# Digital Image Processing Techniques for Land Cover Classification Using Satellite Imagery

**Mahlet Nigussie Tesfaye**
Addis Ababa Institute of Technology, SiTE
rigbe.rmn@gmail.com
Addis Ababa, Ethiopia.
July 09- 2024

## 1 Introduction

### 1.1 Overview

In recent years, the ability to monitor and analyze land cover changes has become increasingly crucial for a range of applications including environmental monitoring, urban planning, and agriculture management. Satellite imagery provides a valuable resource for this purpose, offering a comprehensive view of the Earth's surface over time. This project focuses on applying Digital Image Processing (DIP) techniques to classify land cover types using high-resolution satellite images.

### 1.2 Motivation and Goals

The primary motivation for this project is to develop an efficient and accurate method for classifying land cover from satellite imagery. Accurate land cover classification helps in understanding land use patterns, assessing environmental changes, and supporting decision-making processes in various domains. The goals of this project include:

1. Developing a Robust Classification Model: Implementing and evaluating different Digital Image Processing techniques to classify different land cover types such as forests, urban areas, water bodies, and agricultural lands.
2. Enhancing Classification Accuracy: Utilizing advanced feature extraction methods and classification algorithms to improve the precision of land cover classification.
3. Creating a Comprehensive Workflow: Designing a systematic approach for processing satellite images, extracting relevant features, training classification models, and validating results.

### 1.3 Related Work

Previous studies have explored various methods for land cover classification, including traditional statistical approaches and modern machine learning techniques. Many of these approaches leverage spectral indices, texture features, and advanced classification algorithms to achieve accurate results. However, the integration of different DIP techniques and their impact on classification performance remains an area of active research. This project builds upon these existing methodologies, incorporating a range of Digital Image Processing techniques to address the challenges associated with land cover classification from satellite imagery.

### 1.4 Contributions

This project contributes to the field of remote sensing and land cover classification by:

- Demonstrating the application of Digital Image Processing techniques in a practical setting.
- Comparing different feature extraction methods and classification algorithms to identify the most effective approach for land cover classification.

- Providing a detailed workflow that can be utilized by researchers and practitioners working with satellite imagery for land cover analysis.

## 2 Problem Statement

### 2.1 Challenge Overview

The rapid advancement of satellite imaging technology has made it possible to acquire high-resolution images of the Earth's surface. However, accurately classifying different types of land cover from these images remains a significant challenge. Land cover classification is essential for various applications, including environmental monitoring, urban planning, and resource management. Traditional methods of land cover classification often struggle with issues such as variability in land cover types, differences in image quality, and the complexity of the landscapes.

### 2.2 Specific Problems

1. High Variability in Land Cover Types: Different land cover types, such as forests, urban areas, water bodies, and agricultural lands, exhibit considerable variability in their appearance. This variability can make it challenging to distinguish between them accurately using conventional image processing techniques.
2. Complexity of Satellite Images: Satellite images often contain complex patterns and textures that are difficult to interpret. Factors such as varying lighting conditions, shadows, and different scales of land cover features add to the complexity of the classification task.
3. Limited Resolution and Quality: Despite advancements in satellite technology, the resolution and quality of satellite images may still vary. This can impact the effectiveness of feature extraction and classification algorithms, particularly when dealing with fine-grained land cover details.
4. Integration of Multiple Features: Effective land cover classification typically requires the integration of multiple features such as spectral information, texture, and spatial relationships. Developing a comprehensive approach to combine these features in a meaningful way is a challenge.

## 3 Methodology

The methodology outlines the systematic approach employed in processing and classifying satellite imagery for land cover classification. This process includes data acquisition, preprocessing, feature extraction, classification, and evaluation. Each step is crucial for ensuring accurate and reliable results.

### 3.1 Data Acquisition

- **Satellite Image Retrieval:** Satellite images are obtained from the Sentinel-2 mission, which provides high-resolution optical imagery. Images are selected based on specific geographic coordinates and date ranges to cover the target area and are subsequently downloaded for further analysis.
- **Labeled Data:** Labeled images representing various land cover types (e.g., farms, buildings, water bodies, empty land, and forests) are used as ground truth. These labeled images provide the reference needed to train and validate the classification model.

### 3.2 Data Preprocessing

- **Image Import and Display:** Retrieved satellite and labeled images are imported into the analysis environment. Initial visualization ensures correct loading and spatial alignment.
- **Normalization:** Images are normalized to standardize pixel values across different images and sensor characteristics, mitigating variations due to illumination and sensor noise.

