GIFT UNIVERSITY
Converting Knowledge into Practical Experience

PROPOSAL

(Satellite Image Classification Using Machine Learning Algorithms)

NAME

MUHAMMAD FARAZ – 221980007

IJAZ UL HASSAN – 221980030

HASSAN RASHEED – 221980038

MUHAMMAD HASEEB – 221980004

COURSE

APPLICATIONS OF DATA SCIENCE (A)

INSTRUCTOR

UMER RAMZAN

Proposal for Satellite Image Classification Using Machine Learning Algorithms

Title: Satellite Image Classification Using Google Earth Engine and Machine Learning Algorithms: A Case Study of PAKISTAN

Introduction: Accurate land cover classification is essential for effective environmental management, urban planning, and sustainable development. Traditional methods of land classification are time-intensive and lack scalability. This proposal outlines a project aimed at leveraging machine learning algorithms and satellite imagery to classify land cover in PAKISTAN using the Google Earth Engine (GEE) platform.

Dataset Name:

Landsat 8 OLI/TIRS (Operational Land Imager and Thermal Infrared Sensor)

Objectives:-

Primary Objective:

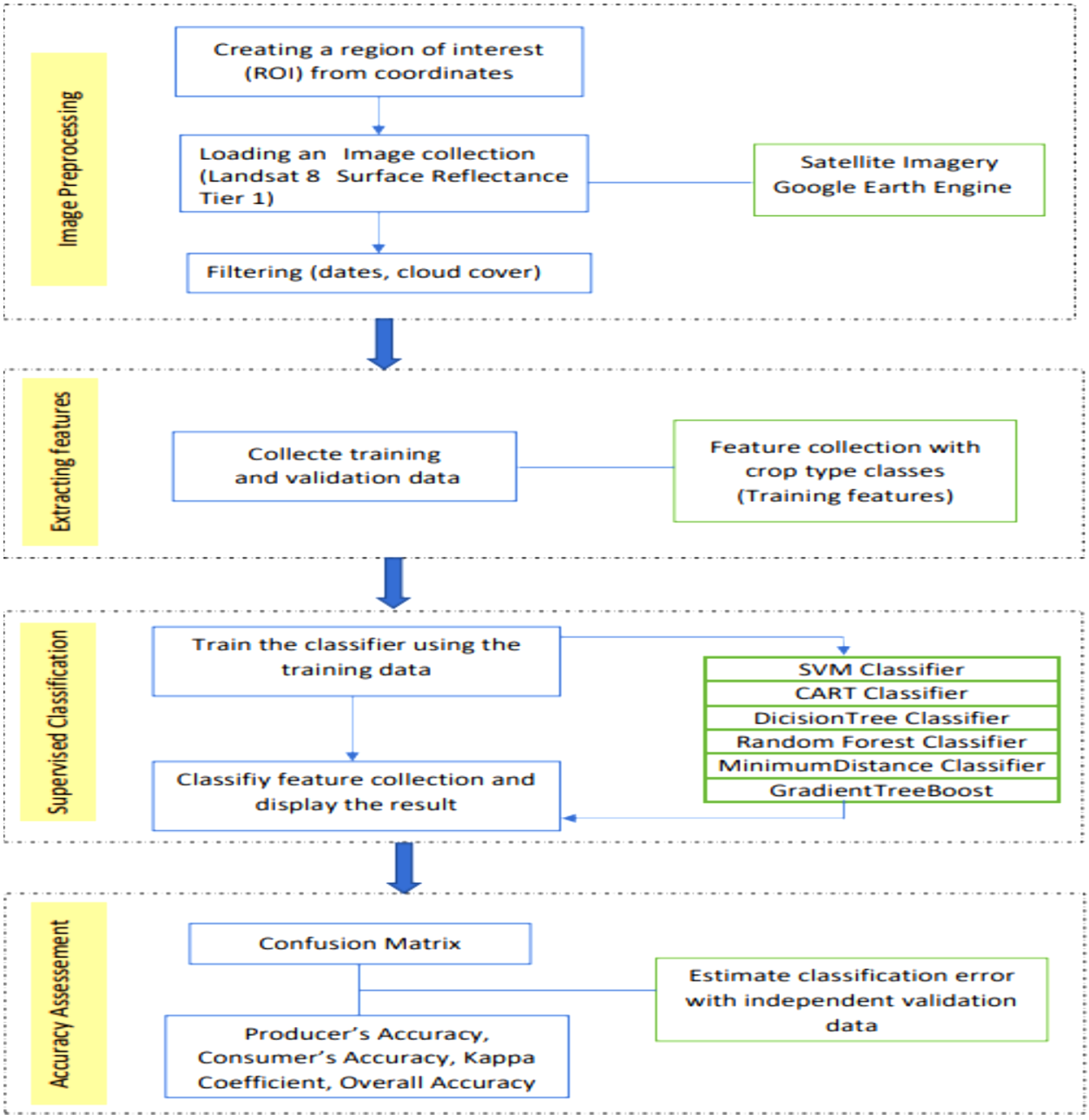
- To classify land cover types in PAKISTAN, including water, forest, barren land, built-up areas, cultivated areas, and sandy regions, using satellite imagery.

