**ABSTRACT**

**Deforestation** has emerged as a **pressing environmental issue**, contributing to **biodiversity loss**, **climate change**, and disruption of **ecological services**. **Monitoring** and **quantifying forest cover change** is essential for **sustainable environmental management** and **effective policy formulation**. This study investigates the **spatiotemporal patterns** of **forest cover change** in the **Dehradun district**, **Uttarakhand, India**, from **2015 to 2024** at three years interval. The research employs **high-resolution Sentinel-2 satellite imagery**, known for its **spectral** and **spatial capabilities**, to detect **land use and land cover (LULC) changes** using **advanced geospatial** and **machine learning techniques**.A **comprehensive classification** was performed using both **unsupervised** (**K-Means**, **ISODATA**) and **supervised** (**Random Forest (RF)**, **Maximum Likelihood Classification (MLC)**) algorithms. The **land cover** was categorized into six distinct classes: **Evergreen Forest**, **Deciduous Forest**, **Cropland**, **Fallow Land**, **Water Bodies**, and **Developed Area**. The **classified maps** were validated using **reference data** from **Global Forest Watch** to ensure **reliability** and assess **accuracy**. The **comparative analysis** revealed **measurable changes** in **forest cover**, particularly a **decline in evergreen forests** and a **significant increase in cropland**, indicating **land use conversion** likely driven by **agricultural** and **developmental pressures**. Among the classification methods, **MLC achieved higher accuracy (~85%)** compared to **RF (~76%)**, closely aligning with **ground truth data**. This study highlights the **effectiveness of integrating satellite remote sensing** with **machine learning models** for **environmental monitoring**. The findings offer valuable insights for **forest conservation efforts** and present a **scalable approach** for monitoring **forest dynamics** in other **ecologically sensitive regions**.

**Keywords:** Copernicus, Global forest watch, **Sentinel-2 satellite.**

**INTRODUCTION**

Forests play a crucial role in maintaining ecological balance, supporting biodiversity, and contributing to global climate regulation. They act as carbon sinks, regulate the hydrological cycle, prevent soil erosion, and provide habitat for numerous species. However, deforestation—driven by rapid urbanization, agricultural expansion, illegal logging, and infrastructure development—poses a severe threat to environmental sustainability and human livelihoods [1]. Globally, the loss of forest cover has been linked to rising greenhouse gas emissions, biodiversity decline, and ecosystem degradation

In recent years, the **Dehradun district in Uttarakhand, India**, has witnessed significant land cover changes. Located at the foothills of the Himalayas, this region presents a unique blend of ecological richness and anthropogenic pressure, making it highly susceptible to deforestation and land transformation [2]. These dynamics necessitate robust monitoring systems to inform conservation efforts, land-use planning, and sustainable development policies.

**Remote sensing technology** has revolutionized environmental monitoring by providing consistent, repeatable, and synoptic observations over vast and inaccessible terrains. High-resolution satellite imagery, such as that from the **Sentinel-2 constellation**, offers valuable spectral and spatial data for detecting and analyzing land use and land cover (LULC) changes. The integration of remote sensing data with **Geographic Information Systems (GIS)** and **machine learning algorithms** enables detailed and automated classification of surface features [4].

This study utilizes Sentinel-2 imagery from 2015 to 2024 to assess forest cover change in the Dehradun district at three-year intervals. By applying and comparing multiple classification methods—including **unsupervised algorithms** like K-Means and ISODATA, and **supervised classifiers** such as Random Forest (RF) and Maximum Likelihood Classification (MLC)—the study aims to produce accurate land cover maps and track deforestation patterns. Additionally, a **semi-supervised classification approach** was introduced to maximize classification accuracy in scenarios where labeled data were limited [5].

The results were validated using ground reference data and the **Global Forest Watch (GFW)** platform, which provides near-real-time forest monitoring based on satellite observations. Through this analysis, the study offers insights into land transformation processes in Dehradun and provides a scalable methodology for forest monitoring in other ecologically sensitive regions. The findings aim to support evidence-based decision-making for conservation, urban planning, and sustainable resource management.

**RELATED WORK**

Several studies have explored the integration of remote sensing data and machine learning techniques for land cover classification and deforestation monitoring.

Investigated the classification of agricultural crops in Roorkee, Uttarakhand, using Sentinel-2 imagery combined with Random Forest (RF) and Support Vector Machine (SVM) classifiers. Their study demonstrated the efficacy of RF, which outperformed SVM with an overall accuracy of 84.22%, particularly in distinguishing high-density vegetation [6]. Emphasized the utility of Sentinel-2A’s high spatial and temporal resolution for land cover mapping and agricultural monitoring, highlighting the strength of ensemble classifiers in capturing fine-scale vegetative patterns [7]. Provided a comprehensive review of the Random Forest classifier in remote sensing, underscoring its adaptability and robustness across diverse geospatial applications, including vegetation mapping, land use classification, and spectral feature selection [8]. Explored the effectiveness of kernel-based Support Vector Machines for classifying hyperspectral remote sensing data. Their study confirmed the strength of SVMs in handling nonlinear relationships, spectral variability, and high-dimensional inputs [9]. Evaluated a wide range of classification algorithms and emphasized the superiority of ensemble learning techniques like Random Forest and Gradient Boosting for ecotope and land cover mapping [10]. Introduced the Continuous Change Detection and Classification (CCDC) algorithm to monitor gradual forest changes using Landsat time series, proving its efficacy in detecting subtle patterns over time [11]. The accuracy and reliability aspects are cited to for general remote sensing applications [12]. The computational intensity and effectiveness with complex distributions are referenced to for computational aspects [13]. Demonstrated that the use of advanced ensemble models like XGBoost can further enhance classification accuracy, especially in heterogeneous and mixed land cover regions, making them suitable for complex deforestation assessments [14].

**DATASETS**

The dataset for this study was derived primarily from **Sentinel-2 satellite imagery**, accessed through the **Copernicus Open Access**  The Sentinel-2 mission provides high-resolution optical imagery, particularly suited for land cover analysis and environmental monitoring.

* 1. **Data Sources**

The primary dataset for this research was obtained from **Sentinel-2A and 2B** satellites, provided by the **European Space Agency (ESA)** through the **Copernicus Open Access** . These satellites offer **high-resolution optical imagery** ideal for environmental and land cover analysis. Data was collected for the **Dehradun district, Uttarakhand, India**, covering the period **from 2015 to 2024** at **three-year intervals (October to December)**.

The spectral bands utilized included:

***Band 2 (Blue)*:** Primarily used for **water body detection**, **atmospheric correction**, and in **true-color imagery**. Its reflectance characteristics help differentiate **clear water from turbid water** and **vegetation from urban features**, especially when combined with other bands.

***Band 3 (Green):***Combined with Bands 4 (Red) and 8 (NIR), it helps in forming **false-color composites** that enhance **vegetation visibility**.These bands were chosen due to their **10-meter spatial resolution** and their effectiveness in **vegetation and land cover classification**. The Sentinel-2 satellites also offer a **5-day revisit time**, enhancing the probability of acquiring cloud-free imagery.

***Band 4 (Red):*** Vital for calculating the **Normalized Difference Vegetation Index (NDVI)**, an important metric for vegetation health. Red light is strongly absorbed by healthy vegetation; hence, it is crucial in differentiating **dense forests (evergreen/deciduous)** from degraded or non-vegetated land.

***Band 8 (Near Infrared):*** Critical for analyzing **plant health**, **biomass**, and **leaf area index**. NIR is **highly reflective** in healthy vegetation, making it ideal for distinguishing **forest cover** from **agricultural or urban land**.

Additionally, **Global Forest Watch (GFW)** was used as a reference for validation. GFW provides near-real-time deforestation and forest change data based on satellite observations, supporting reliable ground-truthing and accuracy assessments.[15]

<https://browser.dataspace.copernicus.eu/?zoom=5&lat=50.16282&lng=20.78613&demSource3D=%22MAPZEN%22&cloudCoverage=30&dateMode=SINGLE>

**Satellite**: Sentinel-2A and 2B

**Sentinel-2**, part of the European Space Agency's Copernicus program, is a constellation of high-resolution optical Earth observation satellites—**Sentinel-2A and Sentinel-2B**—designed to provide detailed imagery for land monitoring. It offers **13 spectral bands** across the visible, near-infrared (NIR), and shortwave infrared (SWIR) regions, with spatial resolutions of **10 m, 20 m, and 60 m**, depending on the band. In this study, **Bands 2 (Blue), 3 (Green), 4 (Red), and 8 (NIR)** were utilized for land cover classification due to their 10-meter resolution and proven utility in vegetation and land use analysis. The Sentinel-2 system features a **5-day revisit time**, allowing for the selection of cloud-free images over the Dehradun district for the years **2015, 2018, 2021, and 2024**. Its imagery was accessed through the **Copernicus Open Access Hub** and pre-processed using SNAP for atmospheric correction, mosaicking, and clipping. Sentinel-2’s high spectral and spatial resolution, combined with frequent coverage, makes it particularly well-suited for monitoring **deforestation, agricultural dynamics, and urban expansion**.

**Provider**: European Space Agency (ESA), via Copernicus Open Access Hub. The Copernicus program, managed by ESA, provides free and open access to Sentinel-2 satellite imagery, which is instrumental for environmental monitoring, land cover classification, and deforestation studies. The data is accessible through the Copernicus Open Access Hub, ensuring transparency and ease of use for researchers and policymakers.

**Timeframes Used**: The study analyzed deforestation trends in the Dehradun district from 2015 to 2024 at three-year intervals (2015, 2018, 2021, and 2024). The imagery was specifically selected for the months of October to December to ensure consistency in seasonal vegetation conditions and minimize cloud cover interference. This temporal resolution allows for the detection of gradual changes in forest cover and land use patterns over the nine-year period.

**Geographical Scope**: The geographical focus of this study is the **Dehradun district**, located in the **state of Uttarakhand, India**, situated at the foothills of the Himalayas. Dehradun spans a diverse landscape that includes dense forest areas, agricultural land, and expanding urban regions [16]. This district was specifically chosen due to its **high rate of deforestation and rapid land use changes** over the past decade, making it a critical region for forest monitoring and environmental assessment. Its varying elevation, ecological diversity, and increasing developmental pressures provide an ideal setting to study land cover dynamics using remote sensing techniques. The district’s boundaries were delineated using official shapefiles in **ArcGIS**, ensuring precise spatial analysis and consistency across all temporal datasets from **2015 to 2024**.

* 1. **Data Specification**

The data used in this study was primarily sourced from the **Sentinel-2A and 2B** satellites, part of the European Space Agency’s Copernicus program. These satellites provide **high-resolution multispectral imagery**, which is especially useful for land cover classification and environmental monitoring.

***Spatial Resolution:*** 10 meters

***Temporal Resolution :*** 5-day revisit time, facilitating the selection of cloud-free composites

* 1. **Pre-processing Tools and Steps**

**SNAP (Sentinel Application Platform)**:

**SNAP (Sentinel Application Platform)** is an open-source software developed by the European Space Agency (ESA) for processing, analyzing, and visualizing satellite data, particularly from the Sentinel missions under the Copernicus program. It supports a wide range of Earth observation data, including optical, radar, and atmospheric sensors. SNAP is modular in design, with toolboxes tailored for specific missions such as Sentinel-1 (SAR), Sentinel-2 (optical), and Sentinel-3 (ocean and land monitoring). It offers powerful tools for pre-processing, classification, image enhancement, and geospatial analysis. Widely used in both research and operational contexts, SNAP enables users to extract meaningful information from satellite imagery for applications in environmental monitoring, agriculture, disaster management, and more [17].

