s41598-025-87796-w.
14) Zhao S, Tu K, Ye S, et al. Land use and land cover classification meets deep learning: A review. *Sensors*. 2023;23:8966. https://doi.org/10.3390/s23218966.
15) Carranza-García M, García-Gutiérrez J, Riquelme JC. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*. 2019;11:274. https://doi.org/10.3390/rs11030274.