**“**ASSESSING DEFORESTATION USING SATELLITE IMAGERY IN DEHRADUN DISTRICT**”**

**MAJOR PROJECT REPORT**

Submitted in partial fulfillment of the requirements for 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: Mr. Ramesh Kumar

(Project Supervisor)

**Mrs Ruchi Tailwal**

(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 Ashish Negi**

(H.O.D.)

Computer Science & Engineering

G.B. Pant Institute of Engineering &

Technology, Pauri Garhwal

(Uttarakhand)

# 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, Mrs. Ruchi Tailwal and 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.**

1. **INTRODUCTION………………………………………..06**
2. **LITERATURE REVIEW..................................................07**
3. **PROBLEM STATEMENT…………………………....…08**
4. **METHODOLOGY……………………………………….09**
5. **SYSTEM REQUIREMENTS……….……..………….....12**
6. **WORK CARRIED OUT AND FUTURE REVIEW…...13**
7. **RESULT AND DISCUSSION………………..………….14**
8. **CONCLUSIONS…………………………………………..24**
9. **REFERENCES……………………………………………29**

**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 five-year period, from 2015 to 2020, 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 201, 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.

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 2020. 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 five-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 2020 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

* 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.

* 1. **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.

* 1. **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.

* 1. **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.

* 1. **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.

* 1. **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**

* 1. This project, **"Assessing Deforestation Using Satellite Imagery in Dehradun district"** aims to leverage Sentinel-2 satellite images to analyze deforestation trends between the years 2015 and 2020.
  2. By comparing forest cover across these years, the study seeks to quantify the extent of deforestation, identify affected regions, and provide a basis for understanding the drivers and impacts of forest loss.
  3. The project's outcomes will contribute to better monitoring and management of forest resources, aiding in conservation and sustainable development efforts.

## METHODOLOGY

## 

Data Collection phase

Data

Pre-Processing phase

Model Training phase

Comparison and analysis phase

Metrics

phase

Classification phase

## 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

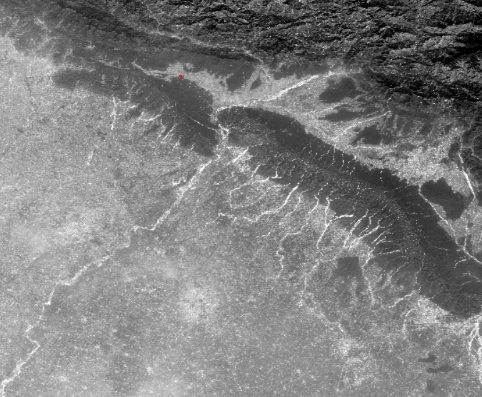
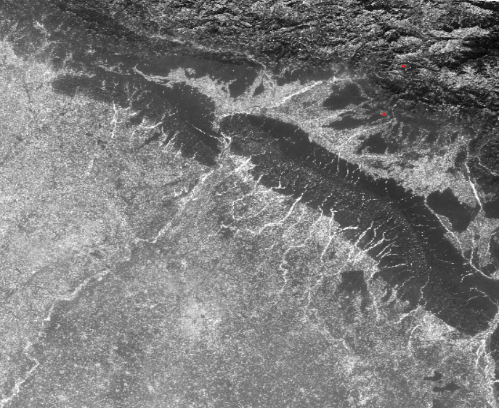
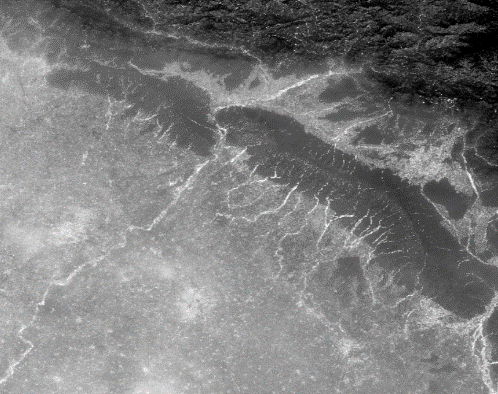
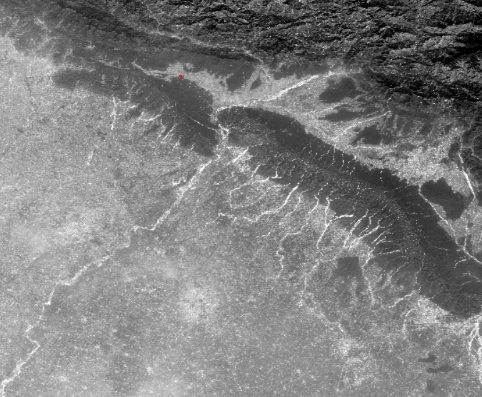
## India Map and Dehradun Shapefile: Download India map and extract Dehradun shapefile using ArcGIS software.

## Sentinel-2 Satellite Imagery: Access Sentinel-2 satellite imagery for October to December 2015 and 2020 from the Copernicus Open Access Hub, covering Dehradun district across multiple tiles.

## 4.2 Data Pre-Processing

## Atmospheric Correction: Apply atmospheric correction to Sentinel-2 MSI files using SNAP software to enhance data accuracy and consistency.