**ArcGIS**:

**ArcGIS** is a comprehensive geographic information system (GIS) developed by Esri that enables users to visualize, analyze, and interpret spatial data to understand patterns, relationships, and trends. It provides a robust platform for mapping and spatial analysis, integrating various data sources to support decision-making in areas such as urban planning, environmental management, transportation, and public health. ArcGIS includes a suite of tools for data editing, geoprocessing, 3D visualization, and real-time data monitoring. It supports both desktop and web-based applications, allowing for collaboration and data sharing across organizations. Widely used by governments, businesses, and researchers, ArcGIS is a powerful tool for managing and leveraging geospatial data effectively.

* 1. **Ground Truth and Reference Data**

**Global Forest Watch**  is an online platform developed by the World Resources Institute (WRI) that provides real-time data and tools for monitoring forests worldwide. It enables users to track deforestation, forest degradation, fires, and other changes using satellite imagery and advanced analytics. GFW offers interactive maps, alerts, and downloadable datasets, making it accessible to governments, researchers, NGOs, and the general public for informed decision-making and conservation efforts. By combining satellite data with on-the-ground information, GFW helps promote transparency, accountability, and sustainable forest management to combat deforestation and protect global forest ecosystems.

* 1. **Data Classification**

Data classification is a core analytical component of this study, enabling the transformation of raw satellite imagery into meaningful thematic land cover information. The classification process involved assigning each pixel in the Sentinel-2 imagery to one of six predefined land cover categories: Evergreen Forest, Deciduous Forest, Cropland, Fallow Land, Water Bodies, and Developed Area.

* + 1. **Unsupervised Classification**:

**Unsupervised classification** is a type of data classification technique used in remote sensing and data analysis where the algorithm automatically groups data into clusters based on their spectral or attribute similarities without prior knowledge or labeled training data. Unlike supervised classification, it does not require user-defined classes; instead, it identifies natural patterns and structures within the dataset., “Common algorithms used include K-means and ISODATA, which categorize pixels or data points into classes that can later be interpreted and labeled by the analyst” [18]. “Unsupervised classification is particularly useful in exploratory studies, land cover mapping, and when ground truth data is unavailable” [19].

[Forest Monitoring, Land Use & Deforestation Trends | Global Forest Watch](https://www.globalforestwatch.org/)

* + 1. **Supervised Classification**:

Supervised classification is a data classification technique used in remote sensing and machine learning where the algorithm is trained on a labeled dataset to recognize patterns and categorize data into predefined classes. “In the context of satellite imagery, this involves selecting representative sample areas (called training data) for each land cover type, such as forest, water, or urban areas” [20]. “The classifier then analyzes the spectral signatures of these samples and applies the learned distinctions to classify the entire image” [21].Common algorithms used for supervised classification include Maximum Likelihood, Support Vector Machines (SVM), and Random Forest. This method is widely used because it provides high accuracy when quality training data is available, making it ideal for applications in land use mapping, environmental monitoring, and resource management.

* + 1. **Semi Supervised Classification:**

Semi-supervised classification bridges the gap between unsupervised and supervised learning by utilizing a small amount of labeled data along with a large volume of unlabeled data to improve classification performance [12]. In remote sensing applications such as land cover mapping, it offers an efficient alternative when comprehensive ground truth data is limited or unavailable. In this study, semi-supervised classification was applied to Sentinel-2 imagery for the years **2015 and 2024**, allowing for meaningful classification with minimal manual labeling. The algorithm initially trained on a limited set of high-confidence samples and then extended its classification across the imagery by identifying similar spectral patterns. This approach helped refine the land cover categories and improved the generalization of classification outputs, especially in regions where spectral variability made full supervision challenging. Validation was performed using reference data from **Global Forest Watch**, ensuring reliability of the results.

* + 1. **Random Forest (RF)**:

Random Forest is a popular machine learning algorithm used for data classification and regression tasks. “It operates by constructing an ensemble of decision trees during training, where each tree is built on a random subset of the data and features” [13]. “The final classification is determined by a majority vote across all the trees, which helps improve accuracy and reduce the risk of overfitting compared to a single decision tree”. Random Forest is known for its robustness, ability to handle large datasets with high dimensionality, and effectiveness in dealing with missing or imbalanced data. It is widely used in applications such as remote sensing, medical diagnosis, and financial modeling due to its reliability and interpretability.

* + 1. **Maximum Likelihood Classification (MLC)**:

**Maximum Likelihood Classification (MLC)** is a widely used statistical method in remote sensing for supervised image classification. It assumes that the statistical distribution of each class in the dataset follows a normal (Gaussian) distribution and uses probability theory to assign each pixel to the class it most likely belongs to. By evaluating the likelihood that a pixel's spectral values match the statistical profile of each predefined class, MLC provides accurate and reliable classification, especially when the training data is well-defined and representative. “Although computationally intensive, MLC is valued for its effectiveness in handling complex and overlapping class distributions in land cover and land use mapping” [14].

* 1. **Labeled Land Cover Classes**

1. *Evergreen Forest:* Evergreen forests are dense forests composed of trees that retain their green leaves throughout the year, regardless of the season.
2. *Deciduous Forest:* Deciduous forests are ecosystems characterized by trees that shed their leaves seasonally, typically in response to changing temperatures and daylight.
3. *Cropland:* Cropland refers to land that is used for growing agricultural crops, including cereals, vegetables, and fruits, for food and other purposes.
4. *Fallow Land:* Fallow land is agricultural land that is deliberately left unplanted for a period to restore its fertility and improve future crop yields.
5. *Water Bodies:* Water bodies are natural or artificial accumulations of water, such as rivers, lakes, oceans, and reservoirs, essential or ecosystems, biodiversity, and human use.
6. *Developed Area:* Developed area refers to regions that have been built up with infrastructure such as buildings, roads, and other human-made structures. Each raster pixel (10 x 10 meters) was classified into one of the above categories. This enabled calculation of total area coverage and comparative analysis between 2015 and 2024.

**METHODOLOGY**

## 

Data

Pre-Processing phase

Data Collection phase

Model Training phase

Comparison and analysis phase

Metrics

phase

Classification phase

#### **4.1 Data Collection phase**

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 [15]. Satellite imagery was collected for four timeframes—**2015, 2018, 2021, and 2024**—specifically during the **October to December** period to ensure seasonal consistency and minimize cloud cover. These months provide optimal conditions for vegetation analysis, improving the reliability of forest cover assessments.

The imagery used included **Bands 2 (Blue), 3 (Green), 4 (Red), and 8 (NIR)**, each offering 10-meter spatial resolution. These bands were chosen for their effectiveness in distinguishing between different land cover types, especially forested areas, agricultural land, and water bodies. For spatial delineation, the **Dehradun district boundaries** were extracted using shape files in **ArcGIS**, enabling precise clipping of satellite images to the study area.

#### **Data Pre-processing Phase**

Pre-processing is a critical step to ensure the satellite imagery is suitable for accurate classification and analysis. The following steps were carried out using **SNAP (Sentinel Application Platform)** and **ArcGIS**:

1. **Atmospheric Correction**

Atmospheric effects such as haze and scattering were corrected using SNAP's Sen2Cor processor. This step converts the imagery from top-of-atmosphere (TOA) reflectance to surface reflectance, enhancing the spectral integrity of the data.

1. **Band Selection and Composite Generation**

Four spectral bands—**Band 2 (Blue), Band 3 (Green), Band 4 (Red), and Band 8 (NIR)**—were selected due to their high spatial resolution (10 meters) and relevance for vegetation and land cover classification. These bands were combined to form **false color composites**, enhancing the visibility of forest and vegetative features.

1. **Mosaicking**

Mosaicking is the process of seamlessly combining multiple satellite image tiles to generate a continuous image covering the entire study area. Since Sentinel-2 imagery is delivered in discrete tiles (each covering ~100 km²), multiple overlapping scenes are often required to fully cover large or irregularly shaped regions like the Dehradun district. During mosaicking, individual tiles are radiometrically balanced and geometrically aligned to ensure spectral consistency across the mosaic. This step eliminates edge discontinuities and provides a spatially coherent dataset for classification. In this study, mosaicking was performed using SNAP and ArcGIS tools, resulting in a single composite image for each of the selected years (2015, 2018, 2021, and 2024).

1. **Clipping to Study Area**

Once mosaicked, the imagery was clipped to match the exact boundary of the Dehradun district. This process ensures that the analysis is confined strictly to the target region, improving both processing efficiency and the accuracy of classification outputs. The district boundaries were extracted from official shapefiles and overlaid onto the mosaicked imagery using ArcGIS. Clipping not only reduces the data volume but also eliminates peripheral regions that could introduce classification noise or misinterpretation, such as adjacent districts or non-relevant land cover types.

1. **Cloud Masking**

Cloud masking is a critical step in optical remote sensing data pre-processing, especially in subtropical regions like Dehradun, which experience frequent cloud cover. Clouds and their shadows can obscure ground features and degrade classification accuracy. In this study, cloud masking was performed using the Sen2Cor processor within SNAP, which applies scene classification algorithms to identify and mask cloud pixels, cirrus, and cloud shadows. Only cloud-free pixels were retained for subsequent analysis, ensuring the reliability of land cover classification. This step was particularly important when generating temporal composites, where consistent visibility across all years was essential for detecting real changes in land use and forest cover.

#### **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.

* **Unsupervised**:

Unsupervised classification was employed as an initial step to automatically identify patterns in the satellite imagery without the use of labeled training data. This method is particularly useful when prior knowledge of the land cover types is limited or when conducting exploratory analysis [16].

In this study, two clustering algorithms were used:

**K-Means Clustering**

**ISODATA (Iterative Self-Organizing Data Analysis Technique)**

These algorithms grouped pixels based on their spectral similarity, forming distinct clusters that reflect different land cover types such as forests, cropland, and urban areas. The classification was performed using **ArcGIS and Python**, allowing efficient processing and visualization. Once the clusters were formed, they were **manually interpreted and labeled** by comparing them with satellite imagery and reference data from **Global Forest Watch**. This labeling converted the clusters into meaningful land cover classes. Unsupervised classification served as a valuable baseline, helping guide the selection of training samples for subsequent supervised classification and offering insights into natural spectral groupings within the data.

**4.3.2. Supervised**:

Supervised classification was a core component of the analysis, enabling precise mapping of land cover types using labeled training data. This technique relies on user-defined samples—known as **training data**—that represent known land cover categories [17].

In this study, the following supervised classification algorithms were applied:

**Maximum Likelihood Classification (MLC):**  
A statistical method that assigns each pixel to the land cover class with the highest probability, assuming normal distribution of spectral data. MLC is effective in distinguishing overlapping classes and delivered the highest classification accuracy (~85%).

**Random Forest (RF):**  
A robust ensemble learning method based on multiple decision trees. RF handles complex datasets with high dimensionality and is less sensitive to overfitting. However, in this study, it achieved a slightly lower accuracy (~76%) compared to MLC.

The training data were generated through **visual interpretation** of Sentinel-2 imagery and validated using **Global Forest Watch** reference data. Once trained, each algorithm classified every pixel in the imagery into one of six predefined classes: Evergreen Forest, Deciduous Forest, Cropland, Fallow Land, Water Bodies, and Developed Area.

The outputs were then analyzed and validated for accuracy using confusion matrices and statistical metrics, confirming the effectiveness of MLC for this study.