Secondary Objectives:

1. To compare the performance of six supervised machine learning algorithms:
    - Random Forest (RF)
    - Support Vector Machine (SVM)
    - Classification and Regression Trees (CART)
    - Gradient Tree Boost (GTB)
    - Decision Tree (DT)
    - Minimum Distance (MD)
2. To improve classification accuracy using spectral indices such as NDVI, NDBI, BSI, and MNDWI.

3. To utilize the computational power of the GEE platform for large-scale data processing and analysis.

Study Area: The project focuses on PAKISTAN, a country with diverse land cover types, including mountains, deserts, forests, and agricultural zones. PAKISTAN's geographical diversity makes it an ideal case study for testing the robustness of machine learning algorithms in land classification.



1. Data Collection:

- Utilize LANDSAT 8 OLI satellite imagery for the year 2021.
- Extract spectral bands and indices from the GEE platform.

2. Preprocessing:

- Apply atmospheric correction to remove noise.
- Filter images by date and cloud cover.

3. Feature Extraction:

- Extract spectral and spatial features from the imagery.
- Add spectral indices such as NDVI, NDBI, BSI, and MNDWI to enhance feature representation.

Index	Equation using Landsat 8 OLI	References
Normalized Difference Vegetation Index (NDVI)	$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}}$	[47], [48]
Modified Normalized Difference Water Index (MNDWI)	$\mathrm{MNDWI} = \frac{\mathrm{Green} - \mathrm{SWIR1}}{\mathrm{Green} + \mathrm{SWIR1}}$	[49]
Bare Soil Index (BSI)	$\mathrm{BSI} = \frac{\mathrm{Green} + \mathrm{NIR}}{\mathrm{Green} - \mathrm{NIR}}$	[50]
Normalized Difference Built-up Index (NDBI)	$\mathrm{NDBI} = \frac{\mathrm{SWIR} - \mathrm{NIR}}{\mathrm{SWIR} + \mathrm{NIR}}$	[51]

4. Model Training and Classification:

- Divide data into 80% training and 20% validation sets.
- Train six supervised machine learning algorithms using GEE's built-in classifiers.

- Evaluate performance using metrics like Overall Accuracy (OA), Kappa Coefficient, User Accuracy (UA), and Producer Accuracy (PA).

5. Validation:

- Use a confusion matrix to validate the accuracy of classified maps.
- Compare the performance of algorithms before and after adding spectral indices.

Expected Outcomes:-

1. **Classification Maps:**
    - High-resolution land cover maps of PAKISTAN showcasing classified regions such as water bodies, forests, and urban areas.
2. **Performance Metrics:**
    - Identification of the most effective algorithm for land cover classification.
    - Improved accuracy metrics with the addition of spectral indices.
3. **Insights:**
    - Recommendations for optimizing machine learning models for large-scale land classification.
    - Demonstration of GEE's efficiency in handling and processing satellite imagery.

Innovations and Contributions:-

- Integration of diverse machine learning algorithms for satellite image classification.
- Use of spectral indices to enhance classification accuracy.
- Application of GEE's cloud-based capabilities for large-scale geospatial analysis.
- Focus on an under-researched geographical area (PAKISTAN), filling a significant research gap.

Proposed Timeline:-

Phase	Duration	Tasks
Data Collection	3 Days	Collect and preprocess LANDSAT 8 imagery.
Feature Extraction	3 Days	Add spectral indices and prepare training data.
Model Training	3 Days	Train and test six machine learning algorithms.
Validation and Analysis	2 Days	Evaluate models and analyze performance.
Report Writing	1 Days	Document findings and prepare final report.

Resources Required

1. **Technical Resources:**
   - Access to Google Earth Engine (GEE) platform.
   - Python libraries for data analysis and machine learning (e.g., Scikit-learn, TensorFlow).
2. **Human Resources:**
   - Data scientist with experience in machine learning.
3. **Data Resources:**
   - LANDSAT 8 OLI satellite imagery.

Conclusion: This project aims to advance the field of remote sensing by combining machine learning algorithms with cloud-based geospatial platforms. The findings will contribute to more efficient and accurate land classification methods, offering valuable insights for environmental management and planning in PAKISTAN and beyond.

Reference:

https://ieeexplore.ieee.org/document/10177754