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



Band 2

Band 3

Band 4

Band 8

**Composition**

## 4.3 Model train phase

**4.3.1.** **Random Forest (RF) Classification**:

**Ensemble Learning**:

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.

**Bootstrapping and Aggregation**:

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.

**Feature Importance**:

Can provide an estimate of feature importance, indicating which variables are most influential in the classification process.

**4.3.2. Maximum Likelihood Classification (MLC)**:

**Probabilistic Approach**: 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.

**Assumptions**: Assumes that the pixel values for each class follow a multivariate normal distribution. This means MLC works best when this assumption holds true.

**Training Data**: 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.

**Output**: Produces a classified image where each pixel is assigned to five of the predefined classes. Additionally, it can provide a probability map indicating the confidence level of the classification.

## Both RF and MLC are trained on the 2015 and 2020 raster data and corresponding level dataset.

## 4.4 Classification phase

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

## 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.

## 4.5 Comparison and Analysis phase

## Compare the classified land cover maps of 2015 and 2020 to identify and analyze changes in forest cover and land use over the five-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 2015 vs 2020.

# 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 100 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**

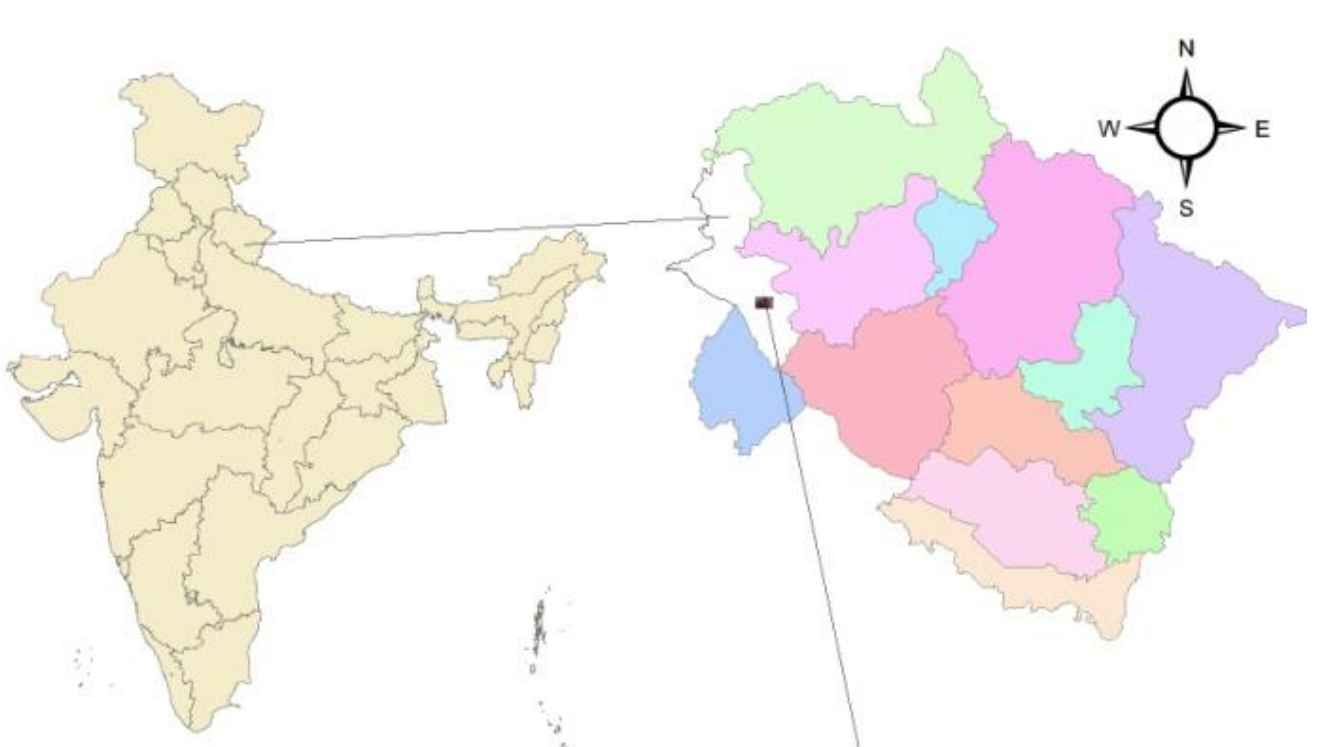
**Work Carried Out**

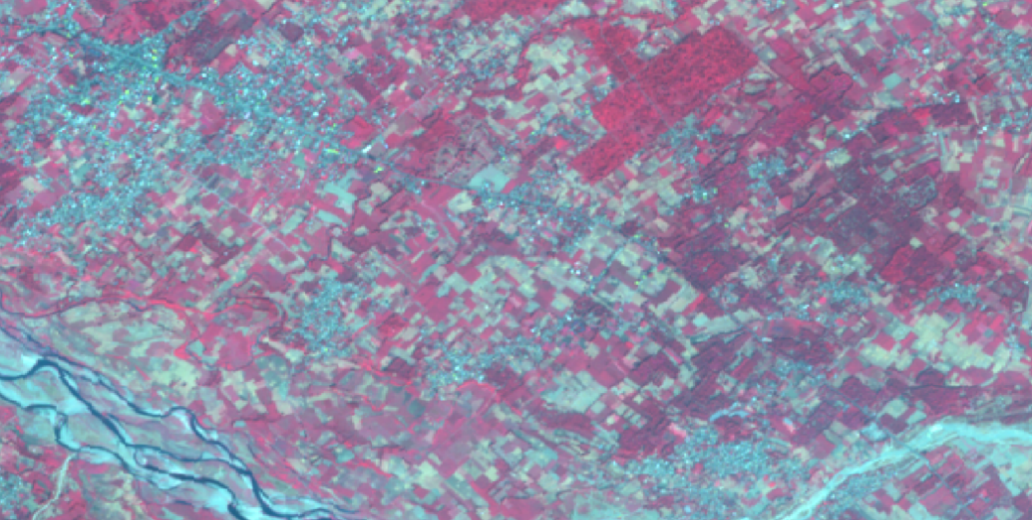
1. **Deforestation Analysis for 2015 Data**
   * Detailed analysis performed to identify deforestation patterns and hotspots.
   * Utilized tools, e.g., GIS, remote sensing, etc.
   * We got land cover classified data for analysis purpose and the area covered by those classes.
2. **Deforestation Analysis for 2020 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.
3. **Data Comparison and Interpretation**
   * Conducted a comparative analysis between 2015 and 2020 datasets.
   * Highlighted differences and identified trends in deforestation.

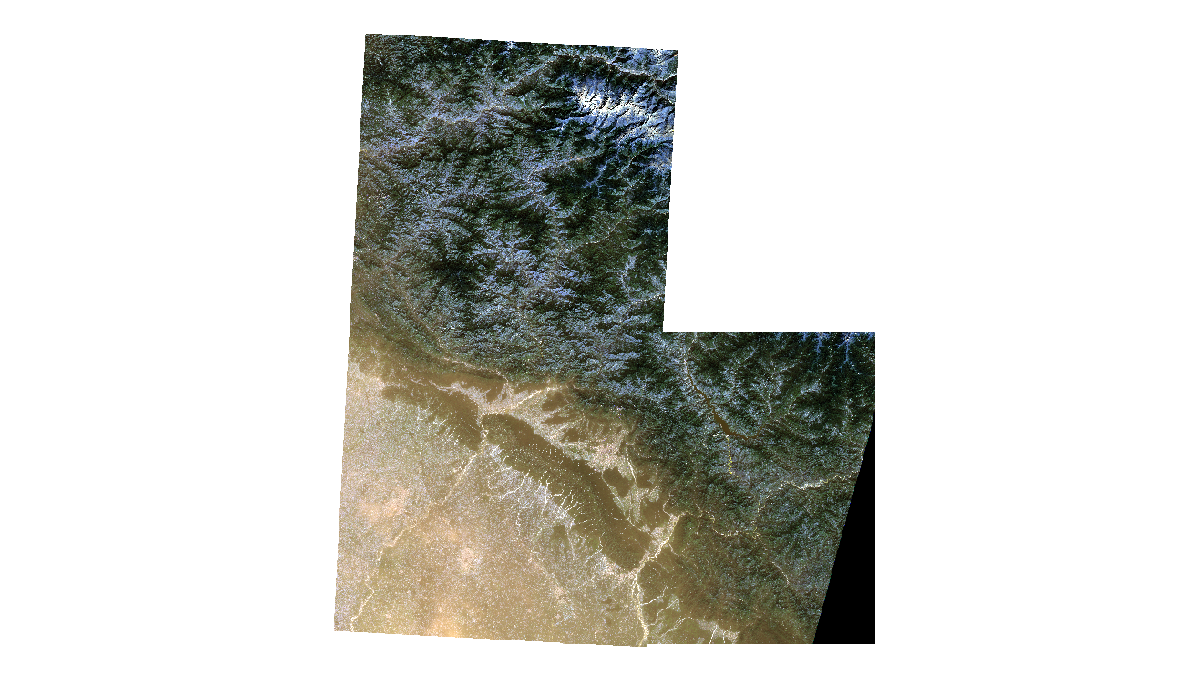
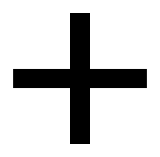
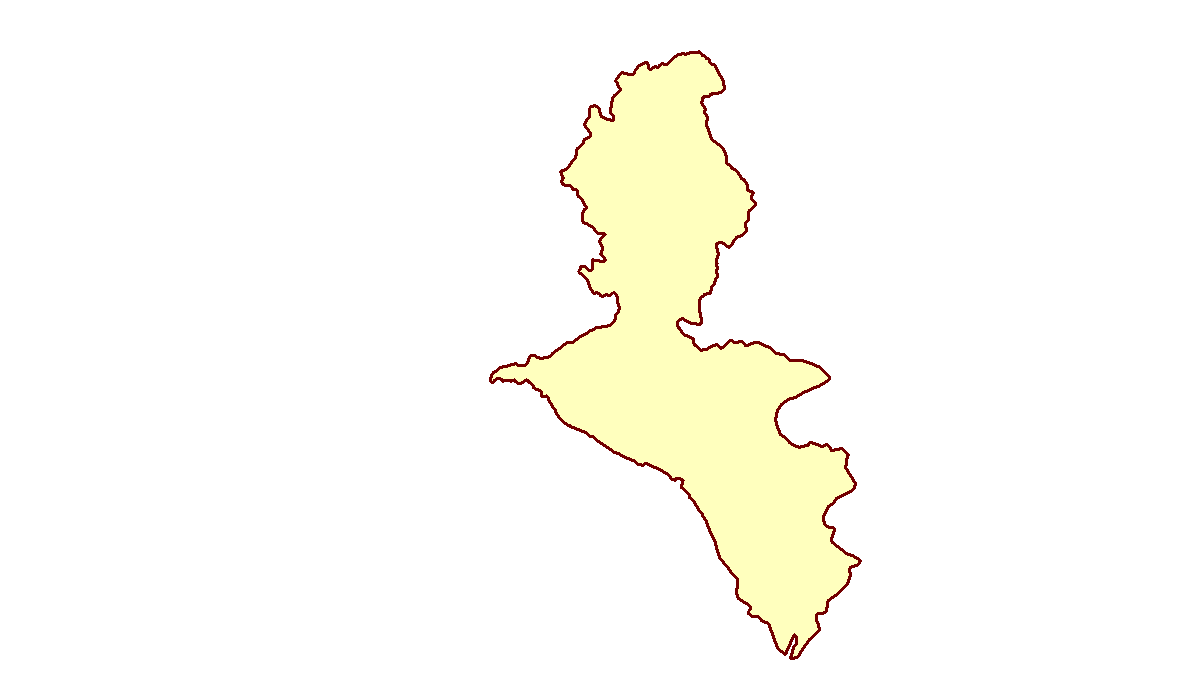
**Future Review**

