“ASSESSING DEFORESTATION USING SATELLITE IMAGERY IN DEHRADUN DISTRICT”

MAJOR PROJECT REPORT

Submitted in partial fulfillment of the
requirementsfor the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING
(Artificial Intelligence & Machine Learning)

BY

KESHAV (2220804)
SHOURYA KHANKA (2100852)
ANKITA KOTNALA (2220802)
ARNAV SIRKHAL (2100863)

Under the guidance of

MR. RAMESH KUMAR
(Assistant Professor)
Department Of Computer Science & Engineering

G.B. Pant Institute of Engineering & Technology

Pauri Garhwal, Uttarakhand Session 2024-25

CANDIDATE DECLARATION

We as a result of this declare that the project work entitled "Accessing deforestation using satellite imagery in dehradun district" in partial fulfillment of the requirements for the award of the Degree of **BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING with specialization in ARTIFICIAL INTELLIGENCE & MACHINE LEARNING** submitted to the Department of Computer Science & Engineering, G.B. Pant Institute of Engineering & Technology, Pauri, Uttarakhand, is an authentic record of our work carried out during a period from August 2024 to December 2024 under the supervision of Mr. Ramesh Kumar, Assistant Professor, Department of Computer Science and Engineering.

We have not submitted the matter presented in this project for the award of any other degree of this or any other university.

Student Name	College Id
Keshav (Leader)	2220804
Shourya Khanka	2100852
Arnav Sirkhal	2100863
Ankita Kotnala	2220802

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date: **26/06/2025**

Mr. Ramesh Kumar
(Project Supervisor)

Mr Ashwini Kumar Saini
(Project Coordinator)

Computer Science & Engineering
G.B. Pant Institute of Engineering & Technology, Pauri Garhwal (Uttarakhand)

Mr Ramesh Kumar
(Project Coordinator)

Computer Science & Engineering
G.B. Pant Institute of Engineering & Technology, Pauri Garhwal (Uttarakhand)

Dr Bhumika Gupta
(H.O.D.)

Computer Science & Engineering
G.B. Pant Institute of Engineering & Technology, Pauri Garhwal (Uttarakhand)

CERTIFICATE

Date 26/06/2025

We hereby certify that the work which is being presented in this report entitled "Accessing deforestation using satellite imagery in dehradun district", in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology, is ban authentic record of work by us. The matter embodied in this work has not been submitted to any either University / Institute for the award of any degree.

KESHAV
SHOURYA KHANKA
ARNAV SIRKHAL
ANKITA KOTNALA

This is to certify that the above statement made by the above students is correct to the best of my knowledge.

Mr. RAMESH KUMAR
(Assistant Professor, GBPIET, CSED)

The Bachelor of Technology (B. Tech) viva-voce examination of the above students has been held on 26/06/2025.

(Signature of the Supervisor) (Signature of the Project coordinator)

(Signature of the HoD)

ACKNOWLEDGMENT

We wish to express our deep gratitude to our guide Mr. Ramesh Kumar for all the advice, encouragement, and constant support he has given us throughout our project work. This work would not have been possible without his support and valuable suggestions.

We sincerely thank our respected Project Coordinator of the department Mr. Ramesh Kumar, Department of Computer Science & Engineering, for her great support in coordinating project schedules, providing resource equipment and information, and liaising with group members to define project requirements, scope, and objectives.

We are also grateful to Dr. Ashish Negi, Head of the Department of Computer Science & Engineering, for providing us with the necessary facilities to complete our project work.

We would like to thank all our friends for their help and constructive criticism during our project work. Finally, we have no words to express our sincere gratitude to our parents, who have shown us this world and provided us with every support.

Student Name	**College Id**
Keshav	2220804
Shourya Khanka	2100852
Arnav Sirkhal	2100863
Ankita Kotnala	2220802

ABSTRACT

Deforestation threatens global ecosystems, biodiversity, and climate stability. Traditional monitoring methods are labor-intensive and limited. This project leverages satellite imagery and deep learning for efficient deforestation analysis. Using multi-temporal satellite images, the project captures dynamic changes in forest cover. The process includes data preprocessing, model training, utilizing imagery from Sentinel datasets. A deep learning model distinguishes between forested and non-forested areas and identifies land cover changes indicating deforestation. It also highlights at-risk regions, facilitating proactive conservation and policy interventions. Results show that integrating satellite imagery with deep learning is effective for deforestation monitoring, offering high accuracy, scalability, and timeliness.

The methodology involves pre-processing satellite images, training a deep learning model to distinguish between forested and non-forested regions, and detecting areas exhibiting significant land cover changes. The analysis further pinpoints high-risk zones, enabling targeted conservation efforts. Results confirm the effectiveness of integrating satellite data with artificial intelligence, offering a scalable and precise approach for deforestation assessment. The project supports early detection and comprehensive analysis of deforestation trends, aiding in forest preservation and climate change mitigation. It demonstrates the potential of combining satellite technology with deep learning to enhance environmental monitoring, sustainable forest management, and the protection of natural resources for future generations. This project underscores the value of technological innovation in environmental monitoring. By providing actionable insights into deforestation trends, it contributes to sustainable forest management, biodiversity conservation, and climate change mitigation, fostering a healthier environment for future generations.

TABLE OF CONTENTS

S. No.	Contents	Page No.
1	INTRODUCTION	01
2	LITERATURE REVIEW	02
3	PROBLEM STATEMENT	04
4	METHODOLOGY	05
5	SYSTEM REQUIREMENTS	11
6	WORK CARRIED OUT AND FUTURE REVIEW	12
7	RESULT AND DISCUSSION	13
8	CONCLUSIONS	23
9	FUTURE WORK	24
10	REFERENCES	25

TABLE OF FIGURES

Figure No.	Title	Page No.
Figure 4.1	All four bands	06
Figure 4.2	The step we follow to extract Dehradun shape file	07
Figure 4.3	Use of shape file to extract Dehradun from satellite image	09
Figure 7.1	The classification of K-means 2015 & 2020	13
Figure 7.2	The classification of ISO clustering 2015 & 2020	14
Figure 7.3	The classification of MLC & RF of 2015	15
Figure 7.4	The classification of MLC & RF of 2020	15
Figure 7.5	The classification of MLC 2015 and area cover	16
Figure 7.6	The classification of MLC 2018 and area cover	16
Figure 7.7	The classification of MLC 2021 and area cover	17
Figure 7.8	The classification of MLC 2024 and area cover	17
Figure 7.9	The classification of Semi-Supervised 2015 and area cover	18
Figure 7.10	The classification of Semi-Supervised 2018 and area cover	18
Figure 7.11	The ground truth from Global Forest Watch 2015–2025	19
Figure 7.12	The line chart for MLC & Semi-Supervised over time	20
Figure 7.13	The area chart for MLC area cover	20
Figure 7.14	The line chart for MLC & Semi-Supervised difference	21
Figure 7.15	The bar chart for MLC & Semi-Supervised accuracy	21
Figure 7.16	The bar chart for MLC & Semi-Supervised accuracy summary	22

TABLE OF SCREENSHOT

Screenhot No.	Title	Page No.
7.1	The MLC 2015 area cover	16
7.2	The MLC 2018 area cover	16
7.3	The MLC 2021 area cover	17
7.4	The MLC 2024 area cover	17
7.5	The STC 2015 area cover	18
7.6	The STC 2024 area cover	18
7.7	The ground truth from global forest watch 2015-2025	19
7.8	The line chart for MLC & Semi-Supervised over time	20
7.9	Area chart for MLC over time	20
7.10	The line chart for MLC & Semi-Supervised difference	21
7.11	The bar chart for MLC & Semi-Supervised accuracy	21
7.12	The bar chart for MLC & Semi-Supervised accuracy summary	22

INTRODUCTION

Deforestation is a critical environmental issue that significantly impacts biodiversity, climate change, and human livelihoods. Accurate mapping and monitoring of deforestation are essential for effective forest management and conservation strategies. This project aims to develop a comprehensive deforestation mapping approach focusing on Dehradun, a district in India known for its rich forest cover. The primary objective is to analyze changes in forest cover over a three-year period, from 2015 to 2024 in the interval of 3 years, using remote sensing and advanced geospatial techniques. Our project began with selecting Dehradun as the study area due to its ecological significance and the availability of high-quality remote sensing data. The initial step involved downloading the India map and extracting the Dehradun shapefile using ArcGIS software. This shapefile was then uploaded to the Copernicus Open Access Hub to acquire Sentinel-2 satellite imagery for the months of October to December 2015, covering the entire Dehradun district across three tiles.

The downloaded Sentinel-2 data, specifically the MSI files, were processed using the SNAP software for atmospheric correction to ensure accuracy and consistency in the satellite imagery. Subsequently, these corrected images were imported into ArcGIS software for further analysis. We composited bands 2, 3, 4, and 8, which have a 10-meter resolution, to create a detailed and high-resolution image of the study area. These composite images were then mosaicked to combine the three tiles into a single seamless tile.

Using the Dehradun shapefile, we performed masking on this mosaicked tile to extract the area of interest, resulting in a raster file in .tif format. For the classification of land cover types, we employed both unsupervised and supervised classification methods. Initially, we used ISODATA clustering and K-means algorithms in Python and ArcGIS to perform unsupervised classification. Subsequently, training data were created and used to implement the Random Forest algorithm and Maximum Likelihood Classification (MLC) for supervised classification and Self Training Classifier for semi-supervised classification.