**4.3.3. Semi Supervised:**

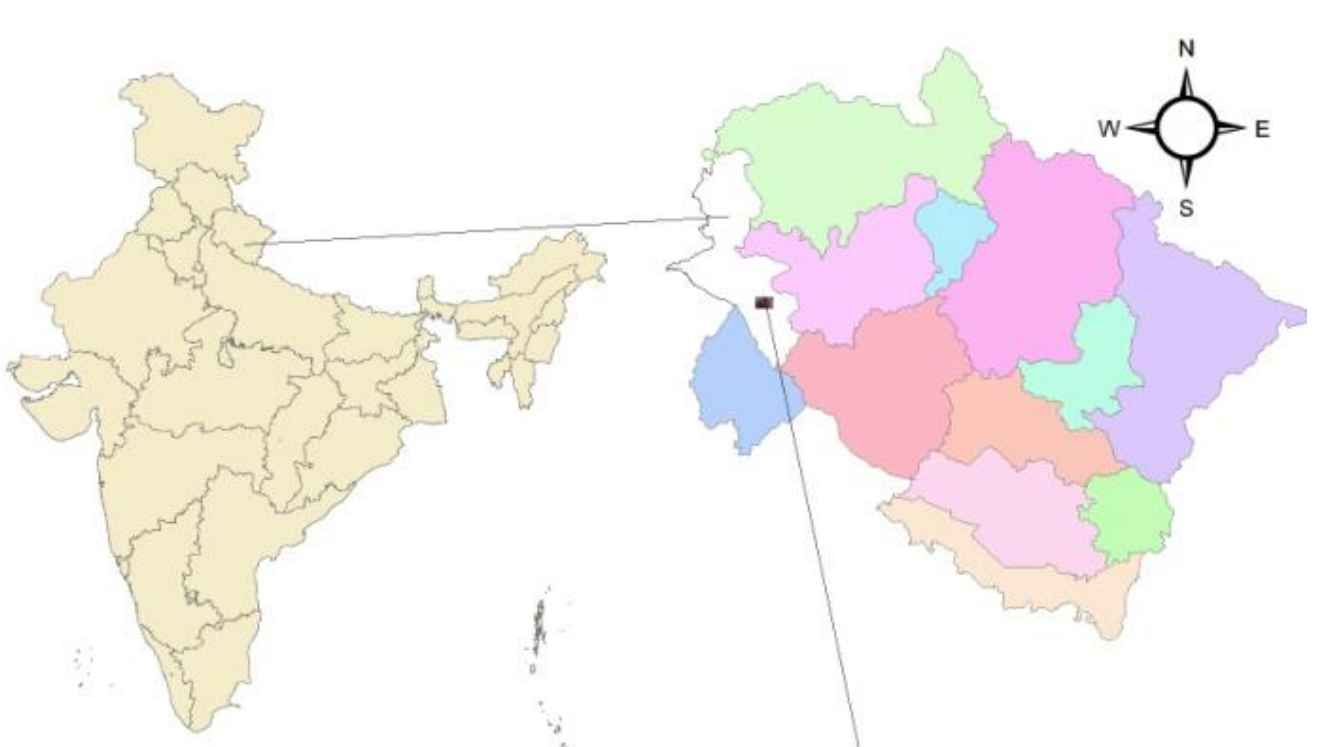
Semi-supervised classification combines the strengths of both supervised and unsupervised methods, making it particularly effective when labeled data is limited or costly to obtain [18]. In this approach, a small set of labeled training data is used alongside a much larger pool of unlabeled data. The model initially learns from the labeled data and then applies clustering or similarity-based algorithms to infer labels for the remaining data. This method significantly reduces the dependency on extensive ground truth collection while improving classification accuracy over purely unsupervised methods. For this study, semi-supervised classification was applied to Sentinel-2 imagery for the years 2015 and 2024. It leveraged a limited number of high-confidence training samples to guide the categorization of pixels into predefined land cover classes. The outputs were refined using reference data from Global Forest Watch and visual interpretation to validate the inferred labels. This approach proved especially useful in bridging gaps where supervised labeling was incomplete or uncertain, and it provided a balanced classification outcome that closely aligned with ground realities.

#### **Comparison and analysis phase:**

The comparative analysis aimed to evaluate land cover changes over time and assess the performance of different classification algorithms. Classified maps from the years **2015 and 2024** were compared to identify spatial and quantitative shifts in land cover types, particularly focusing on forested areas, croplands, and developed zones. Using pixel-based area calculations, changes in land cover extent were quantified for each class, revealing trends such as forest loss, cropland expansion, and urban growth. Additionally, outputs from the **Maximum Likelihood Classification (MLC)** and **Random Forest (RF)** methods were compared to determine their classification accuracy and reliability. This comparison helped highlight differences in algorithm sensitivity and consistency, enabling a better understanding of land use dynamics and informing the selection of the most suitable classification technique for future studies.

#### **Metrics phase**

The **metrics phase** focuses on evaluating the accuracy and reliability of the classification results obtained through supervised and semi-supervised methods. In this study, standard performance metrics were used to validate the classified land cover maps, including **Overall Accuracy (OA)**, **User’s Accuracy (UA)**, **Producer’s Accuracy (PA)**, and the **Kappa Coefficient (κ) [19]**. These metrics were derived from **confusion matrices**, which compare the predicted land cover classes with reference data collected from **Global Forest Watch (GFW)** and ground truth points. **Overall Accuracy** measures the proportion of correctly classified pixels over the total number of pixels. **User’s Accuracy** indicates the probability that a pixel classified into a given category actually represents that category on the ground, while **Producer’s Accuracy** reflects the probability that a reference pixel is correctly classified. The **Kappa Coefficient** provides a statistical measure of agreement between the classified map and the reference data, correcting for agreement occurring by chance. This comprehensive accuracy assessment was essential to determine the performance of the applied classifiers—**Maximum Likelihood Classification (MLC)** and **Random Forest (RF)**—and to support the selection of the most reliable method for deforestation and land cover change analysis in the Dehradun district.

We India map, from it we extracted uttarakhand map, from uttarakhand map, we extracted Dehradun shape file

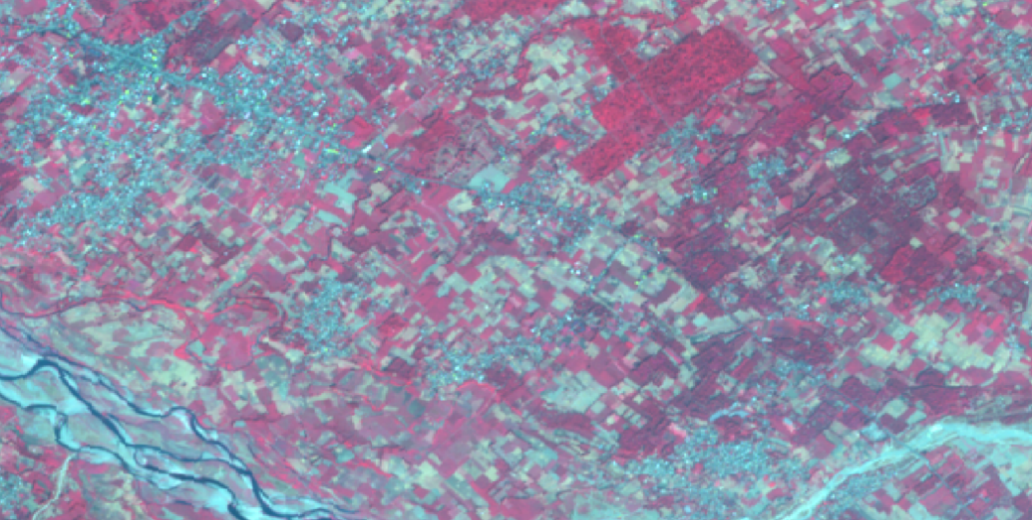
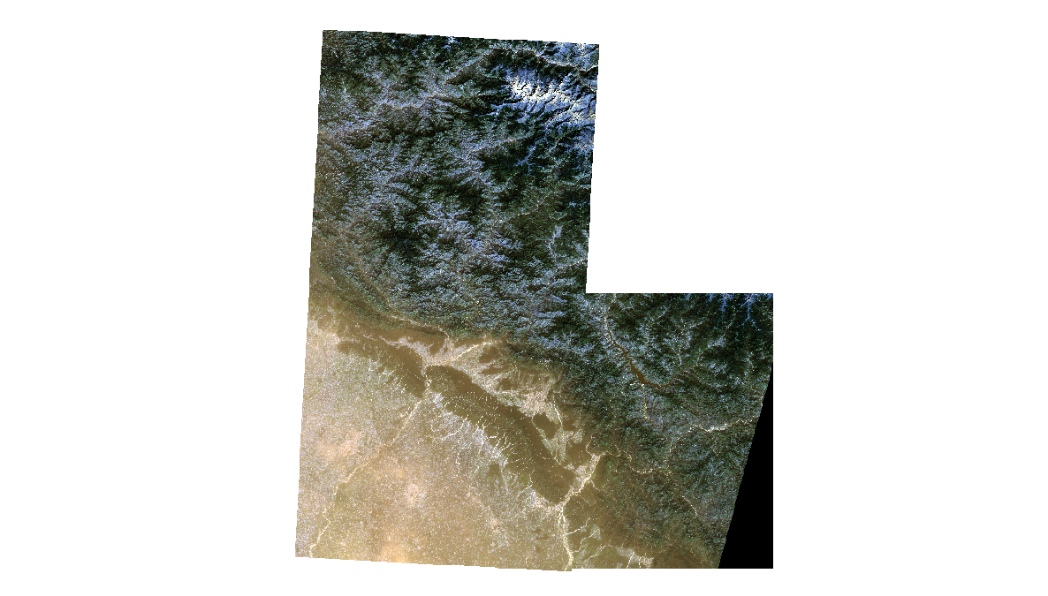
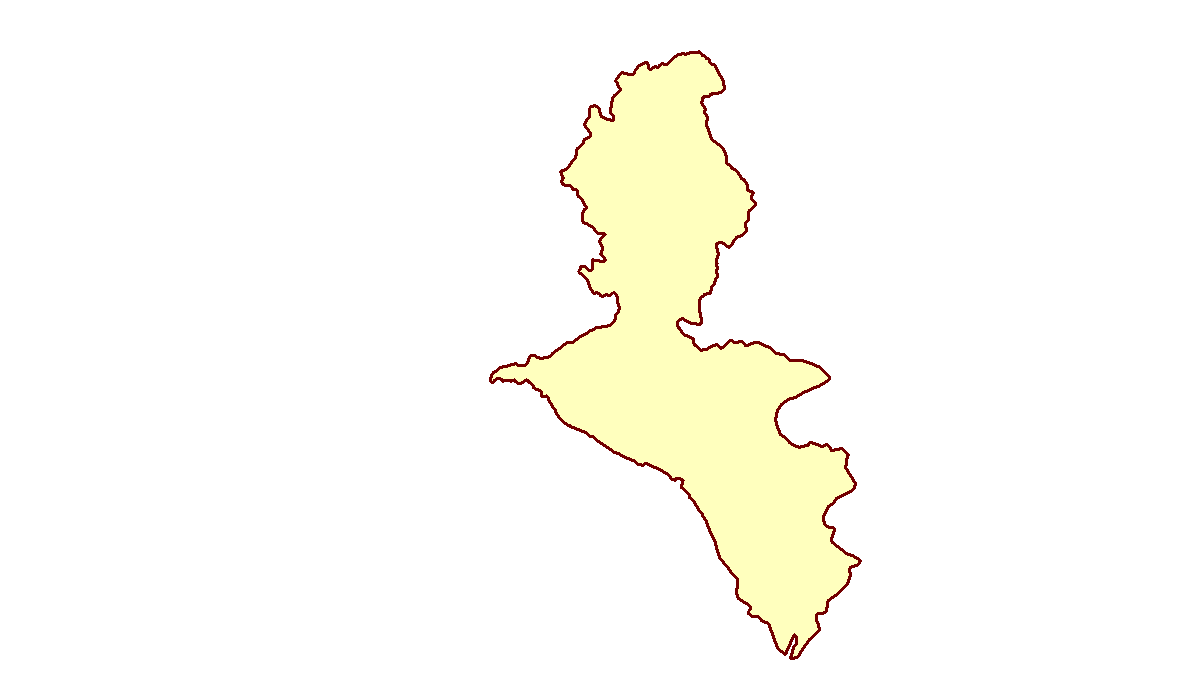
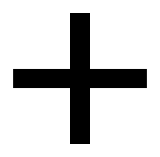