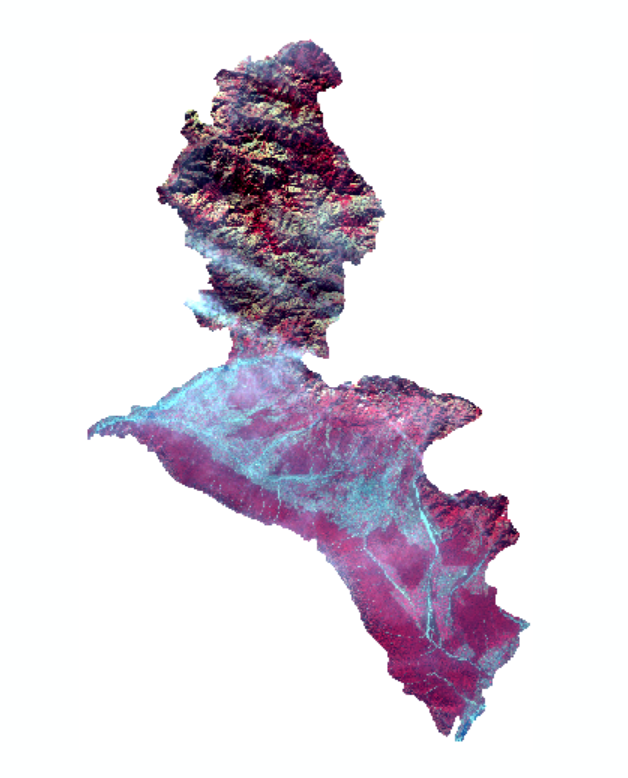
1. **Long-Term Trend Analysis**
   * Extend the study to include data from additional years for a broader trend analysis.
2. **Investigation of Underlying Causes**
   * Explore potential socio-economic, environmental, factors contributing to observed changes.
3. **Impact Assessment**
   * Assess the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

**RESULT & DISCUSSION**

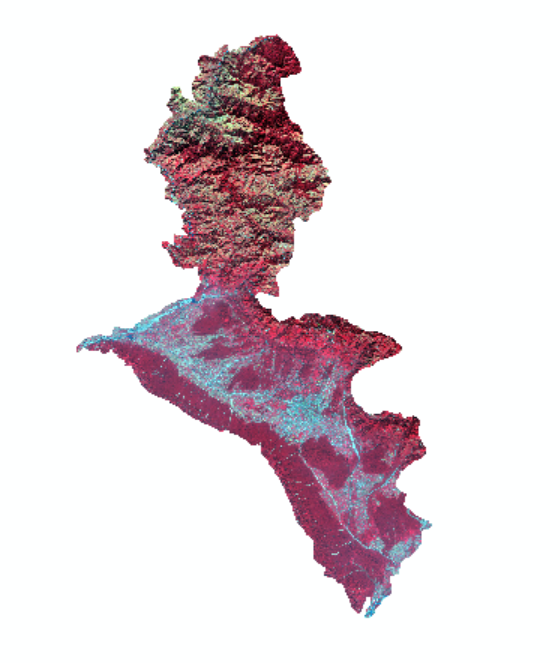




**CLASSIFICATION RESULTS**



2015



2020

**Masking**

The MLC, Random Forest algorithm was employed to classify the preprocessed 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 five of the six land cover classes.

**Land Cover Classes**

**Evergreen Forest:** Represents areas with dense, year-round foliage. These areas are critical for maintaining biodiversity and ecological balance.