- **Alignment:** Spatial correspondence between satellite and labeled images is ensured through alignment or resampling, adjusting the resolution and extent to match each pixel in the satellite image with the corresponding pixel in the labeled images.

### 3.3 Feature Extraction

- **Spectral Features:** Key spectral bands (Red, Green, Blue, and Near-Infrared) are extracted from the satellite images.
  - *Normalized Difference Vegetation Index (NDVI):* Calculated as $(NIR - Red)/(NIR + Red)$, NDVI quantifies vegetation health and density.
  - *Enhanced Vegetation Index (EVI):* Calculated as $2.5 \times (NIR - Red)/(NIR + 6 \times Red - 7.5 \times Blue + 1)$, EVI enhances vegetation signals and reduces atmospheric interference.
  - *Soil-Adjusted Vegetation Index (SAVI):* Calculated as $(1+L) \times (NIR - Red)/(NIR + Red + L)$, where $L$ is a soil adjustment factor, SAVI reduces soil brightness effects.
- **Texture Features:** Texture information is extracted from grayscale images.
  - *Entropy:* Measures the complexity or disorder in an image.
  - *Local Binary Patterns (LBP):* Captures texture by comparing pixel intensity with neighbors.

### 3.4 Classification

- **Feature Combination:** Extracted features (NDVI, EVI, SAVI, entropy, LBP) are combined into a feature matrix representing data input for the classification algorithm.
- **Training and Testing:** Random Forest Classifier is trained using the labeled data. The classifier learns to categorize pixels based on the feature matrix.
  - *Cross-Validation:* Employed to assess model performance and avoid overfitting by splitting data into training and validation sets.

### 3.5 Evaluation

- **Performance Metrics:** Classification results are evaluated using metrics such as accuracy, precision, recall, and F1 score.
  - *Accuracy:* Proportion of correctly classified pixels out of the total.
  - *Confusion Matrix:* Shows true positive, true negative, false positive, and false negative predictions for each class.
  - *Precision, Recall, F1 Score:* Metrics providing detailed evaluation for each class. Precision measures correct positive predictions, recall measures true positives out of actual positives, and F1 score is the harmonic mean of precision and recall.
- **Visualization:** The classified image is visualized to provide a spatial representation of land cover types and to compare classification results with labeled data.

## 4 Proposed Solution

The proposed solution integrates multiple techniques to classify land cover types from satellite images. The solution is designed to effectively distinguish between different land cover types such as farmland, buildings, water bodies, empty land, and forests using high-resolution Sentinel-2 images.

### 4.1 System Components

- **Image Retrieval:** Utilizes Google Earth Engine to obtain high-resolution satellite images of the target area, ensuring high-quality data for analysis.
- **Feature Extraction:** Employs advanced techniques to extract spectral and texture features from the images, including NDVI, EVI, SAVI, entropy, and Local Binary Patterns (LBP).

- **Classification:** Implements a Random Forest Classifier to categorize the images into predefined land cover types based on the extracted features.
- **Evaluation:** Uses accuracy, confusion matrix, precision, recall, and F1 score to evaluate the performance of the classification model and ensure reliable results.

### 4.2 Integration of Components

The system integrates image retrieval, feature extraction, and classification in a streamlined workflow:

- The high-resolution satellite images are first retrieved and preprocessed to ensure they are suitable for analysis.
- Relevant features are extracted from the images, capturing both spectral and texture information.
- The extracted features are fed into an Random Forest Classifier, which is trained and tested on labeled data to classify the images into different land cover types.
- The performance of the classifier is rigorously evaluated using various metrics to ensure accuracy and reliability.

### 4.3 Advantages of the Solution

The proposed solution offers several advantages:

- **High Accuracy:** The use of advanced feature extraction techniques and a robust classification model results in high accuracy in land cover classification.
- **Comprehensive Analysis:** The integration of spectral and texture features provides a comprehensive analysis of the land cover types.
- **Scalability:** The solution can be scaled to analyze larger areas or different locations by adjusting the parameters of the image retrieval process.

## 5 System Architecture

The system architecture section outlines the overall design of the land cover classification system. It includes an overview of the system components, their interactions, and how they work together to achieve accurate land cover classification from satellite images.