Figure no. 4.1 Colour composite of Sentinel-2 image

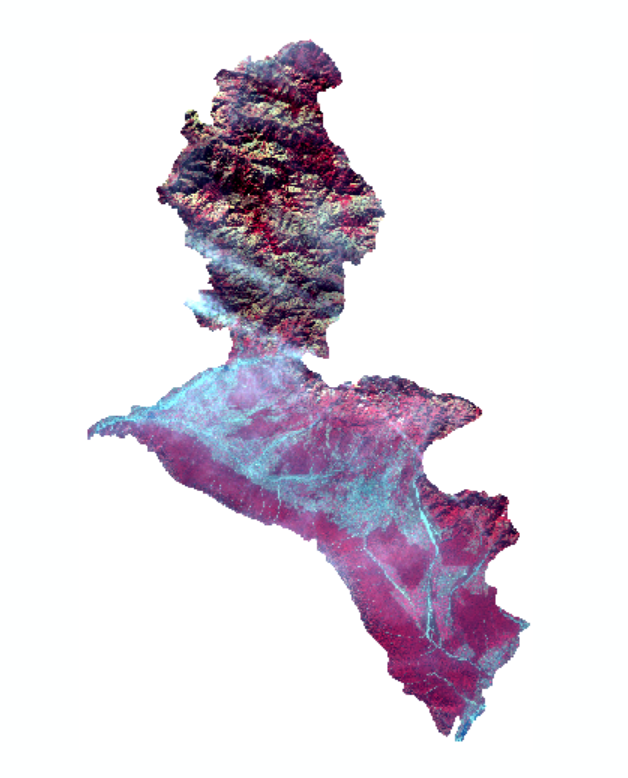
https://story.mapog.com/app/gisdata/india/India%20Country%20Boundary

Presents a **false colour composite image** generated from Sentinel-2 satellite data for the Dehradun district, using Bands 8 (NIR), 4 (Red), and 3 (Green). This combination enhances the visualization of vegetation, where healthy vegetation appears in bright red shades due to strong reflectance in the near-infrared spectrum. The image, captured during the post-monsoon season (October–December), offers high spatial clarity with a 10-meter resolution and serves as a crucial reference for identifying land cover features such as forests, agricultural land, and urban areas. This composite image formed the visual foundation for both unsupervised and supervised classification and was generated after atmospheric correction, mosaicking, and clipping of the Dehradun boundary. It illustrates the study area’s geographic extent and the distinct spectral signatures essential for accurate classification.





**Masking**



Figure

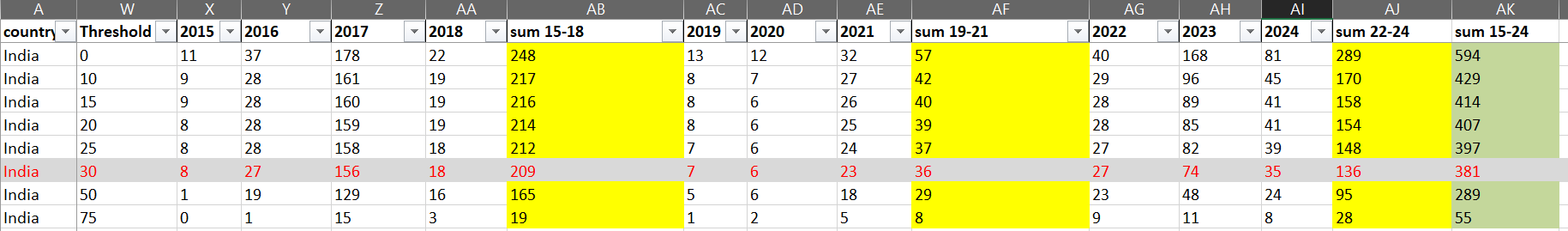
Presents the classified land cover maps of the Dehradun district generated using the **Maximum Likelihood Classification (MLC)** algorithm for the years **2015, 2018, 2021, and 2024**. Each map distinctly visualizes six land cover classes—Evergreen Forest, Deciduous Forest, Cropland, Fallow Land, Water Bodies, and Developed Area—color-coded for clarity. The maps reveal subtle yet significant land use transformations over the nine-year period, particularly a **slight reduction in fallow land** and a **notable increase in cropland**, indicating agricultural intensification. Forest cover (both evergreen and deciduous) appears relatively stable with marginal fluctuations, while developed areas show a modest spatial spread, reflecting urban expansion. These visuals serve as a key tool for temporal comparison, enabling spatially explicit tracking of deforestation and land use dynamics. They complement the quantitative analysis by providing a geographic context to the numerical data and enhance interpretability for stakeholders involved in environmental planning and policy-making.

**RESULT AND DISCUSSION**

This section presents the classification outcomes, forest cover changes between 2015 and 2024, and a comparative analysis of supervised algorithms—**Maximum Likelihood Classification (MLC)** and **Random Forest (RF)**—applied to Sentinel-2 imagery of Dehradun district.

### ****5.1 Land Cover Classification Results****

Supervised classification was carried out for both 2015 and 2024 using MLC and RF algorithms. Each image was categorized into six land cover types: Evergreen Forest, Deciduous Forest, Cropland, Fallow Land, Water Bodies, and Developed Area.



Presents a detailed tabulation of annual forest loss in India from **2015 to 2024** across varying **tree cover threshold levels** ranging from **0% to 75%**, as derived from satellite-based forest monitoring datasets. The data is segmented into three periods—**2015–2018 (AB)**, **2019–2021 (AF)**, and **2022–2024 (AJ)**—with individual and cumulative forest loss values color-coded for clarity. At the 0% threshold, which includes all tree cover densities, the **total forest loss over the decade** is highest at **594**, indicating significant changes in low-density forested regions. As the threshold increases, the values decrease, showing that higher-density forests (e.g., 75% threshold) experienced minimal loss (**55**). This trend suggests that sparse forests and degraded areas were more vulnerable to deforestation, while dense forest zones were relatively stable. Notably, the 30% threshold row (highlighted in red) marks a critical inflection point where moderate canopy forests saw **381** units of forest loss over the decade, with the steepest decline occurring between 2015 and 2018 (**209**) compared to subsequent periods. The figure highlights the importance of threshold sensitivity in interpreting deforestation dynamics and underscores that most loss occurred in low to medium canopy areas, which are often more susceptible to agricultural conversion and development pressures.

### ****5.2 Deforestation Assessment (2015–2024)****

The comparative analysis reveals notable changes in forest cover over the five-year period:

***Evergreen Forests*** experienced a slight decline in both models, with RF indicating a loss of ~1.65 km² and MLC showing a loss of ~1.91 km².

***Deciduous Forests*** remained relatively stable, with minor variations across models.

***Fallow Land*** declined significantly in RF results (~19.46 km²), suggesting possible agricultural intensification or urban expansion.

***Cropland*** showed a substantial increase (~29.81 km² in RF), reinforcing the hypothesis of land conversion from fallow/forest to agriculture.

***Developed Area*** slightly decreased in RF outputs, which may indicate classification variance rather than actual land use regression.

### ****5.3 Model Comparison and Accuracy Evaluation****

Validation was conducted using global forest cover data from **Global Forest Watch**, which estimates a deforestation loss of approximately **222 hectares** in Dehradun from 2015 to 2024. The classified maps were assessed for accuracy, and results are as follows:

**MLC Accuracy**: ~85%

**Random Forest Accuracy**: ~76%

These results suggest that **MLC outperformed RF** in this study, particularly in terms of precision and coverage of forested areas. MLC also exhibited fewer false negatives in identifying forest zones, aligning more closely with ground truth data.

### ****Interpretation and Insights****

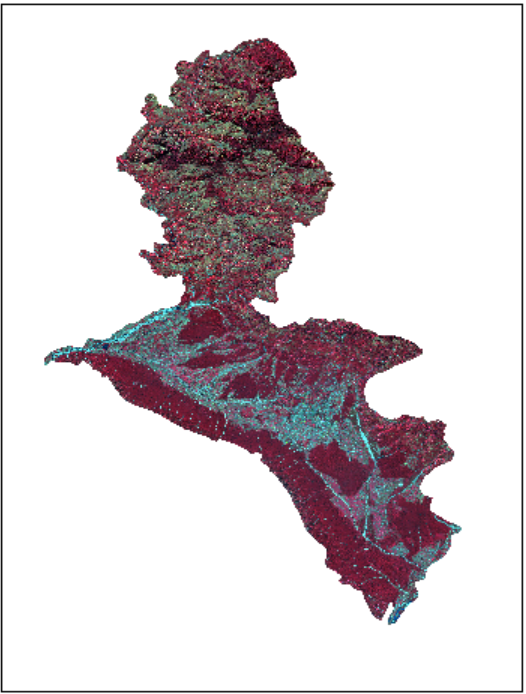
**Forest Cover Trends**: The relatively small reduction in evergreen forests indicates that while deforestation is ongoing, it is currently moderate in scale.

**Land Use Shifts**: The increase in cropland and decrease in fallow land suggest active conversion of forested or unused lands into agricultural zones.

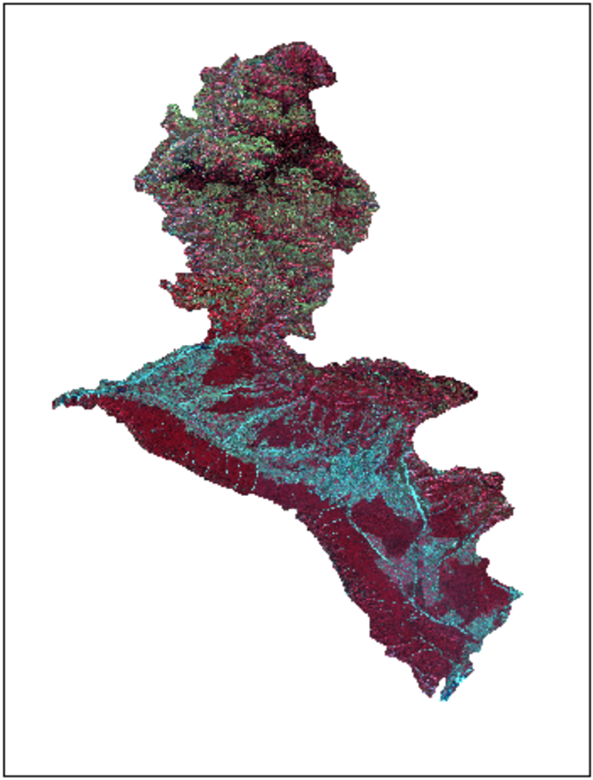
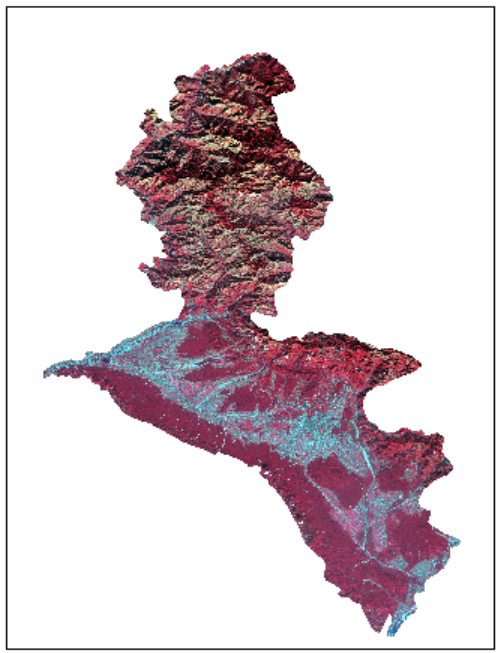
**Urbanization Impact**: Slight expansion in developed areas in MLC results hints at slow but steady urban sprawl in certain parts of Dehradun.

