**Research Preparation Document**

**Project:** 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

**Date:** April 5, 2025

# Literature Review

**Introduction**

Liberia is home to one of West Africa’s last remaining intact rainforests, covering nearly 40% of its landmass. However, deforestation has accelerated in recent decades due to logging, agricultural expansion, mining, and infrastructure development. Between 2002 and 2022, Liberia lost over 300,000 hectares of primary forest (Global Forest Watch). This forest loss threatens biodiversity, undermines carbon sequestration, and destabilizes local livelihoods.

Traditional monitoring methods—such as infrequent field surveys and delayed satellite analysis—are insufficient for early detection and timely response. Advancements in machine learning (ML) and remote sensing now make it possible to design near real-time monitoring systems that are scalable, accurate, and actionable.

The following literature review explores foundational work in remote sensing and AI for deforestation monitoring, highlighting best practices and knowledge gaps that inform GreenLensGroup’s approach.

**Summary of Key Papers:**

* Hansen et al. (2013): Global maps of forest change, Landsat-based.
* Reiche et al. (2018): Sentinel-1 radar for near real-time alerts.
* Belgiu & Csillik (2018): Sentinel-2 and Random Forest, >90% accuracy.
* Zhu et al. (2020): Deep learning review, CNNs outperform classical ML.

**Conclusion**

Machine learning and satellite data are powerful for deforestation monitoring, but Liberia needs region-specific, real-time systems.

# Data Research

**Introduction**

Accurate monitoring depends on high-quality, relevant data.

**Data Sources:**

* Sentinel-2 (optical, 10–20m)
* Landsat 8/9 (30m, historical)
* Global Forest Watch (forest loss labels)
* GADM (boundaries)
* OpenStreetMap (roads, settlements)

**Findings:**

* Class imbalance: forest loss <5% of samples.
* Cloud cover >60% in wet season; dry season more usable.
* 78% of forest loss near roads; mining hotspots identified.

**Conclusion**

Selected datasets provide strong basis, though cloud cover and imbalance remain challenges.

# Technology Review

**Introduction**

The project relies on scalable, open-source tools for data and modeling.

**Tools:**

* Google Earth Engine: planetary-scale data access.
* Python (Pandas, NumPy, Scikit-learn, TensorFlow).
* EfficientNet-Lite: deep CNN for classification.
* Random Forest: robust baseline.
* GDAL/Rasterio: geospatial preprocessing.

**Gaps:**

* Internet dependence of GEE.
* Annual labels from GFW.
* Deployment challenges on edge devices.

**Future Opportunities:**

* Sentinel-1 integration.
* Active learning for labeling.
* Vision Transformers and U-Net segmentation.

# Project Integration & Impact

**Objective:** Detect and classify deforestation in Liberia using ML.

**Models:** Random Forest baseline, EfficientNet-Lite advanced.

**Evaluation:** Target F1 ≥ 85%.

**Innovation:** Liberia-specific, near real-time monitoring.

**Impact:**

* Faster response to deforestation.
* Stronger policy support.
* Empowering conservation in developing regions.

# Conclusion

GreenLensGroup leverages satellite data and AI to create a Liberia-specific, real-time forest monitoring framework. The project contributes to environmental AI by demonstrating the power of accessible tools and open data in tackling global sustainability challenges.