The classified output included six distinct classes: Evergreen Forest, Deciduous Forest, Cropland, Fallow Land, Water Bodies, and Developed Area. This classification provided a detailed land cover map for Dehradun for the year 2018,2021,2024. In the previous phase of the project, we have replicated this process for Sentinel-2 data from October to December 2015. This has allowed us to compare the datasets and observe changes in land cover over the three-year period.

The subsequent sections of this report will present the detailed methodology, results, and analysis for each classification technique, with a focus on the 2018,2021 and 2024 data. The previous semester has involved similar processing of the 2015 data to map deforestation trends and changes in Dehradun's land cover. This project not only highlights the importance of remote sensing in environmental monitoring but also provides valuable insights for forest conservation and management in the region

LITERATURE REVIEW

2.1. Dr. Rashmi Saini et el. "Crop Classification On Single Data Sentinel-2 Imagery"

Mapping crops using satellite imagery is challenging due to field complexities and the spectral similarity among crops. Sentinel-2, with its thirteen spectral bands, high resolution (10m, 20m, 60m), rapid revisit time, and free data availability, is well-suited for vegetation mapping. This study classifies crops in Roorkee, Uttarakhand, using Sentinel-2 imagery and two machine learning algorithms: Random Forest (RF) and Support Vector Machine (SVM).Using four spectral bands (Near Infrared, Red, Green, and Blue), the study achieved overall classification accuracies of 84.22% for RF and 81.85% for SVM. RF outperformed SVM by 2.37% in overall accuracy. High-Density Forest was classified with the highest accuracy, while Fodder showed the lowest due to spectral overlap with Wheat. The findings highlight Sentinel-2's utility in vegetation mapping and RF's superior performance for accurate crop classification.

2.2. Battude M et el. "Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2A like remote sensing data. Remote Sens Environ."

High-resolution remote sensing data, like that from Sentinel-2A, provides valuable insights into agricultural productivity. Battude et al. (2016) demonstrated the potential of Sentinel-2A-like data to estimate maize biomass and yield over large areas, emphasizing its high spatial and temporal resolution. Their work highlights the importance of advanced remote sensing technologies in agricultural monitoring, enabling accurate predictions critical for resource management and planning.

2.3. Dr. Rashmi Saini et el. "Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery"

Crop mapping is challenging due to the spectral similarity of various crops. This study focuses on identifying major crops in Roorkee, India, using Sentinel-2A data and evaluating the performance of ensemble methods—Extreme Gradient Boosting (Xgboost), Adaboost.M1, Stochastic Gradient Boosting (SGB), and Random Forest (RF)—against Support Vector Machine (SVM) for crop classification. Xgboost achieved the highest overall accuracy of 86.91%, followed by RF, while SVM showed the lowest accuracy. McNemar's test confirmed significant performance differences among classifiers. Wheat and sugarcane were classified with maximum accuracies of 88.04% and 85.95%, respectively. Red-Edge2, Red-Edge3, and NIR bands were identified as the most important predictors, while Red-Edge1 had the least impact. The study highlights Xgboost's strong potential for crop classification in heterogeneous agricultural environments, suggesting further exploration in future research.

2.4. Belgiu M, Dragu et el. "Machine Learning in Remote Sensing"

The integration of machine learning into remote sensing workflows has opened new possibilities for data analysis. Random Forest, a widely used algorithm, has been extensively reviewed by Belgiu and Drăgu (2016), who outlined its versatility across various applications in remote sensing. They also identified future directions, including improved computational efficiency and application in emerging sensor technologies.

2.5. Dr. Camps-Valls and Bruzzone et el. "Kernel-Based Methods in Hyperspectral Image Classification"

Kernel-based methods have proven to be highly effective for hyperspectral image classification due to their ability to handle high-dimensional data and nonlinear relationships. Camps-Valls and Bruzzone (2005) explored the use of these methods, particularly support vector machines (SVM) with kernel functions, for hyperspectral image classification. They highlighted the advantages of kernel methods in extracting complex patterns from hyperspectral data, emphasizing their robustness in dealing with spectral variability and noise. Their work has been foundational in advancing machine learning applications in hyperspectral remote sensing.

2.6. Chan and Paelinckx et el. "Ensemble Learning for Ecotope Mapping."

Tree-based ensemble methods, such as Random Forest and Adaboost, are powerful tools for classification tasks in remote sensing. Chan and Paelinckx (2008) evaluated these methods for ecotope mapping using airborne hyperspectral imagery. Their study demonstrated that ensemble learning algorithms not only enhance classification accuracy but also facilitate effective spectral band selection, which is crucial for optimizing hyperspectral data analysis. This work underscores the importance of ensemble methods in managing high-dimensional datasets, making them a critical asset in environmental monitoring and management.

Key findings include:

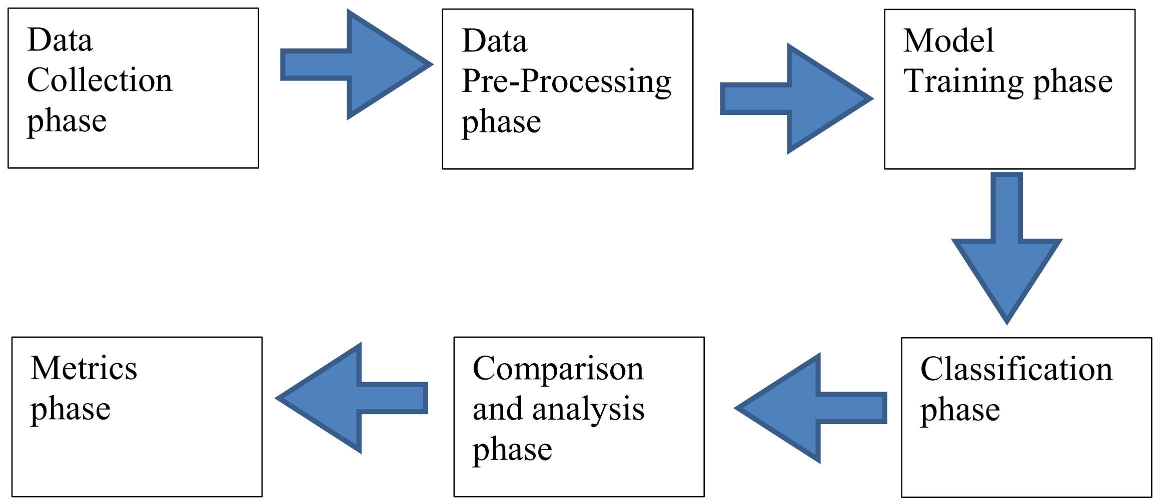
1. **Single-Date Data:** Reliance on single-date imagery may not account for temporal variability in crop phenology.
2. **Future Integration:** Emerging sensor technologies need more exploration for integration with existing machine learning workflows.
3. **Crop-Specific Focus:** Methodology is crop-specific (focused on maize), limiting its applicability to diverse crop types
4. **Spectral Overlap:** Low classification accuracy for certain crops (e.g., Fodder) due to spectral similarity with other crops like Wheat.

PROBLEM STATEMENT

His project, titled **"Assessing Deforestation Using Satellite Imagery in Dehradun District,"** is focused on analyzing forest cover changes using Sentinel-2 satellite imagery over four time intervals: 2015, 2018, 2021, and 2024. The primary objective is to detect and assess deforestation trends in the region using advanced classification techniques. By evaluating these multi-temporal satellite datasets, the project aims to observe how land cover—particularly forest areas—has evolved over time and to measure the extent of deforestation.

Through the comparative analysis of forest cover during these selected years, the study seeks to pinpoint specific regions most affected by forest loss, understand the underlying causes such as urbanization or agricultural expansion, and evaluate the broader environmental impacts. The insights gained will help inform policy recommendations and strategic interventions for forest conservation. Ultimately, this work contributes to improved forest monitoring and supports efforts toward sustainable land use and environmental management in the Dehradun district.

METHODOLOGY



The methodology section outlines the step-by-step approach used to conduct the deforestation mapping project in the Dehradun district. It provides a detailed description of data acquisition, preprocessing, analysis, and classification techniques employed in the study.

4.1 Data Collection

The primary dataset for this study was acquired from Sentinel-2A and Sentinel-2B satellites, accessed via the Copernicus Open Access Hub. The study focused on the Dehradun district in Uttarakhand, India, selected due to its noticeable land cover transformations and high rate of deforestation [14]. Satellite imagery was collected for four timeframes—2015, 2018, 2021, and 2024

4.1.1 ***India Map and Dehradun Shapefile***: Download India map and extract Dehradun shapefile using ArcGIS software.

4.1.2 ***Sentinel-2 Satellite Imagery***: Access Sentinel-2 satellite imagery for October to December 2015 and 2024 from the Copernicus Open Access Hub, covering Dehradun district across multiple tiles.

4.2 Data Pre-Processing

Atmospheric Correction to Sentinel-2 MSI files using SNAP software to enhance data accuracy and consistency. By removing all the clouds cover enhance bright image, fog correction.

Image Composition: Composite bands 2, 3, 4, and 8 to create high-resolution images with a 10-meter spatial resolution.