### 5.1 Overview

The system architecture is designed to process high-resolution satellite images, extract relevant features, classify land cover types, and evaluate the performance of the classification model. The architecture is modular, allowing for flexibility and scalability.

### 5.2 Components

- **Data Acquisition Module:**
  - Responsible for retrieving satellite images from sources such as Google Earth Engine or Sentinel Hub.
  - Handles the downloading of images based on geographic coordinates and date ranges.
- **Preprocessing Module:**
  - Performs normalization, alignment, and initial visualization of images to prepare them for feature extraction.
  - Ensures that the satellite and labeled images are correctly aligned and standardized.
- **Feature Extraction Module:**
  - Extracts both spectral and texture features from the images.

  - Computes indices such as NDVI, EVI, and SAVI, and texture features like entropy and Local Binary Patterns (LBP).
- **Classification Module:**
  - Utilizes a Support Vector Machine Random Forest Classifier to categorize the extracted features into land cover types.
  - Handles training and testing of the classifier using labeled data.
- **Evaluation Module:**
  - Assesses the performance of the classification model using metrics such as accuracy, confusion matrix, precision, recall, and F1 score.
  - Provides visualizations to compare classification results with ground truth data.

### 5.3 Workflow

The workflow of the system architecture is as follows:

- **Image Retrieval:** Satellite images are obtained and downloaded from the data acquisition module.
- **Preprocessing:** Images are normalized, aligned, and visually inspected to ensure they are ready for feature extraction.
- **Feature Extraction:** Spectral and texture features are extracted from the preprocessed images.
- **Classification:** The feature matrix is input to the Random Forest Classifier, which is trained and tested to classify the land cover types.
- **Evaluation:** The classification results are evaluated using various performance metrics, and the results are visualized.

### 5.4 Advantages

The modular design of the system architecture offers several advantages:

- **Flexibility:** Each component can be updated or replaced independently, allowing for easy adaptation to new requirements or technologies.
- **Scalability:** The system can be scaled to handle larger datasets or additional land cover classes by modifying the relevant modules.
- **Efficiency:** The separation of concerns ensures that each module focuses on a specific task, improving the overall efficiency of the system.

## 6 Implementation Details

The implementation details describe the hardware and software environment used for the development and execution of the digital image processing pipeline for land cover classification.

### 6.1 Hardware Specifications

- **Device Name:** DESKTOP-E1503BC
- **Processor:** Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz
- **Installed RAM:** 8.00 GB (7.87 GB usable)
- **System Type:** 64-bit operating system, x64-based processor
- **Pen and Touch:** No pen or touch input is available for this display

### 6.2 Software Specifications

- **Operating System:** Windows 10 (64-bit)
- **Python Version:** Python 3.x (specify exact version used)
- **Libraries:**
  - *geemap*: for accessing and visualizing satellite imagery
  - *NumPy*: for numerical operations
  - *OpenCV*: for image processing tasks
  - *scikit-learn*: for machine learning and classification
- **IDE/Editor:** Jupyter Notebook for Python

### 6.3 Implementation Workflow

- **Python for Image Retrieval:**
  - **Purpose:** Python is used for retrieving satellite images from Google Earth Engine. Then used for processing and classifying the retrieved satellite images.
  - **Process:** Satellite images are accessed and downloaded based on specific geographic coordinates and date ranges. Python scripts handle this retrieval process and save the images to the local drive for subsequent analysis. Once the images are retrieved, preprocessing, feature extraction, and classification is performed. This includes tasks such as normalization, alignment, feature extraction (e.g., NDVI, texture features), and classification using Random Forest Classifier (RFC).
  - **Libraries Used:** *geemap* for Earth Engine access, *NumPy* for data handling, and *OpenCV* for image manipulation.

### 6.4 Implementation Environment

- **Development Environment:** The development and testing were conducted using the specified hardware and software setup.
- **File Storage:** Satellite images and processed data are stored locally.

### 6.5 Challenges and Solutions

- **Memory Constraints:** Limited RAM required efficient handling of image data. Python scripts manage image retrieval and preliminary processing to reduce memory load before data is passed.
- **Processing Time:** The computational power of the i5 processor demanded optimization of the code to ensure timely execution of image processing and classification tasks.