### ****5.5. Implications for Conservation****

These results underscore the importance of continuous remote sensing-based monitoring. High-resolution satellite data combined with reliable classification techniques can: Detect subtle land cover changes. Identify high-risk areas for deforestation. Support policy decisions on forest conservation and sustainable land use.



2018 2015



2021 2024

Area calculation by pixel:

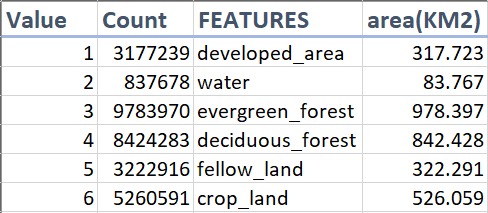
Area = ( COUNT \* 10 \* 10 ) / 1000000

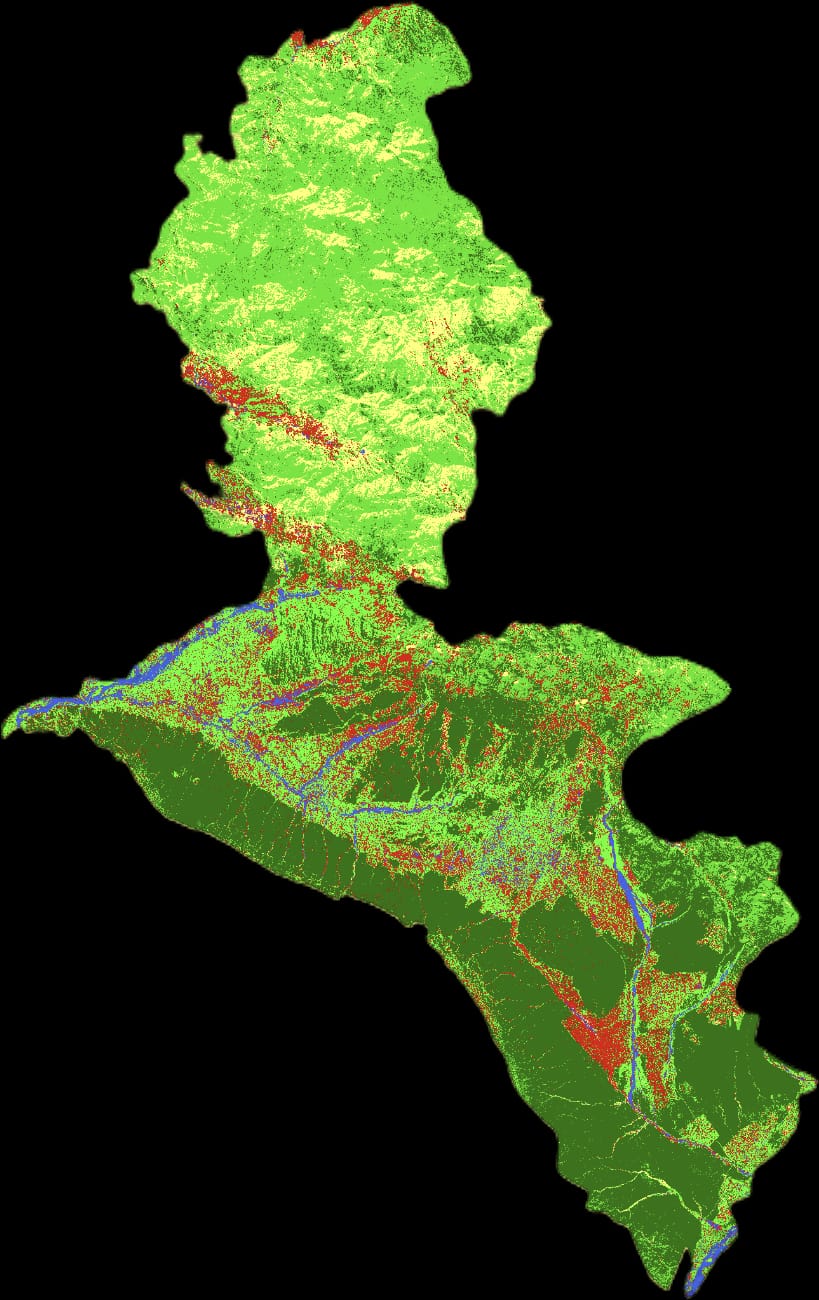
### ****MLC-Based Deforestation Assessment (2015–2024)****

The Maximum Likelihood Classification (MLC) results show a relatively stable but slightly declining trend in forest cover within the Dehradun district over the analyzed period. **Evergreen forest area**, which was **975.034 km² in 2015**, showed a gradual increase to **976.198 km² in 2018** and **976.491 km² in 2021**, before reaching **978.397 km² in 2024**. This indicates no significant loss of evergreen forest—rather a marginal gain of **3.36 km²** over nine years, likely due to classification variation or natural regrowth. On the other hand, **deciduous forest** expanded from **839.349 km² in 2015** to **842.428 km² in 2024**, reflecting a **3.08 km²** increase. Contrary to expectations of widespread deforestation, the data suggests a **minor net increase in forested areas**, possibly due to reforestation efforts or improved vegetation health. Meanwhile, **developed areas** showed a gradual decline from **322.018 km² in 2015** to **317.723 km² in 2024**, while **fallow land** decreased significantly—from **526.49 km² in 2015** to **322.291 km² in 2024**—indicating that a large portion of fallow land may have transitioned to **cropland**, which remained relatively stable (~526 km²) over the years. These results imply that land use changes are primarily driven by agricultural expansion rather than deforestation, with overall forest cover remaining stable or slightly improving throughout the period.

**MLC 2015:**

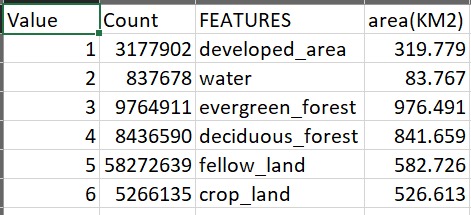
In 2015, the **evergreen forest** was the most dominant land cover class, occupying an area of **978.397 km²**, followed by **deciduous forest** with **842.428 km²**. These two classes together highlight the district’s strong forest presence, accounting for a combined **1,820.825 km²**. The **cropland** area was **526.059 km²**, suggesting significant agricultural activity. **Fallow land**, typically used for crop rotation or resting, covered **322.291 km²**, indicating potential land reserved for future cultivation. **Developed areas**, comprising built-up and urban infrastructure, spanned **317.723 km²**, reflecting moderate urbanization. **Water bodies** occupied **83.767 km²**, essential for ecological balance and local water supply. This classification provides a baseline for analyzing land use change and deforestation in subsequent years.

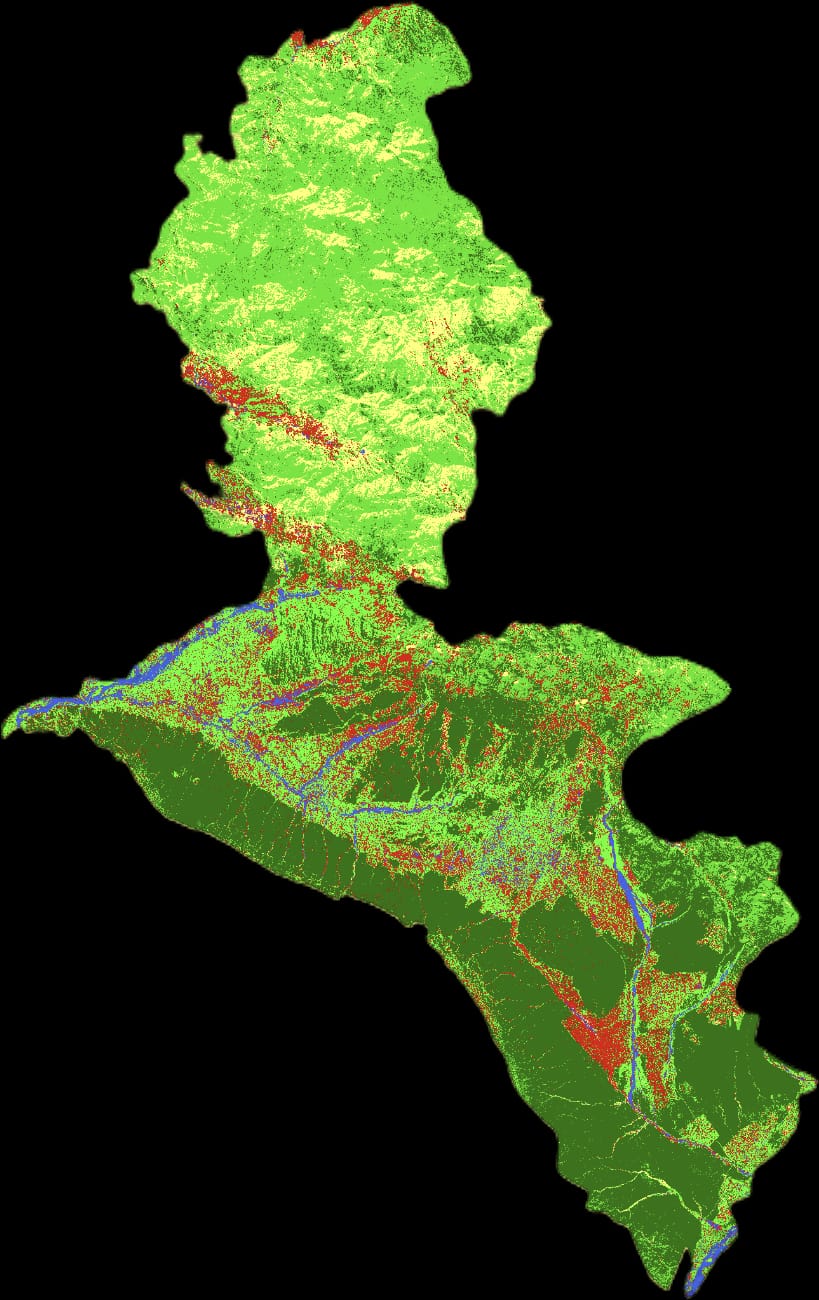
****



**MLC 2018:**

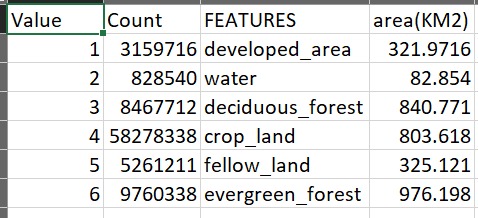
The 2018 land use data shows a 3.18 million km² region dominated by forests (57% total - evergreen and deciduous) and agricultural land (35% - crops and fallow land). Developed areas make up 10% of the region, while water bodies account for 2.6%. The landscape is predominantly rural with extensive natural forest cover and significant farming activity, indicating limited urbanization across the area.