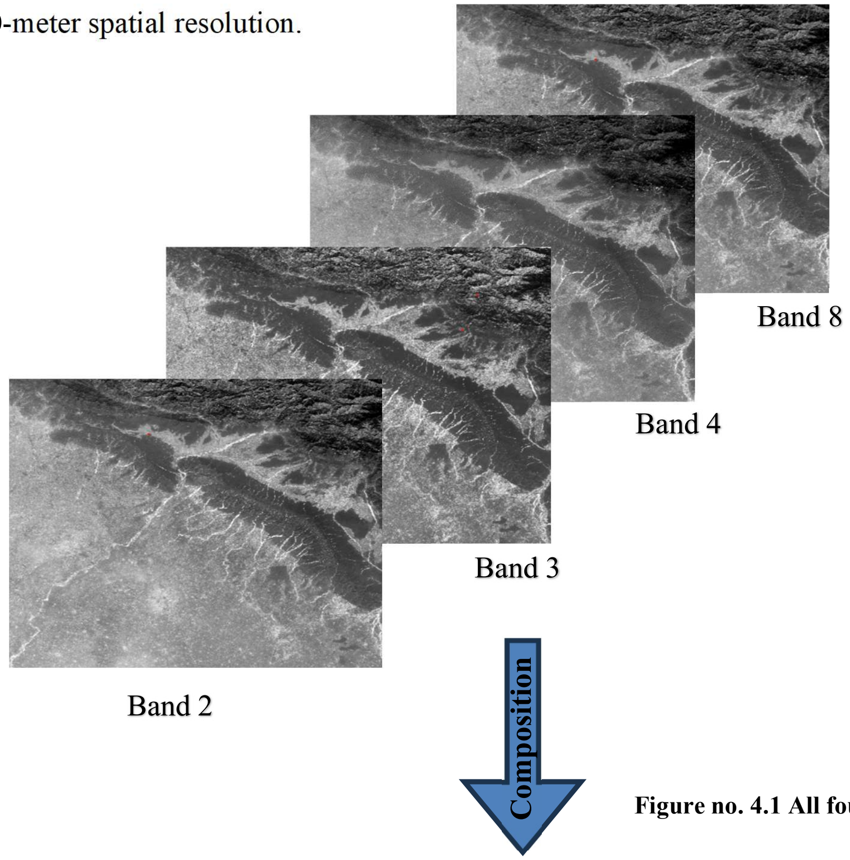
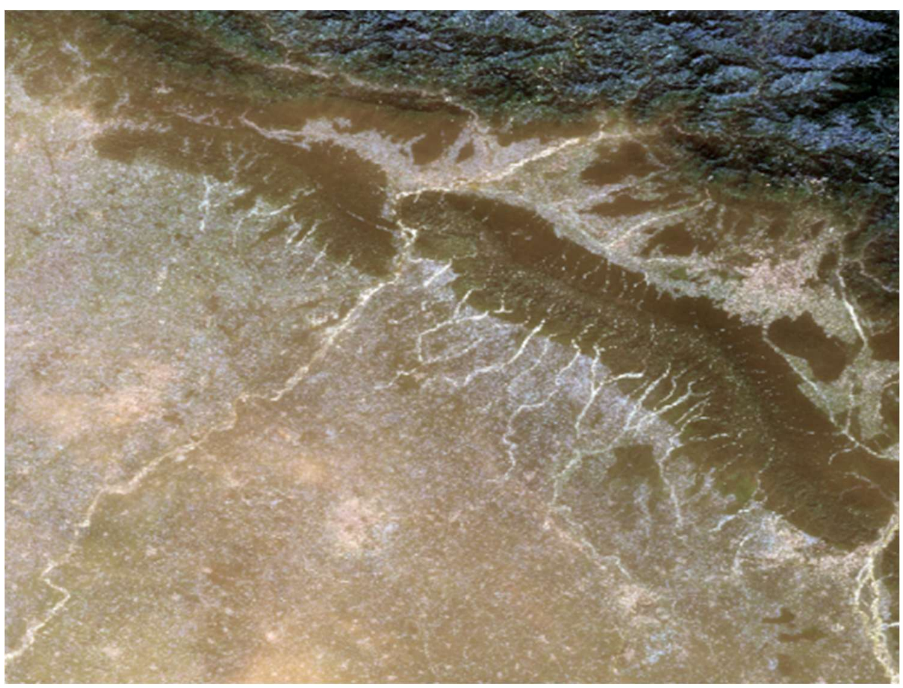


Figure no. 4.1 All four bands



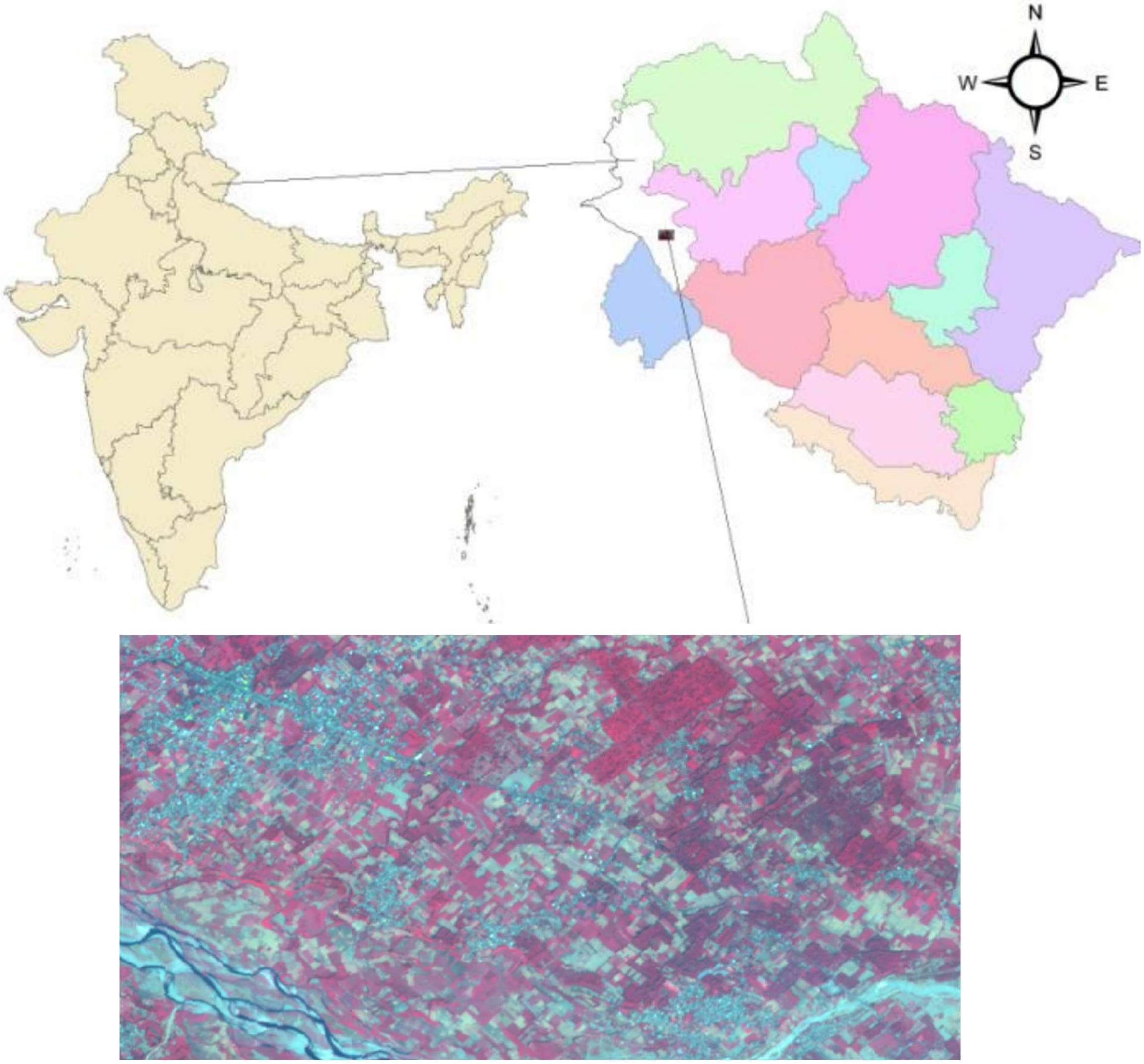


Figure: 4.2 The step we follow to extract Dehradun shape file

4.3 Model train phase

Model training is the fundamental process in machine learning and artificial intelligence where a model learns to make predictions or decisions based on data. During training, the model is exposed to a large dataset that includes input features and corresponding labels (in supervised learning). Through a process of forward prediction and error correction—usually by minimizing a loss function using optimization techniques like gradient descent—the model adjusts its internal parameters (such as weights in a neural network) to improve its accuracy. The goal is to enable the model to generalize well to new, unseen data, not just memorize the training examples.

4.3.1. Random Forest (RF) Classification:

Random Forest builds a 'forest' of many decision trees during training. Each tree is trained on a random subset of the data and a random subset of features. Utilizes bootstrapping (random sampling with replacement) to create subsets of the training data.

Aggregation of the results from multiple trees reduces overfitting and increases generalization. Can provide an estimate of feature importance, indicating which variables are most influential in the classification process.

4.3.2. Maximum Likelihood Classification (MLC):

MLC calculates the probability that a pixel belongs to each class based on the statistical characteristics (mean and covariance) of training data. The pixel is assigned to the class with the highest probability. Assumes that the pixel values for each class follow a multivariate normal distribution. This means MLC works best when this assumption holds true.