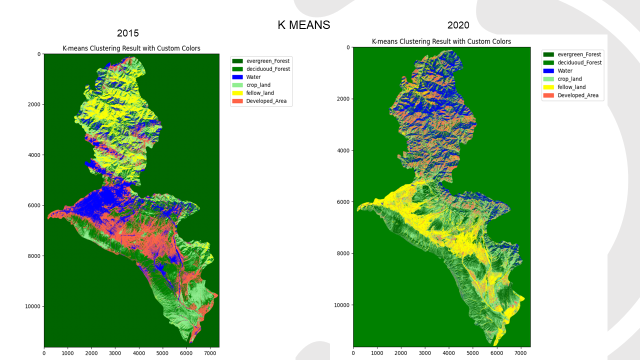
**Deciduous Forest:** Includes regions with trees that shed leaves seasonally. These forests play a vital role in carbon sequestration and habitat provision.

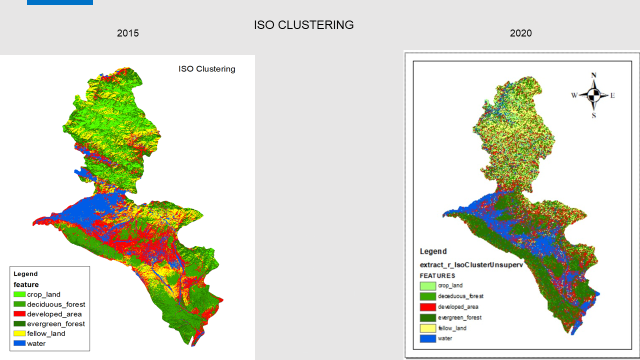
**Cropland:** Denotes agricultural lands used for crop production. These areas are essential for food security and local economies.

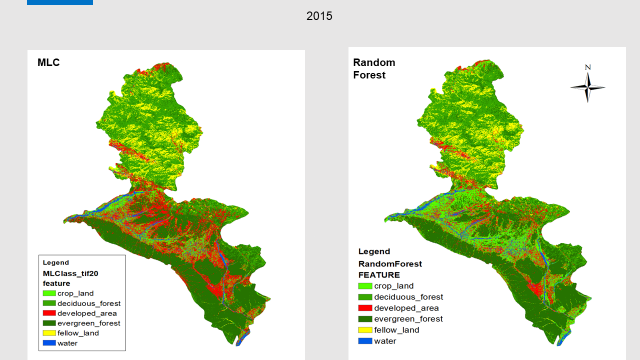
**Fallow Land:** Areas left uncultivated to restore soil fertility. This classification helps in understanding agricultural practices and land use changes.

**Water Bodies:** Includes rivers, lakes, and reservoirs. These are vital for hydrological studies and water resource management.

**Developed Area:** Represents urban and built-up regions. This classification is crucial for urban planning and infrastructure development.

Below is the classification of 2015 and 2020 raster image using k means.

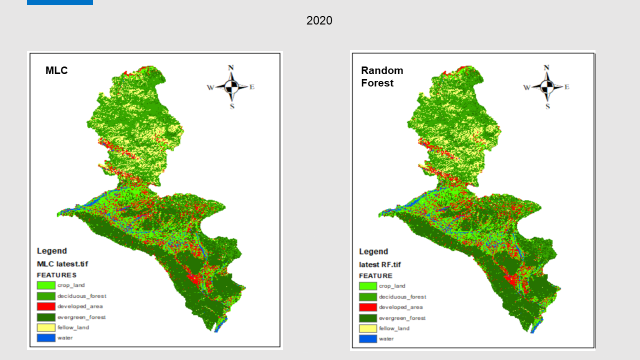
We did unsupervised classification using ISO clustering and calculated the area matrix

****We did supervised classification using MLC and RF on labelled data of 2015.

2015

2020

We did supervised classification using MLC and RF on labelled data of 2020.

****

**COMPARATIVE ANALYSIS**

7.1. MLC previous:

This MLC dataset is of the year 2015.

Here, count shows the pixel of size 10/10.

Area is in kilometres square later we will convert it in hectare.

|  |  |  |  |
| --- | --- | --- | --- |
| VALUE | COUNT | FEATURES | Area(Km2) |
| 1 | 3177239 | developed\_area | 317.7239 |
| 2 | 837678 | water | 83.7678 |
| 3 | 9783970 | evergreen\_forest | 978.397 |
| 4 | 8424283 | deciduous\_forest | 842.4283 |
| 5 | 3222916 | fellow\_land | 322.2916 |
| 6 | 5260591 | crop\_land | 526.0591 |

7.2. MLC latest:

This MLC dataset is of the year 2020.

Here, count shows the pixel of size 10/10.

Area is in kilometres square later we will convert it in hectare.

|  |  |  |  |
| --- | --- | --- | --- |
| VALUE | COUNT | FEATURES | Area(Km2) |
| 1 | 3177903 | developed\_area | 317.7903 |
| 2 | 837678 | water | 83.7678 |
| 3 | 9764908 | evergreen\_forest | 976.4908 |
| 4 | 8437236 | deciduous\_forest | 843.7236 |
| 5 | 3222817 | fellow\_land | 322.2817 |
| 6 | 5266135 | crop\_land | 526.6135 |

7.3. RF previous:

This RF dataset is of the year 2015.

Here, count shows the pixel of size 10/10.