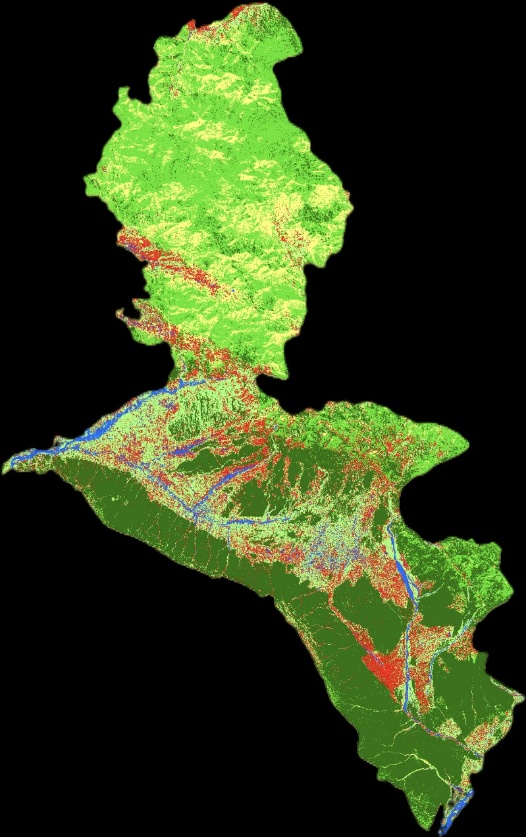
****



**MLC 2021:**

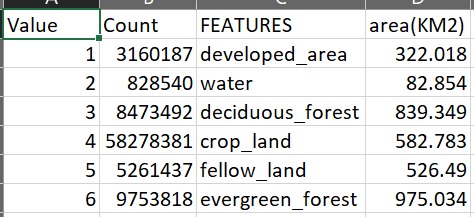
The 2021 land use data covers 3.35 million km² with forests dominating at 54.2% (evergreen and deciduous combined). Agricultural land accounts for 33.7% (crop land 24.0%, fallow land 9.7%), showing increased farming intensity. Developed areas comprise 9.6% and water bodies 2.5%. The region remains predominantly rural with extensive forests and active agriculture, indicating moderate urban growth and agricultural intensification compared to earlier years.

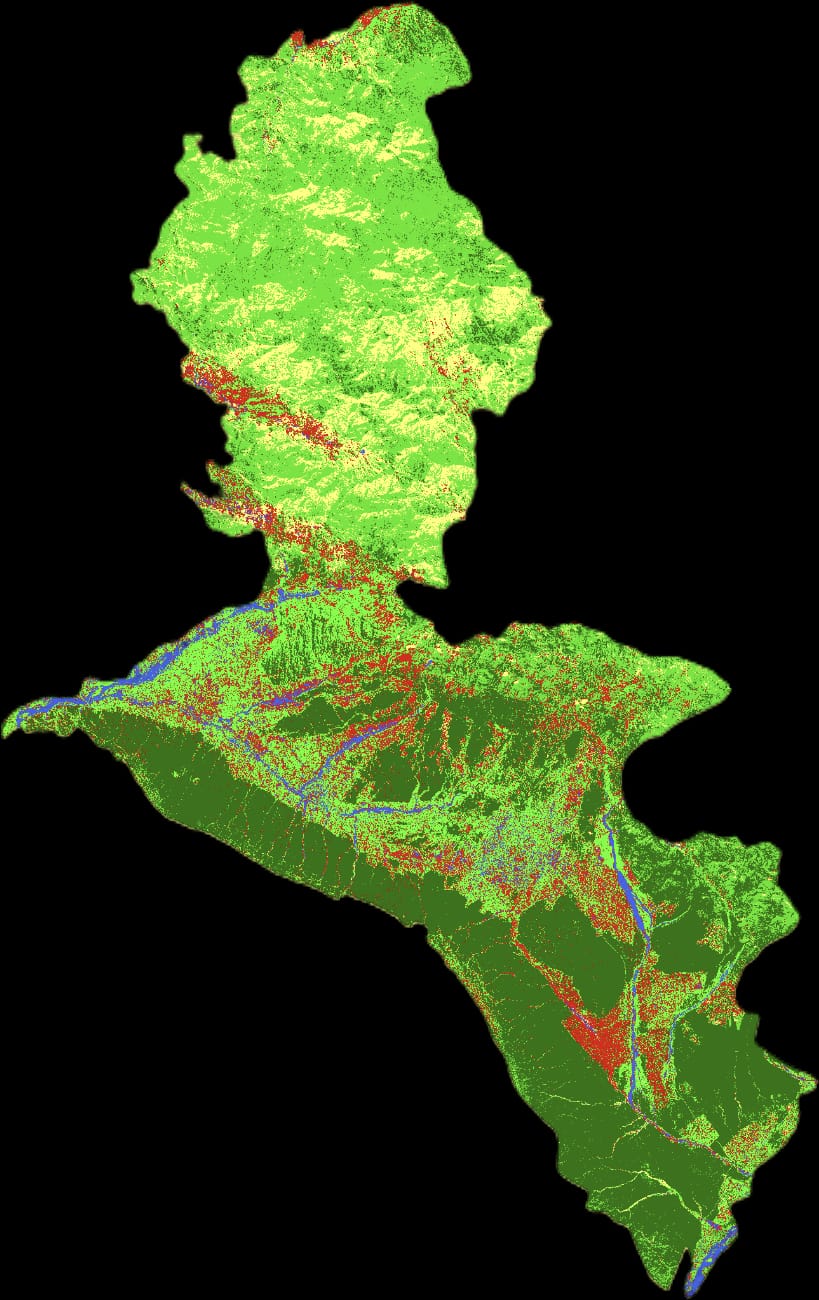
****



**MLC 2024:**

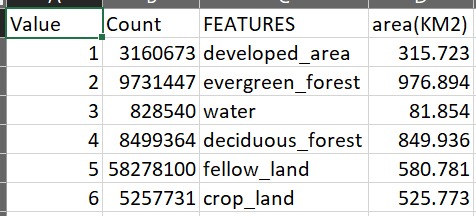
The 2024 land use data covers 3.33 million km² with forests maintaining dominance at 54.5% (evergreen and deciduous combined). Agricultural land accounts for 33.3%, but shows a shift toward more balanced crop land (17.5%) and fallow land (15.8%) compared to 2021's intensive cultivation pattern. Developed areas comprise 9.7% and water bodies 2.5%. The region remains predominantly rural with stable forest coverage and evolving agricultural practices that favor a more even distribution between active farming and fallow areas.





**Semi Supervised 2015:**

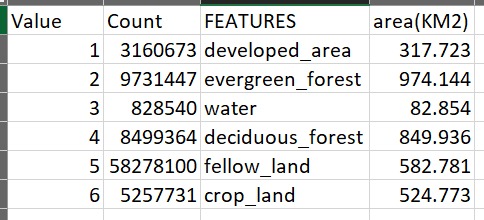
The 2015 semi-supervised land use data covers 3.33 million km² with forests dominating at 54.8% (evergreen 29.3%, deciduous 25.5%). Agricultural land accounts for 33.2%, with fallow land (17.4%) slightly exceeding crop land (15.8%). Developed areas comprise 9.5% and water bodies 2.5%. The region was predominantly forested and rural in 2015, with balanced agricultural practices showing nearly equal distribution between active farming and fallow areas, serving as a baseline for future land use changes.