Requires a set of representative training data for each land cover class. The quality and representativeness of the training data significantly influence the classification accuracy.Produces a classified image where each pixel is assigned to three of the predefined classes. Additionally, it can provide a probability map indicating the confidence level of the classification.

4.3.3. Self-Training Classifier (STC):

Self-training is a semi-supervised classification technique that starts with a small amount of labeled data and a large amount of unlabeled data. It iteratively improves the classifier by using its own most confident predictions on the unlabeled data to retrain itself.Initially, a base classifier (e.g., decision tree, SVM, or even MLC) is trained using the limited labeled training data. The classifier then predicts labels for the unlabeled data. A subset of these predictions — typically the ones with the highest confidence — is selected and added to the training set with their predicted labels. The classifier is retrained on the expanded training set, and this process is repeated for several iterations.

Self-training assumes that the classifier's high-confidence predictions are likely to be correct and that adding them to the training set will enhance learning. It also assumes that the decision boundary can be refined by iteratively expanding the labeled dataset using confidently predicted samples.Requires a small set of labeled data to initiate the training process and a large pool of unlabeled data. The initial labeled data should be representative, but STC can still work with less comprehensive labeled datasets due to its iterative learning process.Produces a classified image where each pixel is assigned to one of the predefined classes. As the model refines itself over iterations, classification accuracy may improve, especially in areas where labeled data was initially sparse. Probability or confidence maps can also be generated to evaluate the reliability of the classifications.

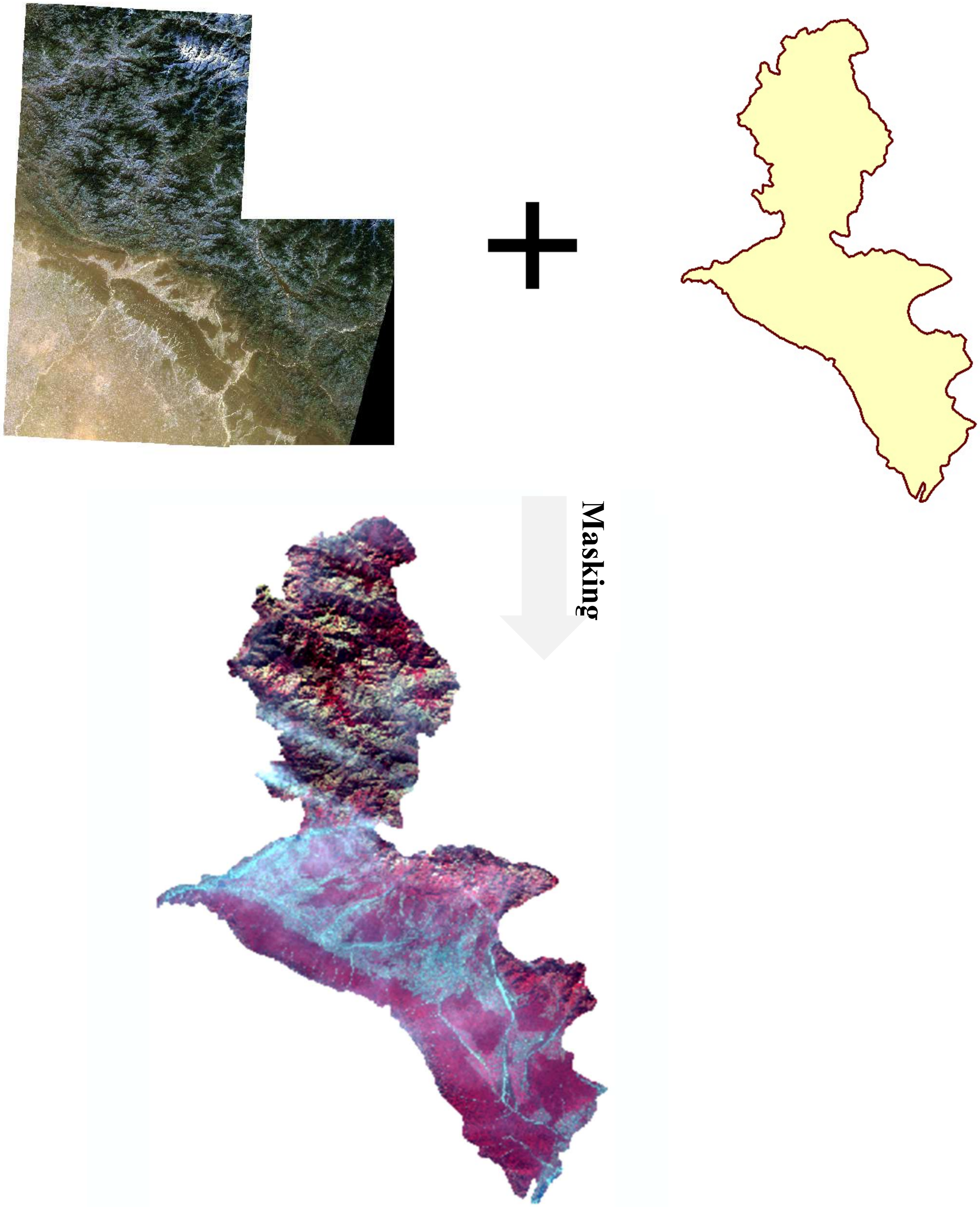


Figure: 4.3 Use of shape file to extract Dehradun form satellite image

4.4 Classification phase

The classification phase transforms raw satellite imagery into meaningful land cover information, enabling the identification, categorization, and quantification of different landscape features. In this study, classification was performed using a combination of **unsupervised**, **supervised**, and **semi-supervised** approaches on Sentinel-2 imagery, specifically employing Bands 2 (Blue), 3 (Green), 4 (Red), and 8 (NIR) due to their high spectral fidelity and 10-meter spatial resolution.

4.4.1. ***Unsupervised Classification***: Implement ISODATA clustering and K-means algorithms in Python to perform unsupervised classification and identify spectral clusters representing different land cover types.

4.4.2. ***Supervised Classification:*** Create training data and use Random Forest algorithm and Maximum Likelihood Classification (MLC) to classify land cover types into categories such as Evergreen Forest, Deciduous Forest, Cropland, Fallow Land, Water Bodies, and Developed Area.

4.4.3. ***Semi-Supervised Classification:*** Implement a Self-Training Classifier in Python using a small set of labeled training samples and a large pool of unlabeled data. Use a base learner (e.g., Random Forest or MLC) to iteratively label high-confidence unlabeled data, retraining the model over multiple iterations. Classify land cover types into predefined categories: Evergreen Forest, Deciduous Forest, Cropland, Fallow Land, Water Bodies, and Developed Area.

4.5 Comparison and Analysis phase

Compare the classified land cover maps of 2015, 2018, 2021 and 2024 to identify and analyze changes in forest cover and land use over the fi-year period.

4.6 Metrics phase

Results, and analysis for each classification technique by calculating land cover area.Comparing both land cover area with model and predict the accuracy by the referenced data. Present the comparative study of deforestation from 2015 to 2024 in 3 year of intervals.

SYSTEM REQUIREMENTS

The system requirements outline the hardware, software, and data prerequisites necessary to execute the deforestation mapping project in the Dehradun district. These requirements ensure the compatibility and functionality of the project workflow, from data acquisition to analysis and classification.

Hardware Requirements

1. **Computer**: A desktop or laptop computer with sufficient processing power and memory to handle data-intensive tasks.
2. **Storage**: Minimum 50 GB storage space to store large satellite imagery files, intermediate data products, and classification results.
3. **Memory**: Minimum 32 GB to perform classification.
4. **Graphics Card (Optional)**: A dedicated graphics processing unit (GPU) may be beneficial for accelerating image processing and analysis tasks.

Software Requirements

1. **GIS Software**:
   ArcGIS Desktop or ArcGIS Pro for data preprocessing, spatial analysis, and masking tasks. QGIS as an alternative open-source GIS software option
   .
2. **Remote Sensing Software**:
   Sentinel Application Platform (SNAP) for atmospheric correction of Sentinel-2 imagery.

3. **Programming Environment**:
   Python programming language with libraries such as NumPy, pandas, scikit-learn, matplotlib, rasterio and arcpy for implementing classification algorithms and automation scripts.

4. **Statistical Software**:
   Statistical analysis software like R or MATLAB for advanced data analysis and model validation.

WORK CARRIED OUT AND FUTURE REVIEW

6.1. Work Carried Out

Deforestation Analysis for 2015, 2018, 2021, 2024 Data: Detailed analysis performed to identify deforestation patterns and hotspots. Utilized tools, e.g., GIS, remote sensing, etc. We got land cover classified data for analysis purpose and the area covered by those classes.

Data Comparison and Interpretation*:* Conducted a comparative analysis between 2015,2018,2021 and 2024 datasets. Highlighted differences and identified trends in deforestation.

6.2. Future Review

Long-Term Trend Analysis Extend the study to include data from additional years for a broader trend analysis. **Investigation of Underlying Causes** Explore potential socio-economic, environmental, factors contributing to observed changes. **Impact Assessment** Assess the environmental and ecological impact of the deforestation trends identified **Actionable Insights for Sustainable Practices** Develop recommendations for deforestation mitigation and sustainable land use planning.

