Capstone Project Concept Note and Implementation Plan

Project Title: GreenLensGroup – Near Real-Time Forest Monitoring in Liberia Using Machine Learning

# Team Members

- Fatu Kromah  
- Alfred Nyeswa  
- Sulaiman Barry  
- Jerome N. Tokpa  
- Joseph S. Lablah Jr

# Concept Note

## 1. Project Overview

Liberia’s rainforests are critical for biodiversity, climate regulation, and community livelihoods, yet they face alarming deforestation from logging, agriculture, and mining. Between 2002 and 2022, over 300,000 hectares of primary forest were lost (Global Forest Watch). Traditional monitoring approaches are slow and insufficient for real-time intervention.  
  
GreenLensGroup addresses this challenge by developing a machine learning–based, near real-time forest monitoring system that integrates optical (Sentinel-2, Landsat) and contextual (GFW, OSM) data. The project directly supports SDG 13 (Climate Action) and SDG 15 (Life on Land) by enabling faster detection and response to deforestation.

## 2. Objectives

- Develop a Liberia-specific machine learning model for classifying forest, non-forest, and forest loss.  
- Integrate multi-source satellite data for near real-time detection.  
- Deploy interpretable tools to support policy and community conservation.  
- Build capacity for sustainable AI-driven environmental monitoring in Liberia.

## 3. Background

Existing global models (e.g., Hansen et al., 2013) provide valuable insights but lack regional customization. Cloud cover, seasonal variability, and limited field validation hinder tropical forest monitoring. Machine learning offers a scalable, automated approach, while deep learning (e.g., EfficientNet-Lite) enhances accuracy in spatial pattern recognition. By tailoring models to Liberia’s unique ecosystems and integrating contextual drivers (roads, mining zones), GreenLensGroup fills a critical gap.

## 4. Methodology

- Data Collection: Sentinel-2, Landsat, GFW labels, OSM, GADM.  
- Preprocessing: Cloud masking, compositing, tiling (512×512).  
- Modeling: Random Forest as baseline; EfficientNet-Lite CNN for advanced classification.  
- Evaluation: Precision, recall, F1-score (target ≥85%), confusion matrix, spatial cross-validation.  
- Deployment: Prototype web dashboard for alerts and visualization.

## 5. Architecture Design Diagram

A block diagram will illustrate data input, preprocessing, model training, classification, and dashboard outputs. Each component will highlight its role and interactions in the system.

## 6. Data Sources

Sentinel-2 and Landsat provide multi-spectral optical data; GFW supplies forest loss labels; OSM and GADM contextualize human activity and boundaries. Together, these datasets form a multi-source pipeline for accurate and timely monitoring.

## 7. Literature Review

Research shows that satellite-driven ML can effectively detect deforestation: Hansen et al. (2013) pioneered Landsat-based global forest mapping, Reiche et al. (2018) proved radar’s value in cloud-heavy tropics, Belgiu & Csillik (2018) validated Random Forest with Sentinel-2, and Zhu et al. (2020) confirmed CNN superiority. GreenLensGroup builds on these works by developing a Liberia-specific, near real-time solution.

# Implementation Plan

## 1. Technology Stack

- Languages: Python, JavaScript (for GEE).  
- Libraries: Pandas, NumPy, Scikit-learn, TensorFlow/Keras.  
- Frameworks: EfficientNet-Lite, Random Forest.  
- Tools: Google Earth Engine, GDAL/Rasterio, QGIS.  
- Deployment: Web dashboard (Flask/Django).

## 2. Timeline

- Weeks 1–2: Data collection & preprocessing.  
- Weeks 3–5: Baseline model (Random Forest).  
- Weeks 6–8: Deep learning model (EfficientNet-Lite).  
- Weeks 9–10: Model evaluation & optimization.  
- Weeks 11–12: Dashboard prototype & deployment.

## 3. Milestones

- Dataset integration completed.  
- Baseline model achieves >80% F1.  
- EfficientNet-Lite surpasses 85% F1.  
- Dashboard operational with Liberia-specific data.

## 4. Challenges and Mitigations

- Cloud cover → Integrate radar (Sentinel-1).  
- Data imbalance → Apply stratified sampling and F1-based evaluation.  
- Limited compute → Optimize with EfficientNet-Lite; use GEE for preprocessing.  
- Ground truth scarcity → Partner with NGOs for field validation.

## 5. Ethical Considerations

Ethical considerations include data privacy during community validation, addressing bias in training samples, ensuring responsible use of deforestation alerts, and promoting open access for conservation stakeholders in Liberia.

## 6. References

- Hansen, M. C., et al. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160), 850–853.  
- Reiche, J., et al. (2018). A Framework for Near Real-Time Forest Monitoring with Sentinel-1. Remote Sensing of Environment, 215, 12–25.  
- Belgiu, M., & Csillik, O. (2018). Sentinel-2 for Land Cover Classification Using Random Forest. Remote Sensing, 10(2), 247.  
- Zhu, X., et al. (2020). Deep Learning in Remote Sensing: A Review. ISPRS Journal of Photogrammetry and Remote Sensing, 162, 156–177.  
- Global Forest Watch. (2023). Tree Cover Loss Data. https://globalforestwatch.org