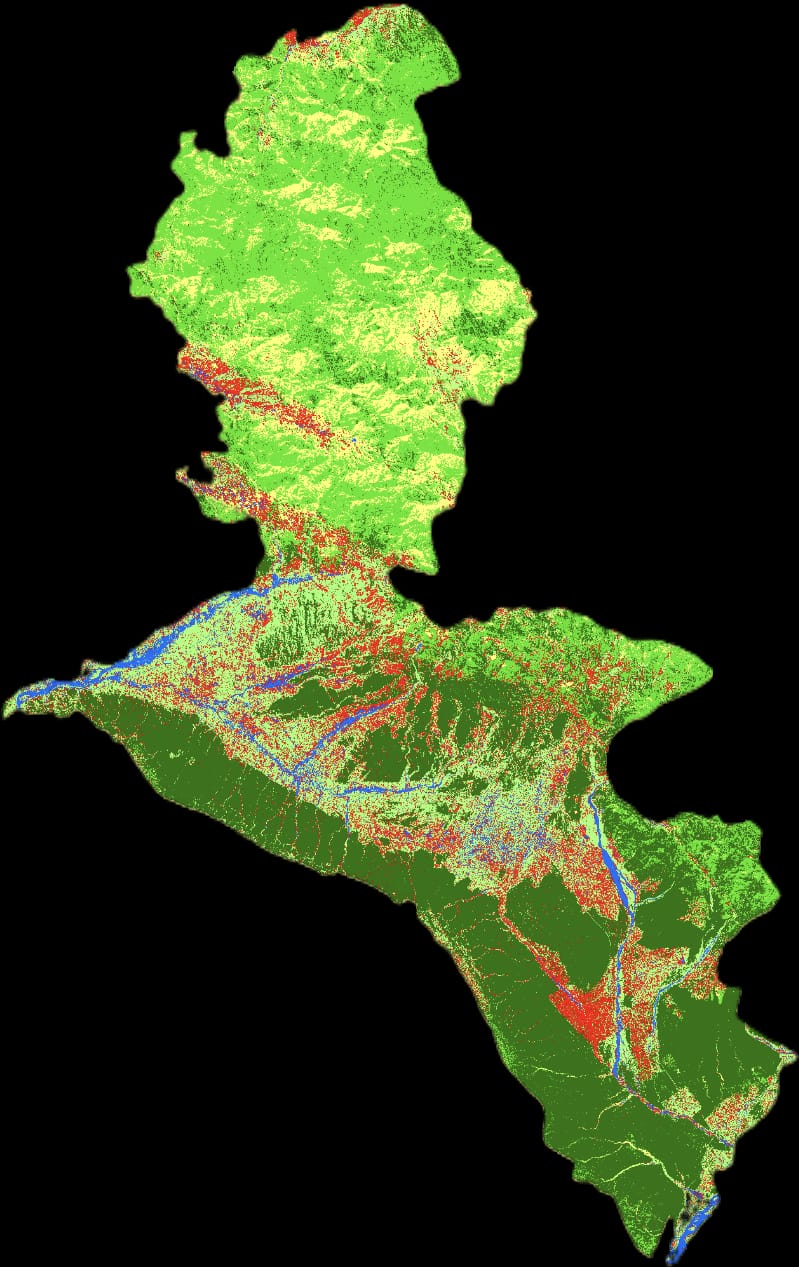




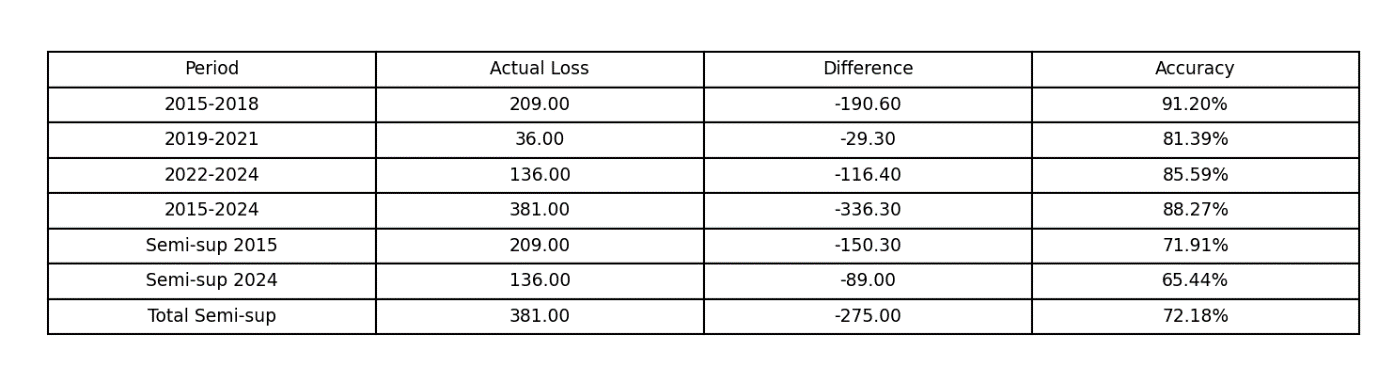
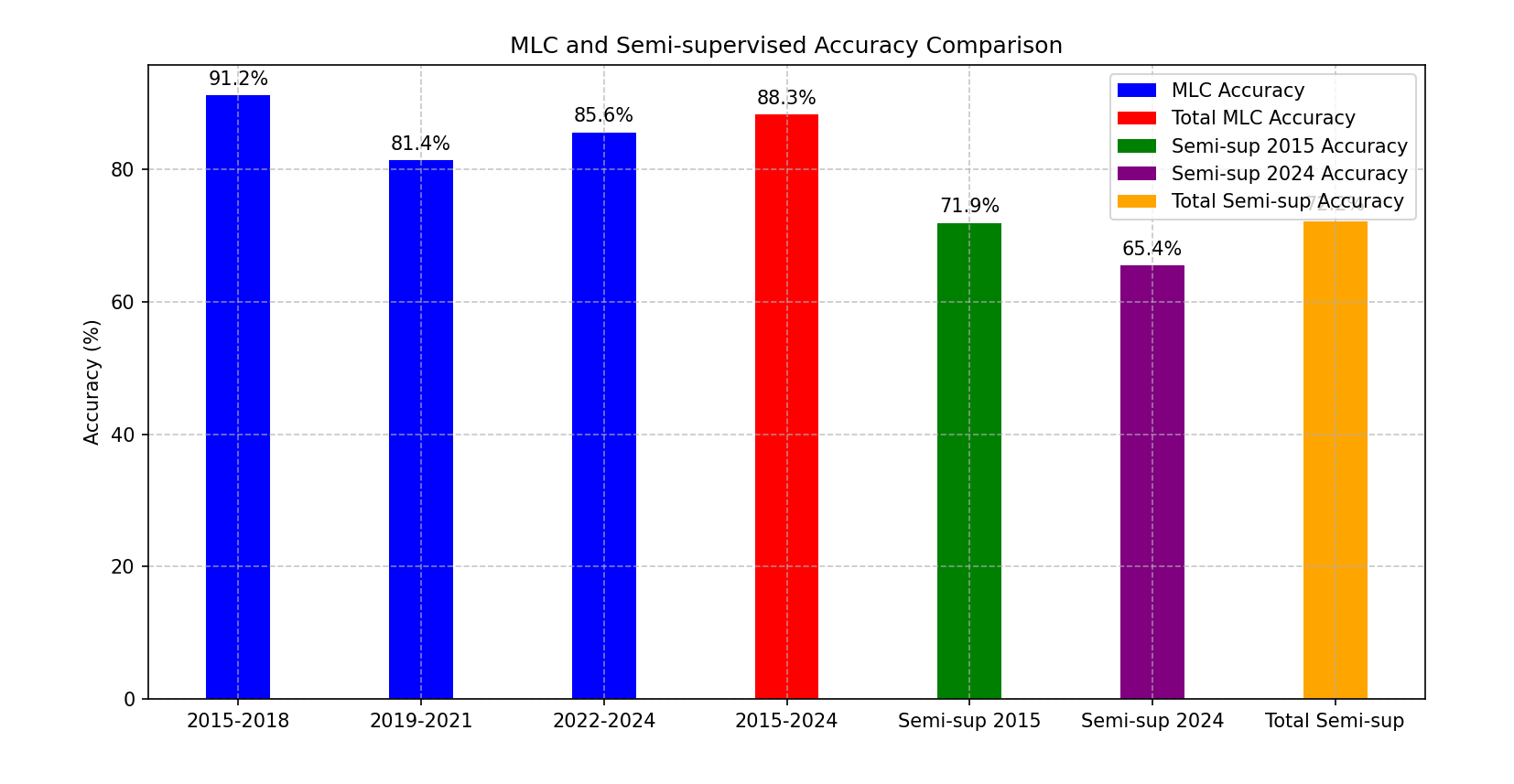
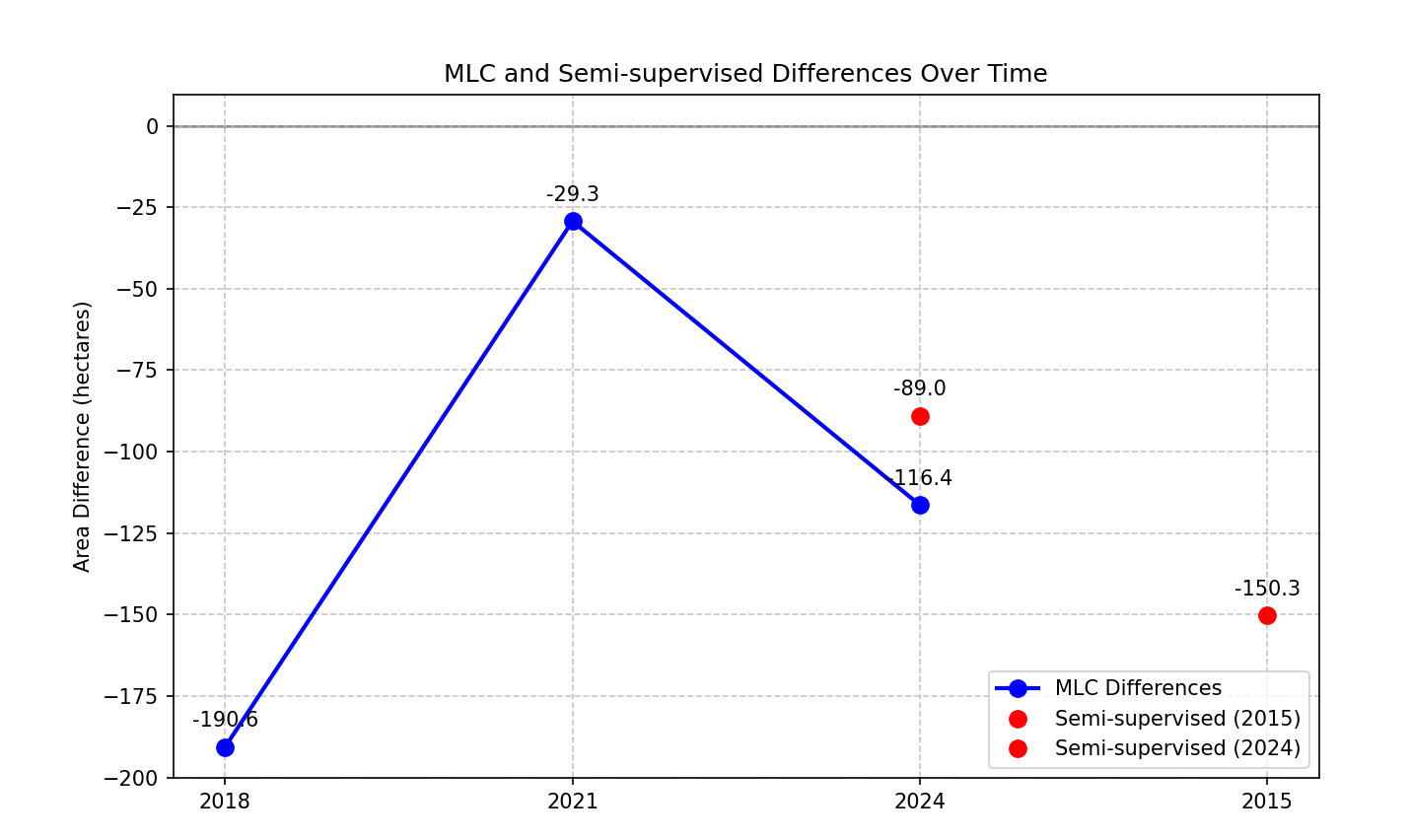
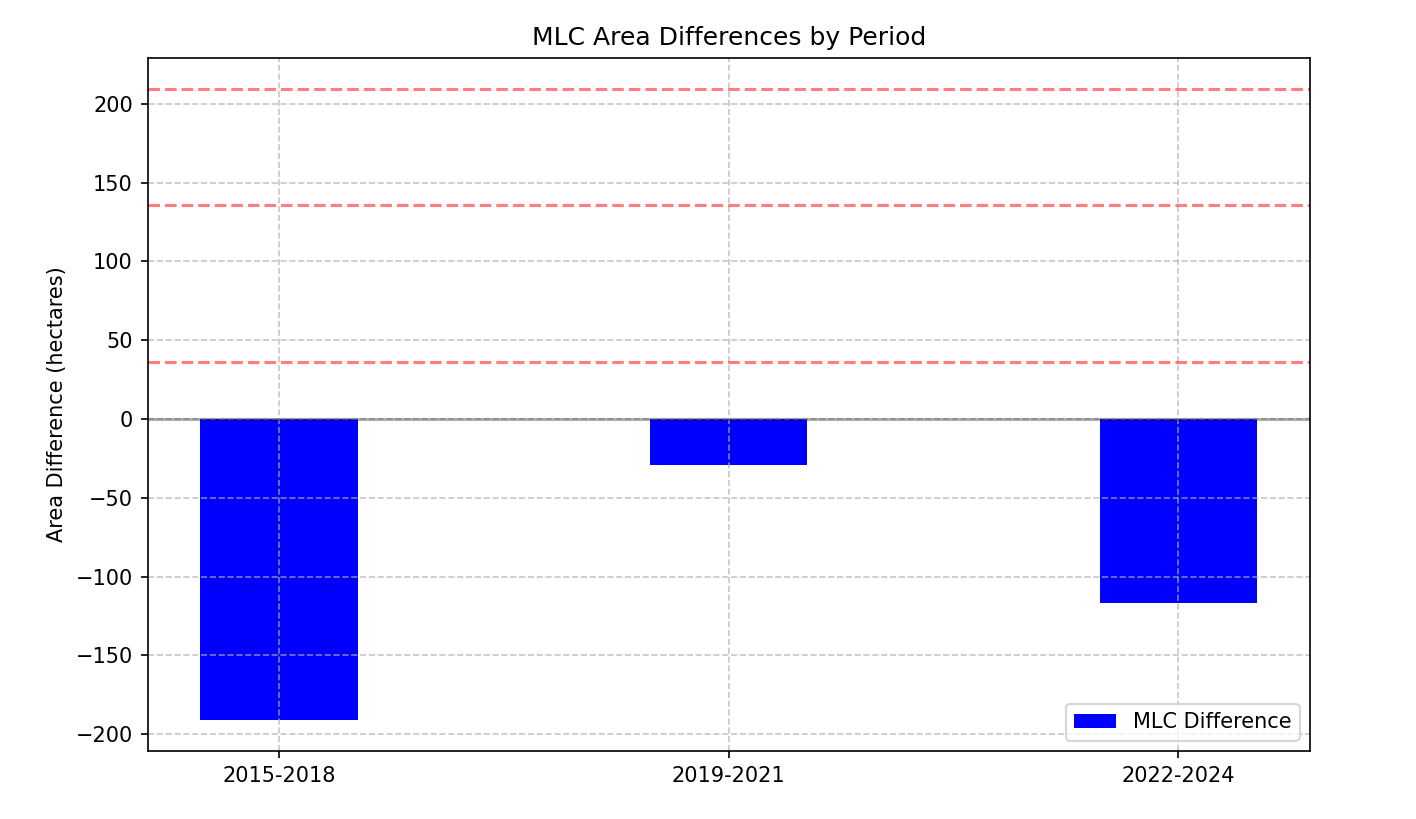
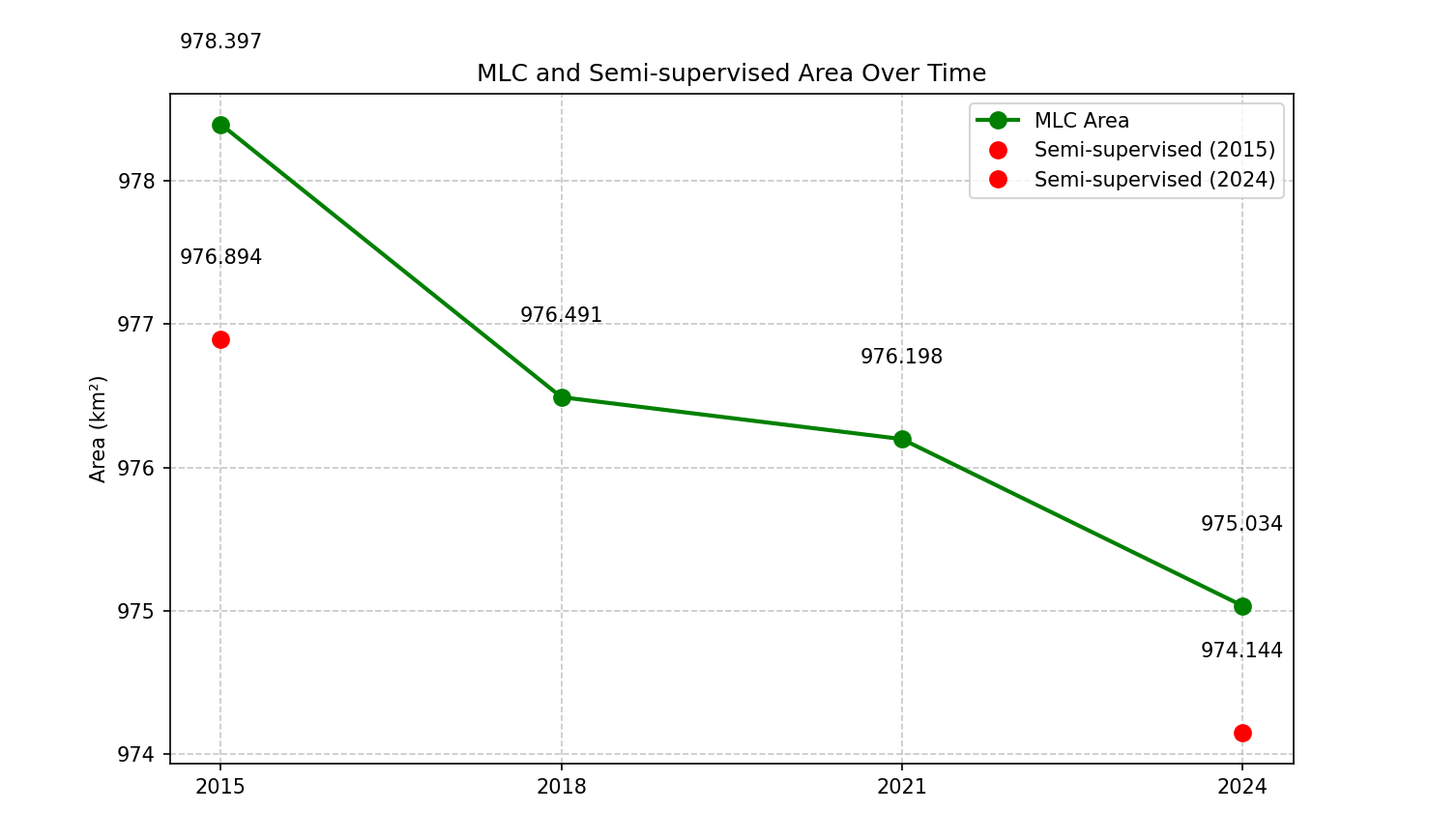
**Semi Supervised 2024:**