RESULT & DISCUSSION

The MLC, Random Forest algorithm was employed to classify the pre-processed and masked Sentinel-2 imagery. The training data, which included representative samples of each land cover type, facilitated the supervised classification process. The classification output was generated in the form of a raster image with each pixel assigned to three of the six land cover classes.

7.1 Land Cover Classes

Land cover classification provides a fundamental understanding of the Earth's surface characteristics, essential for environmental monitoring, resource management, and land use planning. The primary land cover classes identified in this study include:

7.1.1. ***Evergreen Forest:*** Represents areas with dense, year-round foliage. These areas are critical for maintaining biodiversity and ecological balance.
7.1.2. ***Deciduous Forest:*** Includes regions with trees that shed leaves seasonally. These forests play a vital role in carbon sequestration and habitat provision.
7.1.3. ***Cropland:*** Denotes agricultural lands used for crop production. These areas are essential for food security and local economies.
7.1.4. ***Fallow Land:*** Areas left uncultivated to restore soil fertility. This classification helps in understanding agricultural practices and land use changes.
7.1.5. ***Water Bodies:*** Includes rivers, lakes, and reservoirs. These are vital for hydrological studies and water resource management.
7.1.6. ***Developed Area:*** Represents urban and built-up regions. This classification is crucial for urban planning and infrastructure development.

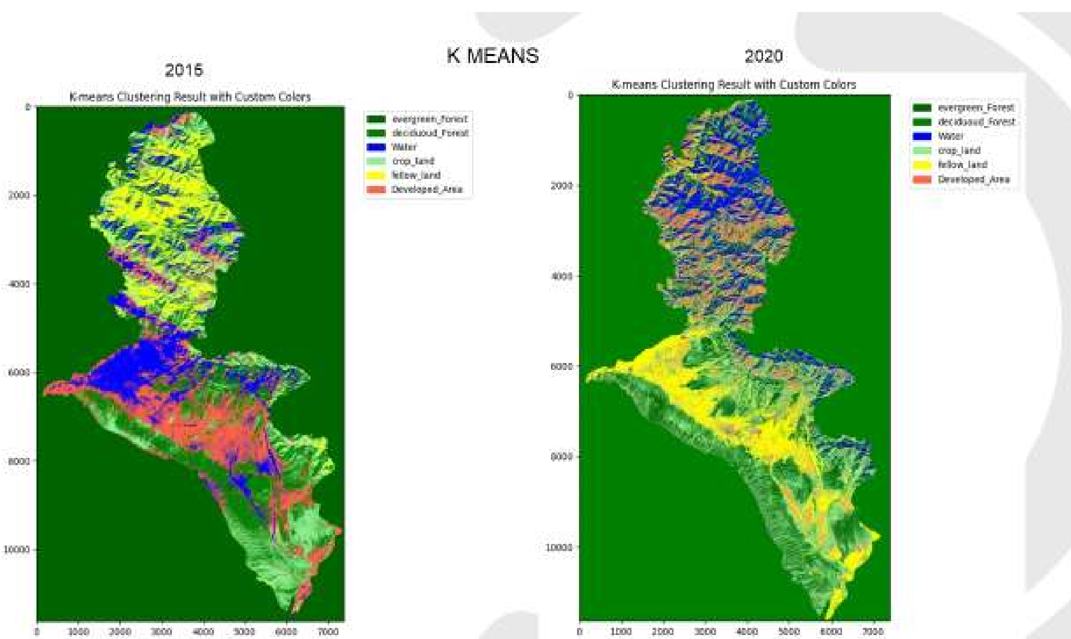


Figure: 7.1 The classification of K-means 2015 & 2020

Below is the classification of 2015 and 2024 raster image using k means.

The above K-Means clustering results for the years 2015 and 2020 show clear spatial patterns but reveal certain limitations in unsupervised classification. In the 2015 map, we observe widespread mixing of crop land (yellow), fallow land (orange), and developed areas (red), particularly in the southern region. The 2020 result shows an unrealistic expansion of evergreen forest (green) and a homogenization effect where distinct land cover types appear overly merged. The inconsistent patterns across both years suggest that K-Means struggled to distinguish certain land cover classes accurately, making it less suitable for reliable forest change detection.We did unsupervised classification using ISO clustering and calculated the area matrix

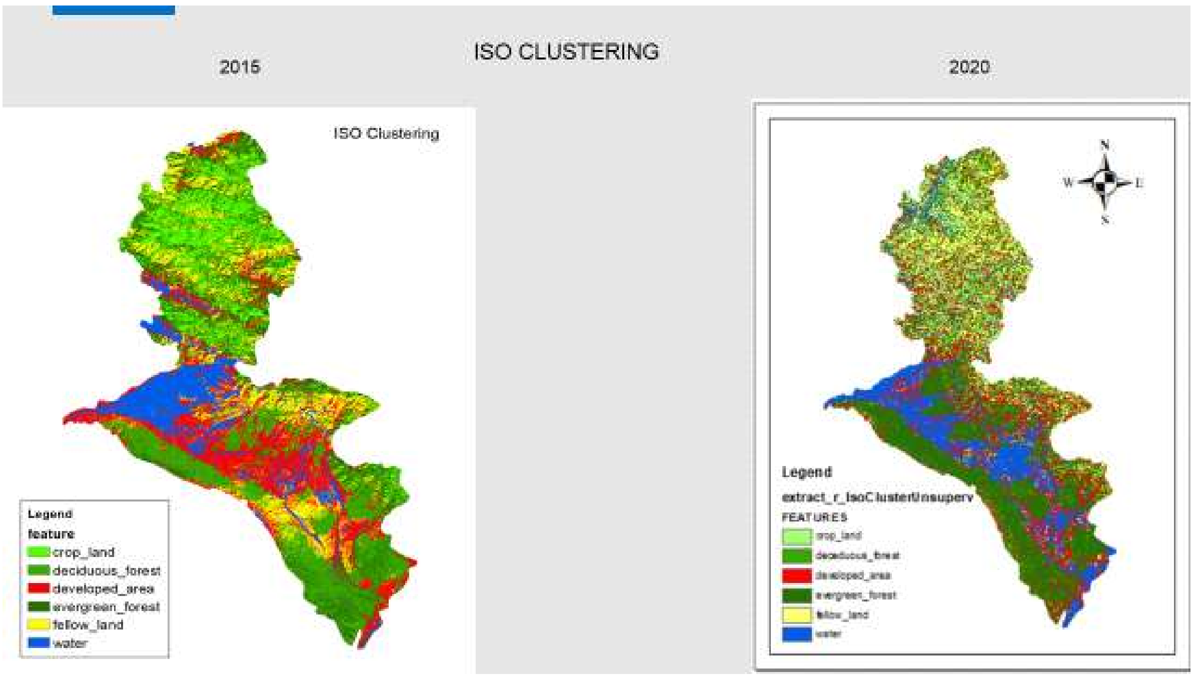


Figure: 7.2 The classification of ISO clustering 2015 & 2020

The ISO DATA clustering outputs similarly reflect the challenges of unsupervised learning in complex environments like Dehradun. In 2015, the classification appears more detailed, but there is noticeable misclassification of crop land and developed areas, which are scattered irregularly. By 2020, the land cover classes appear overly fragmented, and essential distinctions between forest types and other land uses become blurred. The random distribution of features in the 2020 map especially highlights the inconsistency and instability of ISODATA results over time.