Area is in kilometres square later we will convert it in hectare.

|  |  |  |  |
| --- | --- | --- | --- |
| VALUE | COUNT | FEATURES | Area(Km2) |
| 1 | 3167517 | developed\_area | 316.7517 |
| 2 | 837678 | water | 83.7678 |
| 3 | 9944728 | evergreen\_forest | 994.4728 |
| 4 | 8333151 | deciduous\_forest | 833.3151 |
| 5 | 3223640 | fellow\_land | 322.364 |
| 6 | 5199963 | crop\_land | 519.9963 |

7.4. RF latest:

This RF dataset is of the year 2020.

Here, count shows the pixel of size 10/10.

Area is in kilometres square later we will convert it in hectare.

|  |  |  |  |
| --- | --- | --- | --- |
| VALUE | COUNT | FEATURES | Area(Km2) |
| 1 | 3090412 | developed\_area | 309.0412 |
| 2 | 849483 | water | 84.9483 |
| 3 | 9928238 | evergreen\_forest | 992.8238 |
| 4 | 8311467 | deciduous\_forest | 831.1467 |
| 5 | 3028955 | fellow\_land | 302.8955 |
| 6 | 5498122 | crop\_land | 549.8122 |

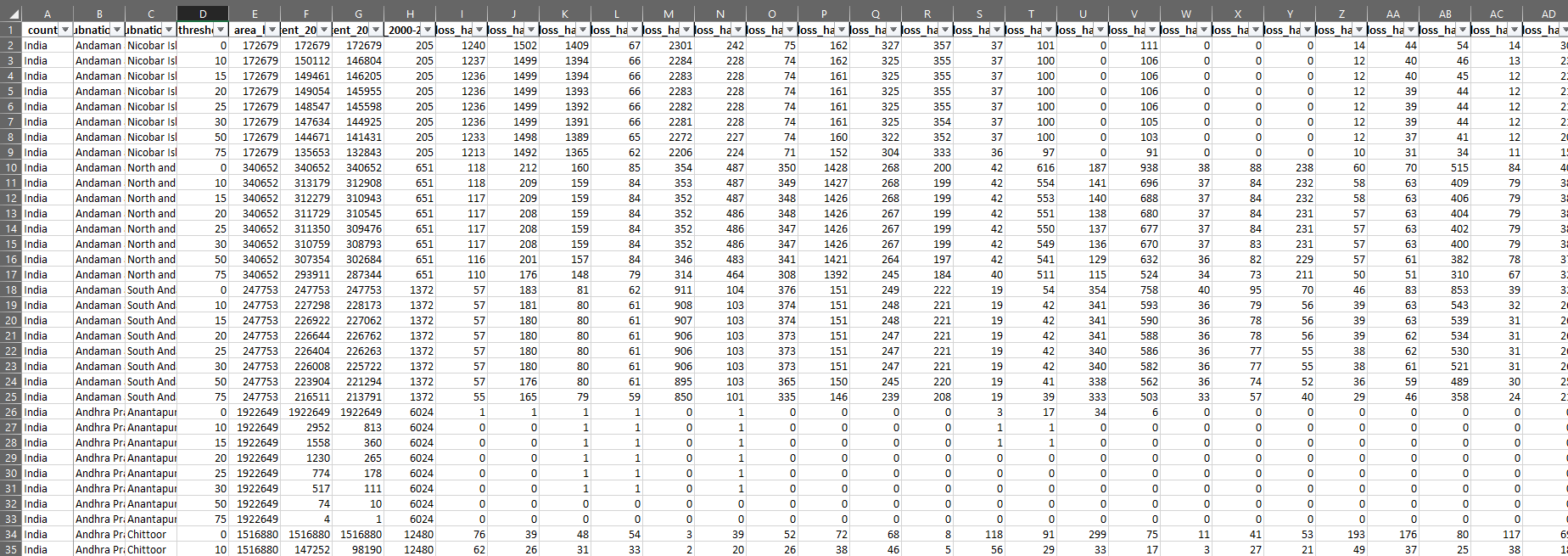
Area calculation by pixel

Area = (COUNT \* 10 \* 10 ) / 1000000

**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:



1. If you aim to capture the maximum deforestation:

**Threshold = 0**: This value generally represents the most inclusive measurement of deforestation, covering even minimal changes in vegetation.

Use this if your priority is to identify all possible areas with deforestation, including minor or degraded areas.

2. If you need a balance between precision and reliability:

**Threshold = 25 or 30**: These thresholds often represent moderate deforestation, avoiding extreme inclusivity (noise) or exclusivity (missing significant deforested areas).

Suitable for general-purpose studies where you need reliable, interpretable data.