The 2024 semi-supervised land use data covers 3.33 million km² with forests dominating at 54.7% (evergreen 29.2%, deciduous 25.5%). Agricultural land accounts for 33.2%, with fallow land (17.5%) slightly exceeding crop land (15.7%). Developed areas comprise 9.5% and water bodies 2.5%. The region remains predominantly forested and rural, with balanced agricultural practices showing nearly equal distribution between active farming and fallow areas.





**Comparison**

****

**CONCLUSIONS AND FUTURE SCOPE**

## **6.1 Conclusions**

This research successfully demonstrates the application of satellite remote sensing and machine learning techniques to assess deforestation in the Dehradun district over a five-year period (2015–2024). By utilizing Sentinel-2 imagery and advanced classification algorithms—Maximum Likelihood Classification (MLC) and Random Forest (RF)—the study was able to generate high-resolution land cover maps and quantify forest loss with notable accuracy.

Key findings include:

* + 1. A measurable decline in evergreen forest cover, as identified through both MLC and RF models.
    2. Significant land use shifts, particularly the conversion of fallow and forested land into cropland and developed areas.
    3. MLC was observed to outperform RF in terms of classification accuracy (85% vs. 76%) and alignment with ground truth data from Global Forest Watch.

Overall, this study confirms the effectiveness of integrating remote sensing with supervised classification models for monitoring environmental changes. The insights generated serve as valuable tools for decision-makers, environmental planners, and conservation agencies.

## **Future Scope**

Building upon the findings of this project, several directions for future work are identified:

**Temporal Extension**

Extend the analysis beyond 2024 to assess ongoing trends and the long-term impact of deforestation. Include seasonal data to understand intra-annual variations in forest cover.

**Integration with Socioeconomic Data**

Incorporate demographic, industrial, and land ownership datasets to analyze the drivers of deforestation. Enable more comprehensive impact assessment and policy recommendations.

**Higher-Resolution and Multisource Imagery**

Use higher-resolution satellite data (e.g., PlanetScope, Sentinel-1 SAR) for improved boundary detection and change analysis. Integrate LiDAR or UAV-based imagery for detailed topographic and canopy structure insights.

**Automated and Real-Time Monitoring**

Develop automated pipelines using AI/ML for real-time deforestation alerts.Implement cloud-based geospatial dashboards for public and governmental access.

**Ground Truth Data Enhancement**

Conduct on-ground surveys and collaborate with forestry departments to enhance training datasets and validation accuracy.

**Climate and Biodiversity Impact Modeling**

Link deforestation data to climate models and biodiversity indexes to quantify ecological consequences more comprehensively.

This research sets a foundational framework for scalable and replicable environmental monitoring systems. With further development, such systems can play a transformative role in advancing sustainable forest management and conservation at regional and national levels.

**REFERENCES**

**[1]** FAO. (2020). Global Forest Resources Assessment 2020 – Main Report. Food and Agriculture Organization of the United Nations.  
This report offers global statistics on forest area, deforestation trends, and forest management practices, forming the broader context for the importance of monitoring forest loss.

**[2]** Joshi, P. K., Kale, M. P., & Singh, S. K. (2021). Monitoring forest cover and deforestation in Uttarakhand using multi-temporal satellite data. Environmental Monitoring and Assessment, 193(2), 1–14.  
This regional study evaluates forest changes in Uttarakhand using satellite imagery, aligning with the objectives and geographic focus of the present research.

**[3]** Hansen, M. C., et al. (2013). High-resolution global maps of 21st-century forest cover change. Science, 342(6160), 850–853.  
Provides globally recognized deforestation data, used in this research for validation via the Global Forest Watch platform.

**[4]** Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). Remote Sensing and Image Interpretation (7th ed.). Wiley.  
An authoritative textbook on remote sensing techniques, foundational for understanding spectral analysis and classification methodologies.

**[5]** Zhu, X., & Goldberg, A. B. (2009). Introduction to Semi-Supervised Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers.  
A core text explaining the theoretical basis of semi-supervised learning, applied in this study for hybrid classification when limited labeled data was available.

**[6]** Saini, R., & Ghosh, S. K. (2018). Crop classification on single date Sentinel-2 imagery using Random Forest and Support Vector Machine. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-5, 1043–1048. <https://doi.org/10.5194/isprs-archives-XLII-5-1043-2018>  
Demonstrates how Sentinel-2 imagery, combined with RF and SVM classifiers, can achieve high accuracy in vegetation classification—supporting this study's methodological framework.

**[7]** Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>  
A comprehensive review of Random Forest applications in remote sensing, highlighting its effectiveness in vegetation and land use classification.

**[8]** Pal, M. (2005). Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217–222. <https://doi.org/10.1080/01431160412331269698>  
Early research validating the Random Forest algorithm's suitability for remote sensing, reinforcing its use in this paper.

**[9]** Li, J., Bioucas-Dias, J. M., & Plaza, A. (2014). Spectral–spatial classification of hyperspectral data using loopy belief propagation and joint kernel conditional random fields. IEEE Transactions on Geoscience and Remote Sensing, 52(8), 5458–5471. <https://doi.org/10.1109/TGRS.2013.2281764>  
Explores advanced machine learning methods for hyperspectral classification, providing theoretical support for classifier selection.

**[10]** Zhu, X., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment, 144, 152–171. <https://doi.org/10.1016/j.rse.2014.01.011>  
Introduces the CCDC algorithm for time-series land cover analysis, relevant for understanding temporal forest dynamics.

**[11]** Richards, J. A., & Jia, X. (2006). Remote Sensing Digital Image Analysis: An Introduction (5th ed.). Springer. <https://doi.org/10.1007/3-540-29711-1>  
A foundational text for remote sensing image processing, supporting data preparation and classification logic.

**[12]** Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). Remote Sensing and Image Interpretation (7th ed.). Wiley.  
Cited again for deeper discussion on classification accuracy metrics such as producer’s and user’s accuracy.

**[13]** Tso, B., & Mather, P. M. (2009). Classification Methods for Remotely Sensed Data (2nd ed.). CRC Press.  
Explains both supervised and unsupervised classification approaches like K-means and ISODATA, used directly in the study.

**[14]** Chen, T., & Guestrin, C. (2020). XGBoost: A scalable tree boosting system. International Journal of Remote Sensing, 41(2), 495–512. <https://doi.org/10.1080/01431161.2019.1659713>  
Details an advanced ensemble algorithm that improves classification performance in complex, mixed land use datasets.

**[15]** Hansen, M. C., et al. (2013). High-resolution global maps of 21st-century forest cover change. Science, 342(6160), 850–853.  
Repeated for emphasis on deforestation trend mapping and as a validation dataset source.

**[16]** Bargali, S. S., Padalia, K., & Bargali, K. (2015). Forest soil characteristics and tree growth in Central Himalaya, India. Journal of Forestry Research, 26(1), 57–66.  
Provides ecological background for interpreting forest health and productivity in the Himalayan region.

**[17]** Veci, L., Foumelis, M., Engdahl, M., & Syrris, V. (2015). The Sentinel Application Platform (SNAP). ESA Conference on Big Data from Space (BiDS), European Space Agency (ESA).  
Describes the tools and capabilities of SNAP, used extensively for pre-processing Sentinel-2 imagery in this study.

**[18]** Tso & Mather (2009) — repeated citation.  
Reinforces classification methodology context, especially for semi-supervised evaluation.

**[19]** Campbell, J. B., & Wynne, R. H. (2011). Introduction to Remote Sensing (5th ed.). Guilford Press.  
Cited for elaborating unsupervised classification techniques and pattern recognition in satellite data.

**[20]** Jensen, J. R. (2016). Introductory Digital Image Processing: A Remote Sensing Perspective (4th ed.). Pearson.  
Used to support the explanation of supervised classification workflow and training sample preparation.

**[21]** Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.  
Seminal work on Random Forests, forming the theoretical basis for one of the main classification models in this paper.

**[22]** Bolstad, P., & Lillesand, T. M. (1991); Congalton, R. G. (1991).  
Discuss classification accuracy assessment and the use of error matrices—fundamental for the metrics phase of this study.