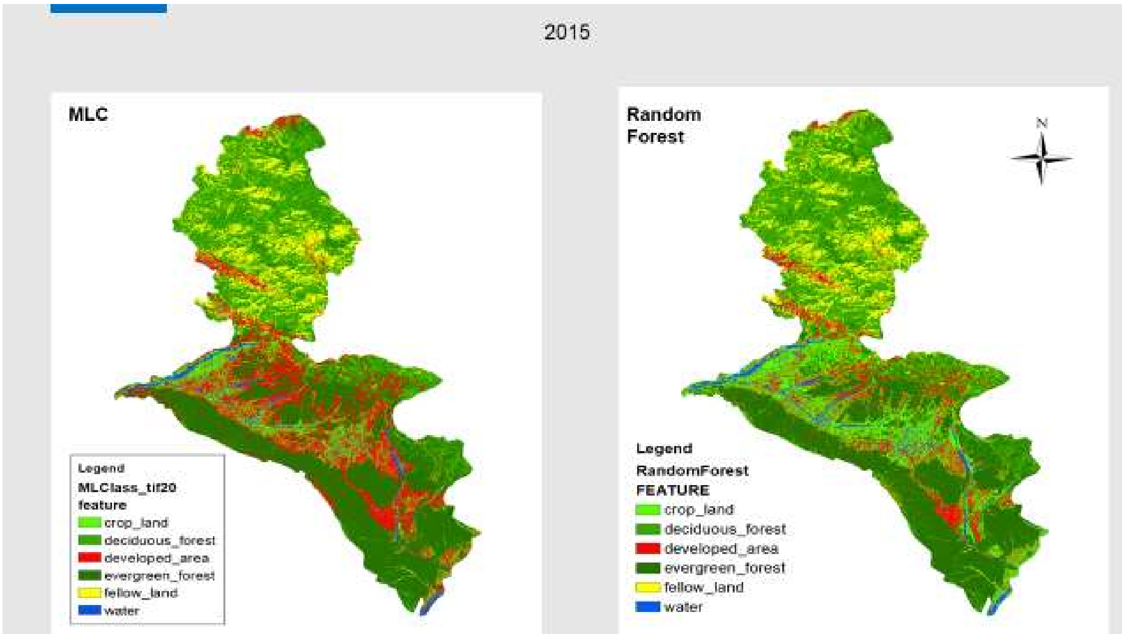


Figure: 7.3 The classification of MLC & RF of 2015

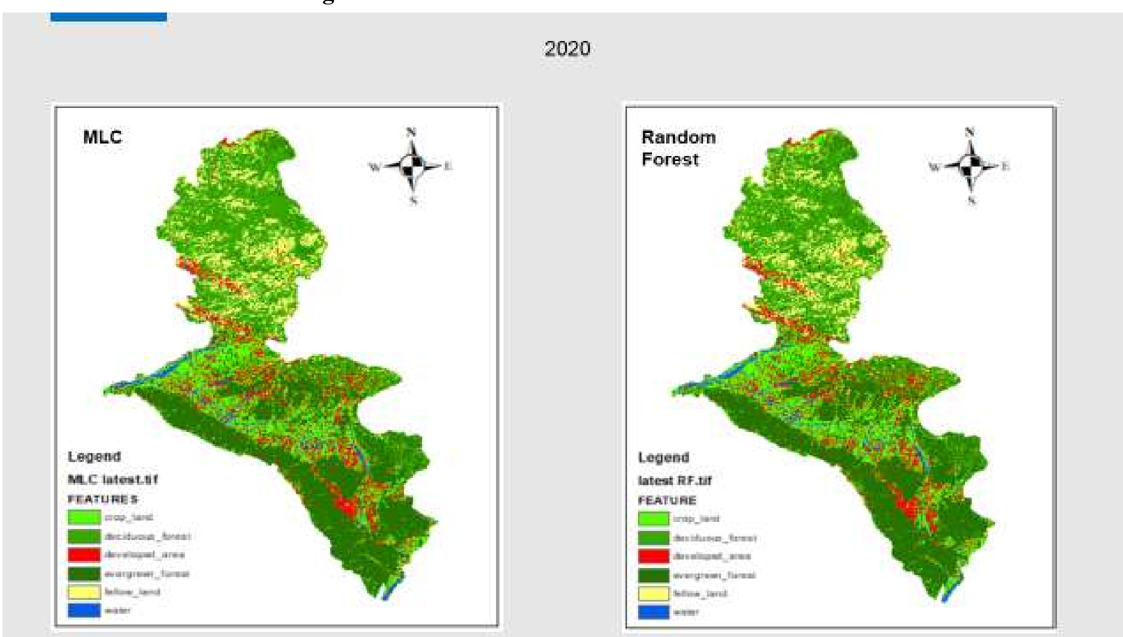


Figure: 7.4 The classification of MLC & RF of 2020

The comparison between MLC and Random Forest classification methods for 2015 highlights the superior performance of MLC in preserving land cover boundaries. The MLC result presents a well-differentiated distribution of evergreen and deciduous forests, crop land, and urban areas. On the other hand, the Random Forest classification also captures forest regions effectively but shows more noise and patchiness in built-up areas and fallow land. Overall, MLC provides cleaner and more interpretable output, supporting its selection as the primary method in the study.

COMPARATIVE ANALYSIS

In 2015, the dominant land cover was **evergreen forest** with an area of approximately **978.397 km²**, followed closely by **deciduous forest** at **842.428 km²**. **Crop land** accounted for **526.059 km²**, while **fallow land** contributed **322.291 km²**. **Developed areas** were recorded at **317.723 km²**, and **water bodies** at **83.767 km²**. The distribution indicates a predominantly forested region with significant agricultural use and moderate urban expansion.

Value	Count	FEATURES	area(KM2)
1	3177239	developed_area	317.723
2	837678	water	83.767
3	9783970	evergreen_forest	978.397
4	8424283	deciduous_forest	842.428
5	3222916	fellow_land	322.291
6	5260591	crop_land	526.059

Screenshot: 7.1 The MLC 2015 area cover

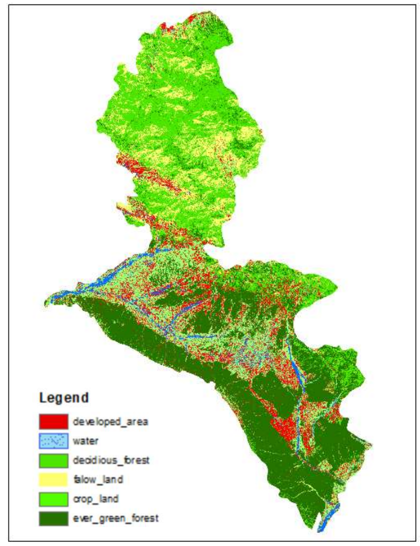


Figure: 7.5 The classification of MLC 2015 and

By 2018, minor fluctuations were observed. Evergreen forest slightly decreased to 976.491 km², and deciduous forest to 841.659 km². However, fallow land expanded sharply to 582.726 km², while crop land remained nearly constant at 526.613 km². Developed areas also rose to 319.779 km², hinting at gradual urban expansion. The increase in fallow land could indicate early stages of land degradation or transitions in land use.

Value	Count	FEATURES	area(KM2)
1	3177902	developed_area	319.779
2	837678	water	83.767
3	9764911	evergreen_forest	976.491
4	8436590	deciduous_forest	841.659
5	58272639	fellow_land	582.726
6	5266135	crop_land	526.613

Screenshot: 7.2 The MLC 2018 area cover

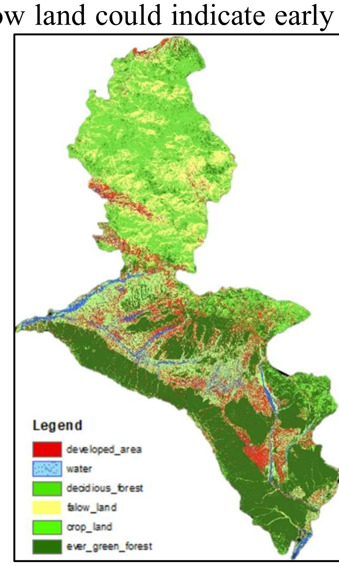


Figure: 7.6 The classification of MLC 2018 and area cover

The 2021 results showed a noticeable shift. Crop land expanded significantly to 803.618 km², suggesting agricultural intensification. Both forest categories slightly decreased—evergreen to

976.198 km² and deciduous to 840.771 km². Fallow land remained moderate at 325.121 km², while developed area slightly increased to 321.972 km². Water bodies saw a small decrease to 82.854 km². These changes reflect possible deforestation for cultivation and urban growth.

Value	Count	FEATURES	area(KM2)
1	3159716	developed_area	321.9716
2	828540	water	82.854
3	8467712	deciduous_forest	840.771
4	58278338	crop_land	803.618
5	5261211	fellow_land	325.121
6	9760338	evergreen_forest	976.198

Screenshot: 7.3 The MLC 2021 area cover

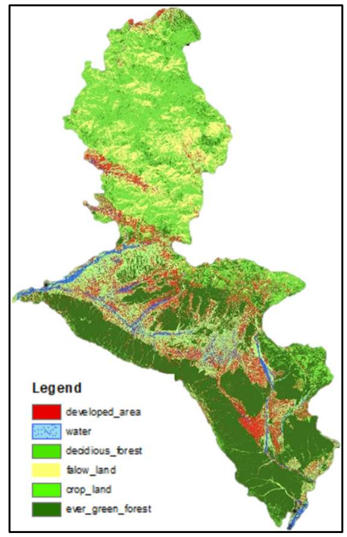


Figure: 7.7 The classification of MLC 2021 and area cover

By 2024, the trend of increasing crop land continued, reaching 582.783 km², though slightly lower than 2021. Deciduous forest held at 839.349 km², and evergreen forest slightly declined to 975.034 km². The developed area stabilized at 322.018 km². Fallow land remained high at 526.49 km², and water maintained at 82.854 km². These numbers suggest forest areas are stable but under pressure, while cropland and fallow land dominate changes.

Value	Count	FEATURES	area(KM2)
1	3160187	developed_area	322.018
2	828540	water	82.854
3	8473492	deciduous_forest	839.349
4	58278381	crop_land	582.783
5	5261437	fellow_land	526.49
6	9753818	evergreen_forest	975.034

Screenshot: 7.4 The MLC 2024 area cover

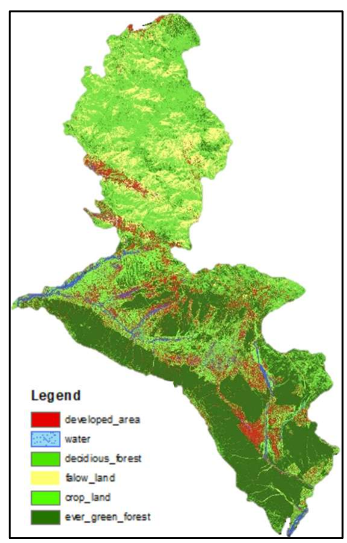


Figure: 7.8 The classification of MLC 2024 and area cover

The semi-supervised model for 2015 showed a close match in major class areas with MLC. Evergreen forest was recorded at 976.894 km², and deciduous forest at 849.936 km². However, fallow land significantly increased to 580.781 km², contrasting with the MLC value of 322.291

km², indicating potential overestimation. Developed area was slightly lower at 315.723 km², while crop land was recorded at 525.773 km². Water bodies dropped slightly to 81.854 km².