3. If your study focuses on severe deforestation:

**Threshold = 50 or 75**: These thresholds are stricter and capture only areas with significant loss of forest cover.

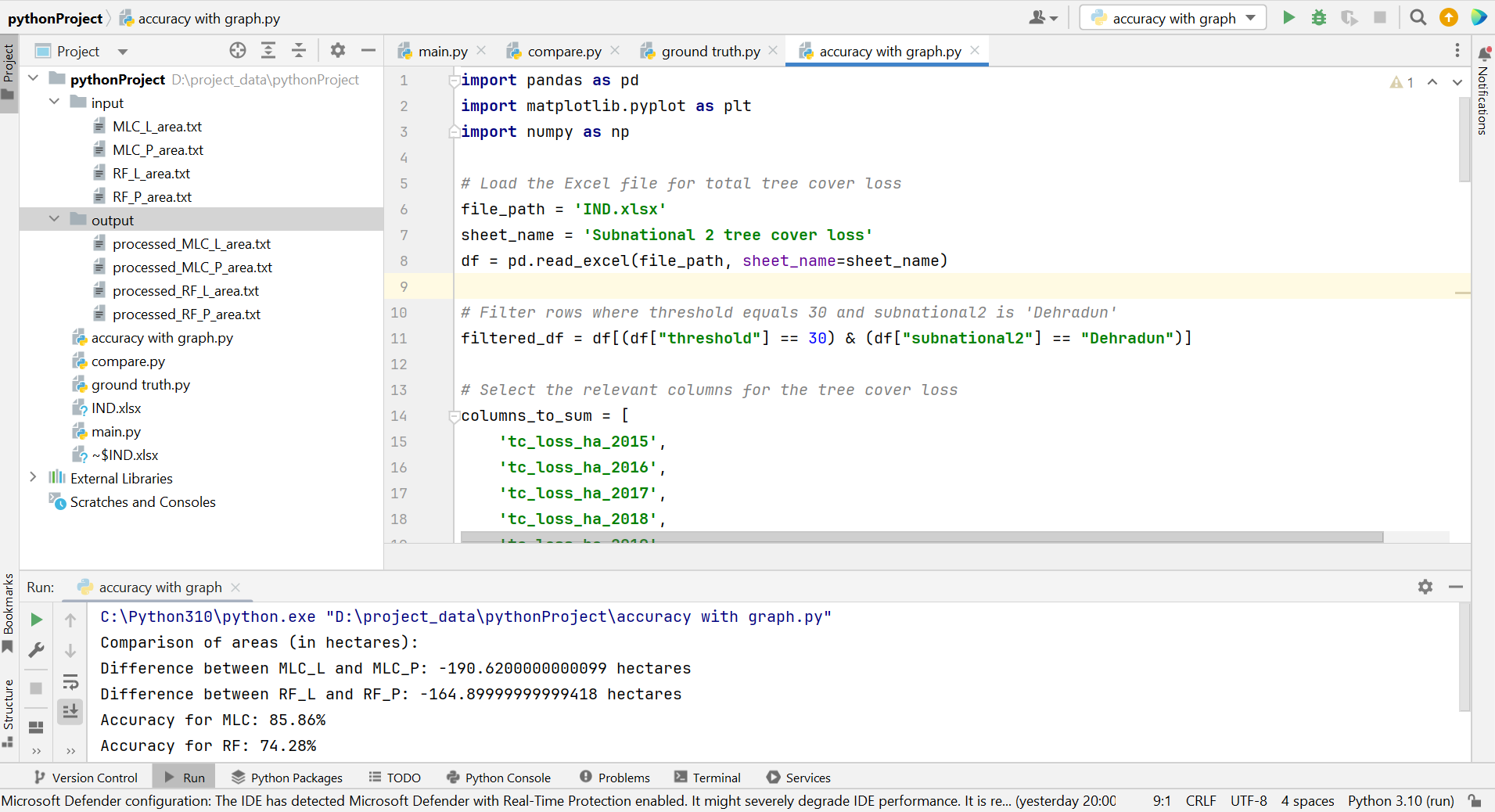
Use this if your goal is to identify high-priority zones for intervention or policy-making.

This Python script analyzes the "Outputs" dataset, comparing it against a reference dataset from global forest research. The purpose of the script is to evaluate the accuracy of forest cover predictions and calculate the total forest cover area.

**Functionality:**

1. **Input Data Handling**: Load and preprocess the Outputs dataset and the global forest research dataset.
2. **Comparison**: Compare the Outputs dataset with the reference data to identify matching and mismatching areas.
3. **Accuracy Calculation**: Compute the percentage of correctly classified areas based on overlaps and mismatches between the datasets.
4. **Cover Area Calculation**: Calculate the total forest cover area based on the Outputs dataset.
5. **Visualization (optional)**: Provide visual insights such as a confusion matrix or maps for better understanding.

**Key Steps in the Script:**

1. Import necessary libraries like pandas, numpy, and geospatial libraries (e.g., geopandas, rasterio).
2. Load datasets, ensuring they are in compatible formats (e.g., shapefiles, GeoTIFFs).
3. Perform spatial or categorical comparisons depending on the data type.
4. Use statistical methods to evaluate accuracy (e.g., precision, recall, F1 score).
5. Compute the total forest area by summing values (e.g., pixel areas in raster datasets or polygon areas in vector datasets).
6. Output results in a user-friendly format such as CSV reports or visual graphs.

**OUTPUT ACCURACY VISUALIZATION**

**1. Forest Cover Comparison Visualization (MLC vs. Random Forest):**  
The first visualization provides a comprehensive comparison of forest cover predictions generated by the Maximum Likelihood Classification (MLC) and Random Forest models for both previous and latest datasets.