## 7 Results

### 7.1 Image Retrieval and Preprocessing

To begin the land cover classification task, we first retrieved satellite images from the Sentinel-2 collection using Google Earth Engine. The process involved the following steps:

- **Location and Radius Definition:** The target area was centered at coordinates (latitude: 11.57628, longitude: 37.42476) with a specified radius of 3,048 meters (approximately 10,000 feet). This area of interest was defined as a bounding box geometry.
- **Image Collection Filtering:** We queried the Sentinel-2 image collection for the specified location, filtering the images by date range (May 1, 2024, to May 10, 2024) and limiting the results to images with less than 10% cloud cover. This filtering ensures that the selected images are of high quality and suitable for land cover analysis.
- **Image Selection and Visualization:** The size of the filtered collection was checked to ensure that suitable images were available. If images were found, the first image in the

collection was selected for further analysis. Visualization parameters were defined for rendering the image using the Red, Green, and Blue (RGB) bands, with a gamma correction of 1.4 to enhance image contrast. The image and the bounding box were then added to an interactive map for visualization.

The successful retrieval and visualization of a Sentinel-2 image for the specified location are critical to the land cover classification task, providing the raw data required for subsequent analysis.

## 7.2 Classification Performance

With the satellite image retrieved and preprocessed, the Random Forest classifier was applied to classify different land cover types within the image. The classification performance was evaluated using several metrics, including precision, recall, F1-score, and overall accuracy.

### 7.2.1 Precision, Recall, and F1-Score

The classification report provides a detailed overview of the precision, recall, and F1-score for each land cover class, which include Farm, Building, Water, Empty Land, and Forest.

- **Farm:** The precision for the Farm class is 0.85, meaning that 85% of the pixels predicted as Farm are correctly classified. The recall is 0.89, indicating that 89% of all actual Farm pixels were correctly identified by the model. The F1-score, which balances precision and recall, is 0.87.
- **Building:** The Building class achieved a precision of 0.95 and a recall of 0.93, resulting in a high F1-score of 0.94. This indicates that the model performs well in identifying and correctly classifying Building pixels.
- **Water:** The Water class had a perfect precision of 1.00, meaning that all pixels classified as Water were correct. The recall was also 1.00, demonstrating that the model successfully identified all Water pixels. Consequently, the F1-score is also 1.00, reflecting flawless classification for this class.
- **Empty Land:** For the Empty Land class, the precision was 0.84, while the recall was slightly lower at 0.81, leading to an F1-score of 0.83. This indicates a good but not perfect performance, with some Empty Land pixels being misclassified.
- **Forest:** The Forest class achieved a high precision of 0.97 and a recall of 0.99, resulting in an F1-score of 0.98. This suggests that the model is highly effective in identifying Forest pixels.

### 7.2.2 Overall Accuracy

The overall accuracy of the model, which represents the proportion of correctly classified pixels across all classes, was found to be 0.91. This means that 91% of the pixels in the test set were correctly classified by the Random Forest model, demonstrating strong performance in this land cover classification task.

### 7.2.3 Macro and Weighted Averages

The macro average F1-score, which averages the F1-scores across all classes, was 0.92. This metric treats all classes equally, regardless of their size. The weighted average F1-score, which takes into account the support (number of true instances) of each class, was 0.91. This indicates that the model performs consistently well across both large and small classes.

## 7.3 Confusion Matrix Analysis

The confusion matrix provides a more detailed view of the model's performance by showing the actual versus predicted classifications for each class.

- **Farm:** Out of 10,495 actual Farm pixels, 9,306 were correctly classified, while 1,069 were misclassified as Empty Land, and a small number were misclassified into other classes.

- **Building:** Of the 6,497 actual Building pixels, 6,047 were correctly classified. A total of 232 pixels were misclassified as Farm, and 146 as Empty Land.
- **Water:** The Water class had 9,043 pixels, with 9,016 correctly classified and only 27 misclassified, reflecting the model's near-perfect accuracy for this class.
- **Empty Land:** For the 8,071 actual Empty Land pixels, 6,525 were correctly classified, while a notable number (1,348) were misclassified as Farm.
- **Forest:** The Forest class, consisting of 3,775 pixels, saw 3,730 correctly classified, with only minor misclassifications.

The confusion matrix highlights the challenges in distinguishing between Farm and Empty Land, as well as between Farm and Building. However, the model performs exceptionally well in distinguishing Water and Forest from other classes.