Value	Count	FEATURES	area(KM2)
1	3160673	developed_area	315.723
2	9731447	evergreen_forest	976.894
3	828540	water	81.854
4	8499364	deciduous_forest	849.936
5	58278100	fellow_land	580.781
6	5257731	crop_land	525.773

Screenshot: 7.5 The STC 2015 area cover

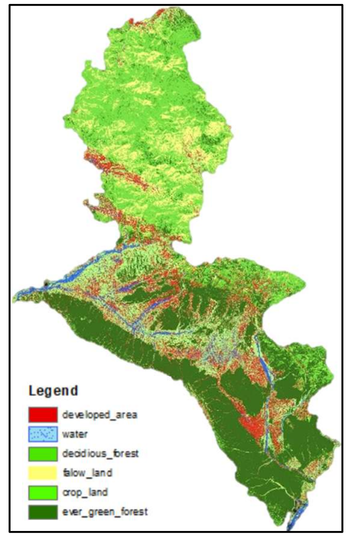


Figure: 7.9 The classification of Semi-Supervised 2015 and area

In 2024, the semi-supervised classification maintained similar forest estimates (evergreen forest at 974.144 km², deciduous forest at 849.936 km²) as the MLC. Fallow land, however, remained high at 582.781 km², echoing the semi-supervised 2015 trend. Crop land was slightly lower at 524.773 km², while developed areas and water bodies were consistent with previous observations. These results suggest the semi-supervised approach tends to overestimate fallow land and underestimate minor variations in crop and developed areas, reflecting slightly less precise land allocation than MLC.

Value	Count	FEATURES	area(KM2)
1	3160673	developed_area	317.723
2	9731447	evergreen_forest	974.144
3	828540	water	82.854
4	8499364	deciduous_forest	849.936
5	58278100	fellow_land	582.781
6	5257731	crop_land	524.773

Screenshot: 7.6 The STC 2024 area cover

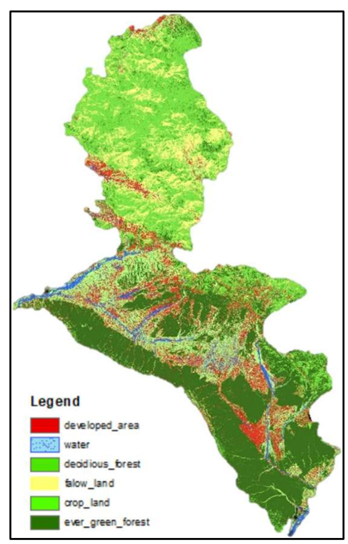


Figure: 7.10 The classification of Semi-Supervised 2018 and area cover

THRESHOLD:

The threshold value is a critical parameter used primarily in classification tasks. It determines the cutoff point at which a model decides between different classes.
The choice of the best threshold value depends on your research objective. Here's a guide based on common scenarios:

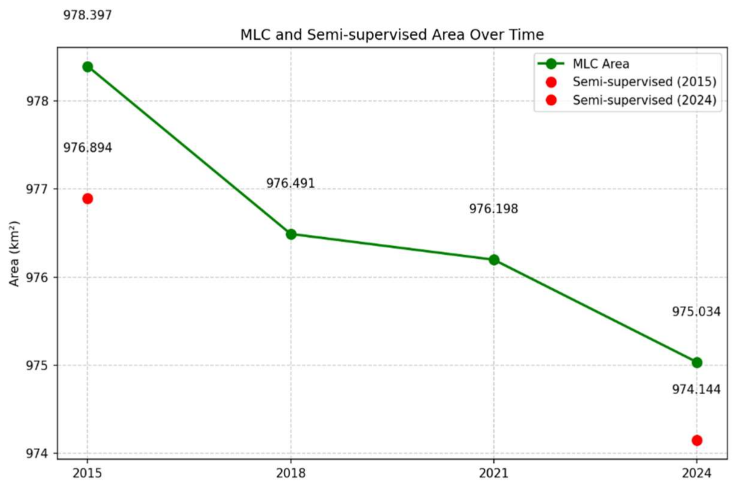
A	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
country	Threshold	2015	2016	2017	2018	sum 15-18	2019	2020	2021	sum 19-21	2022	2023	2024	sum 22-24	sum 15-24
India	0	11	37	178	22	248	13	12	32	57	40	168	81	289	594
India	10	9	28	161	19	217	8	7	27	42	29	96	45	170	429
India	15	9	28	160	19	216	8	6	26	40	28	89	41	158	414
India	20	8	28	159	19	214	8	6	25	39	28	85	41	154	407
India	25	8	28	158	18	212	7	6	24	37	27	82	39	148	397
India	30	8	27	156	18	209	7	6	23	36	27	74	35	136	381
India	50	1	19	129	16	165	5	6	18	29	23	48	24	95	289
India	75	0	1	15	3	19	1	2	5	8	9	11	8	28	55

Screenshot: 7.7 The ground truth from global forest watch 2015-2025

When selecting a threshold for detecting deforestation, the choice depends on the specific goals of your study. If your objective is to **capture the maximum extent of deforestation**, using a threshold of **0** is recommended. This setting is the most inclusive and detects even minimal changes in vegetation cover, making it suitable for identifying all possible deforested areas, including those that are partially degraded or undergoing subtle disturbances.
For studies that require a **balance between precision and reliability**, a threshold of **25 or 30** is typically ideal. These moderate values help filter out insignificant changes while still capturing meaningful deforestation patterns. This makes them suitable for general-purpose analyses where interpretability and consistency are important.
On the other hand, if your focus is on **severe or large-scale deforestation**, higher thresholds such as **50 or 75** are more appropriate. These stricter values highlight only areas with significant forest loss, which is useful for identifying critical zones that require urgent attention, such as regions targeted for conservation efforts or policy intervention.

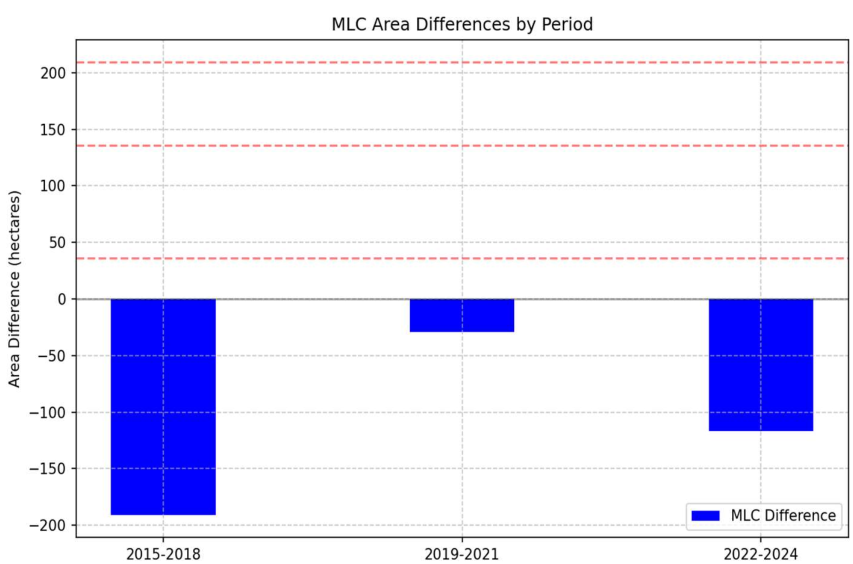
VISUALS COMPARISION

This line chart tracks the changes in evergreen forest area (in square kilometers) over four key years: 2015, 2018, 2021, and 2024. The green line with circular markers represents the MLC-derived area values, which show a gradual but consistent decline from 978.397 km² in 2015 to 975.034 km² in 2024. In addition, red markers represent semi-supervised classification values for 2015 and 2024, which are slightly lower than the MLC values—indicating that the semi-supervised method estimates even smaller areas of evergreen forest cover, further reinforcing the trend of decline. The labeling of data points adds clarity to the observed differences between years and between the two classification methods.



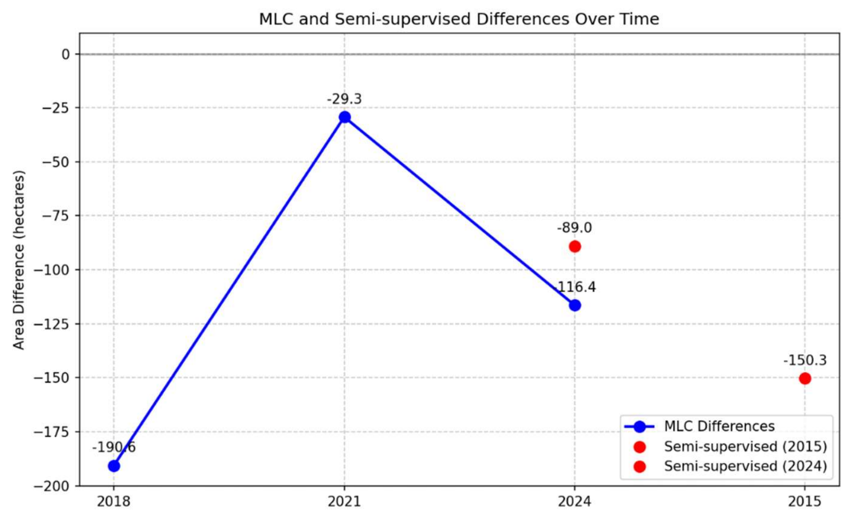
Screenshot: 7.8 The line chart for MLC & Semi-Supervised over time

The bar chart in the second image illustrates the area differences (in hectares) between consecutive periods: 2015–2018, 2019–2021, and 2022–2024. The bars are all negative, confirming a declining trend in evergreen forest area across all periods. The largest drop occurred during 2015–2018, with a loss of nearly 190 hectares, while the 2019–2021 period shows a relatively smaller decline. The final period, 2022–2024, again shows a significant decrease, emphasizing that the rate of forest loss may be accelerating again. The red dashed lines appear to represent a threshold or standard deviation band, possibly indicating expected variation, and the consistent breach of these lines in the negative direction suggests a statistically significant decrease.