**Highlights:**

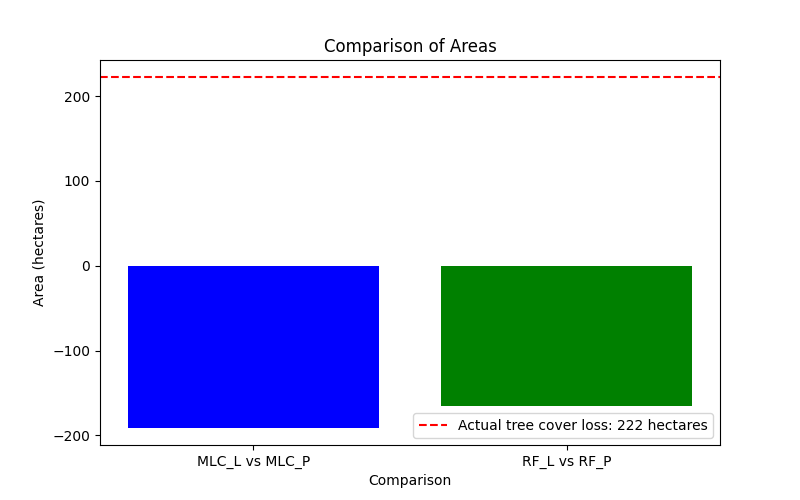
Areas shown in negative (marked in red) represent regions where deforestation has occurred, indicating a reduction in forest cover.

The visualization demonstrates that the total area covered by the MLC model is larger compared to the Random Forest model.

This suggests that the MLC model is more effective in identifying forested regions, highlighting its superiority in generating accurate forest cover predictions.

**Interpretation:**

The negative areas clearly outline the extent of deforestation, emphasizing the regions that require attention.

The MLC model provides better coverage, which could signify a lower tendency for false negatives compared to Random Forest.

**2. Accuracy Comparison Bar Chart (MLC vs. Random Forest):**  
The second visualization is a bar chart comparing the classification accuracy of the MLC and Random Forest models.

**Accuracy Results:**

**MLC Model**: Achieves an accuracy of approximately 85%, indicating strong performance in predicting forest cover.

**Random Forest Model**: Records a lower accuracy of around 76%, suggesting it is less effective in capturing the correct classifications compared to the MLC model.

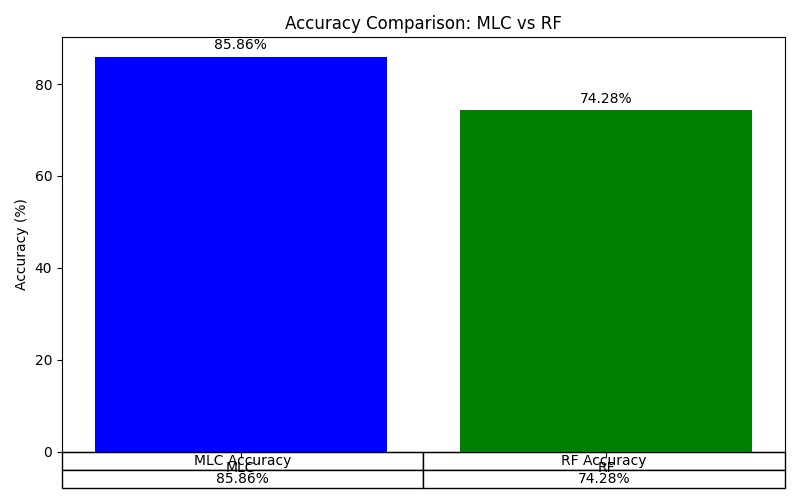
**Highlights:**

The clear difference in bar heights visually emphasizes the superior accuracy of the MLC model.

The bar chart makes it easy to understand and compare the performance of the two models in a concise and intuitive way.

**Interpretation:**

The higher accuracy of the MLC model reinforces its capability to provide reliable forest cover predictions.

The performance gap suggests that the choice of classification algorithm plays a significant role in achieving accurate and meaningful environmental insights.

# CONCLUSIONS

The deforestation mapping project conducted in the Dehradun district from 2015 to 2020 utilized Sentinel-2 satellite imagery and advanced classification algorithms to assess land cover changes. The study produced several key findings. Firstly, the use of Maximum Likelihood Classification (MLC) and Random Forest techniques effectively identified six land cover classes with approximately 90% accuracy. Secondly, the analysis highlighted significant trends, revealing substantial forest loss driven by urbanization and agricultural expansion, thereby identifying critical deforestation patterns. Additionally, the project’s systematic methodology offers a replicable framework for similar studies, leveraging open-source tools and readily accessible satellite data. Furthermore, the results underline important policy implications, emphasizing the need for urgent conservation measures and sustainable land use strategies to guide policymakers in prioritizing intervention areas. Overall, this study underscores the critical role of remote sensing in monitoring deforestation and supporting evidence-based conservation efforts.

**FUTURE WORK**

This project sets the stage for ongoing monitoring and research. Future work will involve:

1. Extending the analysis to more recent satellite data to track ongoing changes and refine deforestation trends.
2. Integrating additional environmental variables and ground truth data to enhance classification accuracy and robustness.
3. Collaborating with local authorities and conservation organizations to implement data-driven conservation strategies and monitor their effectiveness.

**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., Sansfive, 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.