### 7.4 Summary

In summary, the Random Forest classifier demonstrated strong performance in classifying different land cover types in satellite images. The model achieved a high overall accuracy of 91%, with particularly high precision and recall for the Water and Forest classes. While the classification of Farm and Empty Land posed some challenges, the results are generally robust and suggest that the model is well-suited for land cover classification tasks.

The process of image retrieval and preprocessing using Google Earth Engine was successful, providing high-quality satellite images with minimal cloud cover, which were essential for the subsequent classification task. Future improvements could involve addressing the confusion between Farm and Empty Land classes, possibly through enhanced feature engineering or incorporating additional data such as spectral indices or higher-resolution imagery.

## 8 Deployment

The deployment section outlines how the developed land cover classification system can be implemented in real-world settings. It includes considerations for operational setup, maintenance, and potential for scaling.

### 8.1 System Deployment

- **Environment Setup:** The system is deployed in an environment where Python is installed.
- **Dependencies:** Ensure all necessary libraries and tools are installed. For Python, this includes ‘geemap‘, ‘earthengine-api‘, and other dependencies for image retrieval and preprocessing.
- **Configuration:** Configure the system with parameters such as geographic coordinates, date ranges, and classification classes. These parameters should be set based on the specific area and types of land cover being analyzed.
- **Data Integration:** Set up the system to handle the integration of retrieved satellite images with labeled data. Ensure that the preprocessing steps align images correctly and that features are extracted and classified accurately.

### 8.2 Operational Considerations

- **Data Handling:** Implement mechanisms for managing large volumes of satellite imagery data, including storage solutions and data backup.
- **Processing Time:** Be aware of the processing time required for image retrieval, preprocessing, and classification. Optimize code and utilize efficient algorithms to handle large datasets effectively.
- **Error Handling:** Implement error handling and logging to capture and address any issues that arise during the processing and classification phases.

- **User Interface:** Develop a user-friendly interface, if applicable, for users to interact with the system, configure parameters, and view results. This could be a web-based interface or a standalone application.

### 8.3 Maintenance

- **Software Updates:** Regularly update the software and libraries to incorporate improvements and fix any bugs. Ensure compatibility with the latest versions.
- **Data Updates:** Periodically update the dataset with new satellite images to maintain the relevance and accuracy of the classification results.
- **Performance Monitoring:** Monitor system performance and accuracy over time. Adjust and fine-tune the classification model as needed based on new data and evolving requirements.

### 8.4 Scaling and Future Enhancements

- **Scalability:** Design the system to scale efficiently to handle larger geographic areas or additional land cover classes. Consider cloud-based solutions or distributed processing for handling extensive datasets.
- **Future Enhancements:** Explore opportunities for enhancing the system, such as integrating advanced machine learning techniques, incorporating additional satellite data sources, or improving feature extraction methods.
- **Integration with Other Systems:** Investigate potential integrations with other geographic information systems (GIS) or environmental monitoring tools to provide a more comprehensive analysis.

## 9 Conclusion

In this project, we developed a land cover classification system utilizing high-resolution Sentinel-2 satellite imagery combined with advanced digital image processing techniques. The system achieved an impressive accuracy of 91

The project employed a systematic methodology, starting with the retrieval of satellite images and progressing through data preprocessing, feature extraction, and classification. Key techniques such as NDVI, EVI, and SAVI were used to extract valuable spectral information, while texture features like entropy and Local Binary Patterns (LBP) provided additional discriminative power. Despite the overall success, some challenges were encountered, particularly with the quality of labeled data and classification results. These issues underscore the importance of high-quality input data and accurate preprocessing to achieve reliable outcomes.

The implications of this work are significant for various applications, including environmental monitoring, urban planning, and resource management. The system's ability to accurately classify land cover types offers valuable insights into land use patterns and supports decision-making processes in these domains. Additionally, the integration of satellite imagery with advanced classification techniques provides a practical tool for analyzing large geographic areas and obtaining detailed land cover information.

Looking ahead, there are several avenues for future work. Enhancing the quality of labeled data and refining preprocessing techniques will be crucial for addressing current limitations and improving classification accuracy. Exploring more advanced machine learning and deep learning methods could further enhance the system's performance and handle complex land cover types. Expanding the system to incorporate additional features, such as temporal analysis of land cover changes or integration with other remote sensing data sources, would provide a more comprehensive analysis. Moreover, developing scalable solutions and automating parts of the process could facilitate real-time land cover monitoring and analysis, making the system more adaptable to various applications and larger datasets.