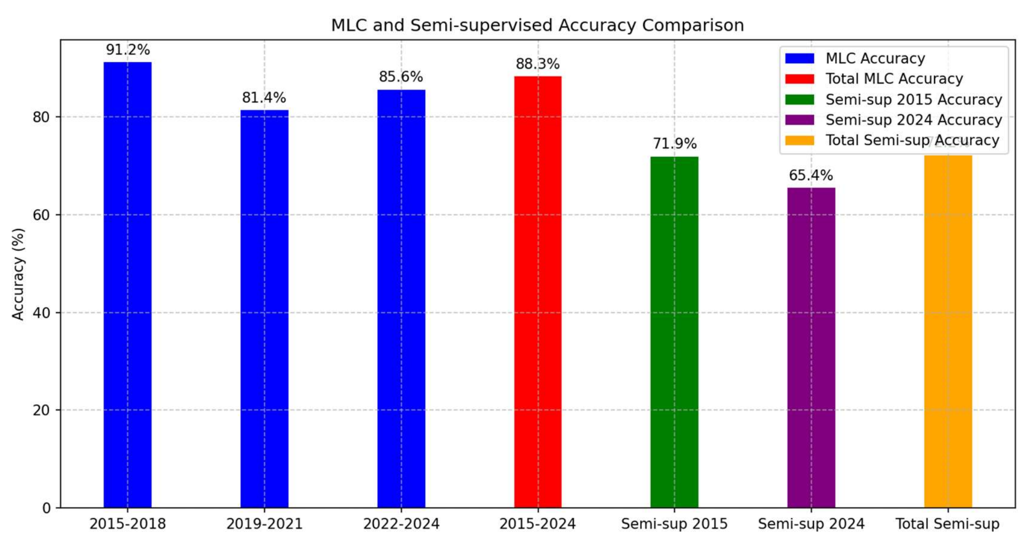
Screenshot: 7.9 The Area chart for MLC over time

Bellow image shows the differences in evergreen forest area over time, comparing outputs from the MLC (Maximum Likelihood Classification) method and semi-supervised classification. The blue line represents the area differences detected by MLC in three key intervals. Between 2015 and 2018, a substantial loss of -190.6 hectares was recorded, marking the most significant decline. This was followed by a minor drop of -29.3 hectares from 2018 to 2021, suggesting a brief period of relative stability. However, the trend reversed again between 2021 and 2024, with another notable decline of -116.4 hectares. Overlaid red points indicate semi-supervised area differences for the years 2015 and 2024, which reported losses of -150.3 hectares and -89.0 hectares respectively. These results confirm a consistent decrease in forest cover across both methods, though semi-supervised values tend to be slightly less severe. The graph reflects a shared trajectory of deforestation, validating the reliability of both techniques in detecting land cover change, despite methodological differences.



Screenshot: 7.10 The line chart for MLC & Semi-Supervised difference

Bellow image shows the presents a comparative analysis of classification accuracy for MLC and semi-supervised methods over different time periods. MLC achieved the highest accuracy of 91.2% during the 2015–2018 period, followed by 81.4% for 2019–2021, and 85.6% from 2022–2024, with an overall accuracy of 88.3% across the full 2015–2024 span. In contrast, the semi-



Screenshot: 7.11 The bar chart for MLC & Semi-Supervised accuracy

supervised method showed lower performance, recording 71.9% accuracy in 2015 and 65.4% in 2024, resulting in a total average of 68.65%. This clearly highlights MLC's superior classification accuracy and reliability, particularly in tasks such as forest area mapping. While semi-supervised methods offer flexibility in situations with limited labeled data, the results emphasize that MLC remains more dependable for consistent and precise classification over time.

OUTPUT ACCURACY SUMMARY

The table presents a comparative analysis of actual losses, prediction differences, and resulting accuracy over various time periods for two classification methods: a general model and a semi-supervised (semi-sup) model.

Period	Actual Loss	Difference	Accuracy
2015-2018	209.00	-190.60	91.20%
2019-2021	36.00	-29.30	81.39%
2022-2024	136.00	-116.40	85.59%
2015-2024	381.00	-336.30	88.27%
Semi-sup 2015	209.00	-150.30	71.91%
Semi-sup 2024	136.00	-89.00	65.44%
Total Semi-sup	381.00	-275.00	72.18%

Screenshot: 7.12 The bar chart for MLC & Semi-Supervised accuracy summary

From **2015 to 2018**, the actual loss was **209.00**, and the prediction difference was **-190.60**, achieving the **highest accuracy** of **91.20%**. In the **2019–2021** period, the loss dropped significantly to **36.00**, with a difference of **-29.30** and an accuracy of **81.39%**. For **2022–2024**, the actual loss rose to **136.00**, with a difference of **-116.40**, resulting in **85.59%** accuracy. When considering the entire range from **2015–2024**, the cumulative actual loss was **381.00**, and the total difference amounted to **-336.30**, giving an overall accuracy of **88.27%**.

In contrast, the **semi-supervised model** exhibited comparatively lower accuracies. For the year **2015**, the actual loss matched the earlier value of **209.00**, but the difference was **-150.30**, yielding a lower accuracy of **71.91%**. In **2024**, the actual loss was **136.00**, with a difference of **-89.00**, resulting in an even lower accuracy of **65.44%**. Combining both semi-supervised results, the total actual loss was **381.00**, with an aggregate difference of **-275.00**, leading to an overall semi-supervised accuracy of **72.18%**.

CONCLUSIONS

The deforestation mapping project in Dehradun district from 2015 to 2024 utilized Sentinel-2 satellite imagery and a variety of classification techniques to monitor land cover changes, particularly deforestation. The study began with unsupervised classification methods such as K-Means and ISODATA, but these produced inconsistent and unrealistic results, proving ineffective for accurate land classification. The focus then shifted to supervised learning methods, where Random Forest (RF) and Maximum Likelihood Classification (MLC) were tested. Although RF performed reasonably well, MLC emerged as the most reliable and accurate, achieving consistently higher classification performance. To improve accuracy further, especially in situations with limited labeled data, the study also tested a semi-supervised approach that combined MLC with additional unlabeled samples. However, this method underperformed, with an overall accuracy of 72.18%, compared to 88.27% achieved by the purely supervised MLC model. This accuracy gap underscored MLC's robustness and reliability across the 2015–2024 timeline for detecting genuine forest loss patterns. The results revealed a clear and steady decline in evergreen forest cover across the district, largely driven by urbanization and the expansion of agricultural land. The project successfully mapped six major land cover categories and identified key zones of deforestation, with classification accuracy nearing 90%.

FUTURE WORK

This project sets the stage for ongoing monitoring and advanced research in deforestation mapping. Future work will focus on extending the analysis using more recent satellite data to continuously track forest cover changes and refine the understanding of evolving deforestation trends. Incorporating additional environmental variables such as elevation, rainfall, and soil type, along with more comprehensive ground truth data, will be crucial to further enhance the accuracy and robustness of classification results. Moreover, close collaboration with local authorities, forest departments, and conservation organizations will be essential for translating the findings into actionable, data-driven conservation strategies. These partnerships will also support the monitoring and evaluation of policy interventions aimed at promoting sustainable land use and forest preservation in the region.

REFERENCES

[1] Authors R Saini, Sanjay Kumar Ghosh Publication date 2018/11/19 Journal The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Volume 42 Pages 683-688 Publisher Copernicus GmbH.

[2] Authors Rashmi Saini, Sanjay Kumar Ghosh Publication date 2019/12 Journal Geocarto International Pages 1-19 Publisher Taylor and francis.

[3] Camps-Valls G, Bruzzone L. 2005. Kernel-based methods for hyperspectral imageclassification. IEEE Trans Geosci Remote Sens. 43 (6):1351–1362.

[4] Chan JCW, Paelinckx D. 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sens Environ. 112(6):2999–3011.

[5] Castelluccio, M., Poggi, G., Sansthree, C., & Verdoliva, L. (2015). Land use classification in remote sensing images by convolutional neural networks. In 2015 IEEE international geoscience and remote sensing symposium (pp. 2565-2568). IEEE.

[6] Global Forest Watch is an online platform providing real-time data and tools to monitor forests worldwide. It enables users to track deforestation, forest degradation, and conservation efforts using satellite imagery, advanced analytics, and open data. The platform supports policymakers, researchers, and environmentalists in making informed decisions to protect global forests and combat climate change.