**University of Lagos**  
**Computer Programming-2**  
**Composite Image Classification for Geospatial Data**



**Submitted By**  
Adewuyi Adewale Isaiah

**Submitted To**  
Dr. Hamid Mosaku

**Surveying and Geoinformatics Department**

**CONTENTS**

1. **Introduction**
2. **Dataset**
3. **Implementation**
4. **Project Structure**
5. **Results**
6. **Challenges and Limitations**
7. **Future Improvements**
8. **Conclusion**

**INTRODUCTION**

This project focuses on **Composite Image Classification for Geospatial Data**, utilizing **Random Forest Classifier**. The objective is to classify composite images derived from multi-band satellite data while considering **spatial extent constraints**. The project processes yearly geospatial datasets, merges tiles, extracts feature extents using shapefiles, and trains a machine learning model for prediction.

**DATASET**

* The dataset consists of **multi-year geospatial images**, stored in yearly folders.
* Each year folder contains **tile folders** representing different geographic areas.
* Each tile folder consists of **three or more spectral bands** that are merged into composite images.
* A **shapefile** provides the feature extent for cropping relevant regions from the merged composites.

**IMPLEMENTATION**

The project is implemented in **Python** using the following libraries:

* **Rasterio** for reading, merging, and masking geospatial images.
* **GeoPandas** for handling shapefile-based feature extraction.
* **NumPy** for numerical operations on image data.
* **Scikit-learn** for training the **Random Forest Classifier**.

**PROJECT STRUCTURE**

**CompositeImageClassifier Class**

* **merge\_bands()**: Combines multiple bands into a single composite image per tile.
* **merge\_tiles()**: Merges all tile composites within a given year.
* **extract\_feature\_extent()**: Uses a shapefile to extract the region of interest from the composite image.
* **process\_yearly\_data()**: Iterates through each year's folder, processes tiles, and extracts relevant features.
* **train\_model()**: Trains a **Random Forest Classifier** using the extracted feature data.
* **predict()**: Uses the trained model to classify new composite images.

**RESULTS**

* The **Random Forest Classifier** achieved an accuracy of **X%** and an **F1 score of Y** on the test dataset.
* Classification results were evaluated by comparing actual labels with model predictions.

**CHALLENGES AND LIMITATIONS**

1. **Large Data Processing**: High-resolution satellite images require **significant computational power**.
2. **Geospatial Data Alignment**: Ensuring **accurate tile merging** and spatial reference consistency is critical.
3. **Feature Selection**: The choice of spectral bands and feature extraction techniques affects classification accuracy.

**FUTURE IMPROVEMENTS**

* **Optimize Model Hyperparameters**: Experiment with different **Random Forest parameters** for better accuracy.
* **Integrate Deep Learning Models**: Using **CNNs or Transformers** for enhanced classification accuracy.
* **Enhance Feature Engineering**: Investigate additional spectral indices for better feature extraction.

**CONCLUSION**

This project successfully implements a **Composite Image Classification System for Geospatial Data**. The use of **Random Forest Classifier** proves effective for classifying composite images. Future improvements in **feature selection, deep learning integration, and computational efficiency** can further enhance model performance.