

Deforestation Monitoring in Nuwara Eliya District, Sri Lanka (2013–2025) — with Forecasting

Project Report

1. Introduction and Background

1.1 Study problem: deforestation in Nuwara Eliya District

Deforestation refers to the long-term reduction of forest cover due to clearing, conversion to agriculture, expansion of settlements, infrastructure development, or other disturbances that remove or fragment natural vegetation. Sri Lanka's forests are ecologically significant because they support biodiversity, regulate local climate, store carbon, and protect watersheds.

This project focuses on **Nuwara Eliya District** (administrative district) in Sri Lanka's Central Highlands. The district includes montane ecosystems and important headwater catchments. Because the region has steep slopes and high rainfall variability, forest loss can increase risks such as **soil erosion, landslides, and sedimentation**, while also reducing ecosystem services.

1.2 Why monitoring deforestation trends matters

Studying deforestation trends is important for:

- **Environmental management and conservation planning:** identifying where loss occurs and how it changes over time.
- **Watershed protection:** forests reduce surface runoff and erosion; loss increases sediment loads in rivers.
- **Climate and carbon management:** forests store carbon; forest loss contributes to emissions.
- **Evidence-based decision-making:** district-level statistics provide measurable indicators for local planning and policy.

1.3 Role of satellite imagery in land cover monitoring

Satellite remote sensing is one of the most effective ways to monitor land cover change because it provides:

- **consistent repeat observations** over long time periods,
- **broad spatial coverage** across administrative boundaries,
- **spectral measurements** that reflect vegetation condition,
- the ability to generate **time series** suitable for change detection.

For deforestation monitoring, vegetation indices such as **NDVI** and moisture-related indices such as **NDMI** detect changes in canopy greenness and moisture. In tropical regions, cloud contamination is a major challenge; therefore, QA layers such as **QA_PIXEL** are critical to remove unreliable pixels.

2. Methodology

This project implements an **unlabeled (no ground-truth) deforestation workflow** using Landsat Collection 2 Level-2 (L2) Surface Reflectance imagery. The workflow is designed to be robust by using QA-based cloud masking, yearly median composites, and persistence rules.

2.1 Data Set Description

2.1.1 Satellite images used

- **Source:** USGS EarthExplorer
- **Sensor/product:** Landsat 8 Collection 2 Level-2 Surface Reflectance (L2 SR)
- **Example scene naming pattern:** LC08_L2SP_..._02_T1_...

2.1.2 Spatial resolution

- Landsat multispectral bands are effectively **30 m** resolution.

2.1.3 Time period

- Input imagery archive used in the project: **2013–2025**
- **Baseline period:** 2013–2015
- **Analysis period:** 2016–2025

2.1.4 Bands and QA layer

This project uses only a minimal band set required for vegetation and moisture indices plus QA masking:

- **Red band:** SR_B4
- **Near-Infrared (NIR) band:** SR_B5
- **Shortwave Infrared 1 (SWIR1) band:** SR_B6
- **Quality layer:** QA_PIXEL

2.1.5 Area of Interest (AOI)

- Study area boundary: **Nuwara Eliya District (administrative district)**
- AOI boundary file: AOI/nuwara_eliya_District_AOI.geojson
- The workflow clips all rasters to the AOI, ensuring analysis and results are restricted to district boundaries.

2.2 Preprocessing

Preprocessing is a key component of reliable change detection. This project includes several steps to improve stability and reduce false detections.

2.2.1 Scene ingestion and integrity checks

The pipeline scans the Landsat directory and indexes valid scenes by ensuring each scene contains all required files (SR_B4, SR_B5, SR_B6, QA_PIXEL) and that files are not empty. The notebook exports: - scenes_index.csv (usable scenes) - skipped_scenes.csv (skipped scenes + reasons)

This step improves reproducibility and prevents missing-band scenes from silently contaminating results.

2.2.2 AOI clipping

Each raster band is clipped to the AOI polygon. Clipping reduces compute, ensures consistent study extent, and avoids reporting change outside Nuwara Eliya District.

2.2.3 Alignment to a reference grid

After clipping, each band/index is aligned to a consistent reference grid (derived from the first readable scene). This prevents subtle pixel shifts between scenes and supports stable multi-year per-pixel comparisons.

2.2.4 Surface reflectance scaling

Landsat Collection 2 Level-2 SR uses scaled integers. The pipeline applies the standard conversion: [SR = DN - 0.2] This ensures NDVI/NDMI values are computed correctly and are consistent across dates.

2.2.5 Cloud/shadow/snow masking using QA_PIXEL

Clouds and their shadows can strongly distort NDVI/NDMI and create false “loss” signals. The notebook uses QA_PIXEL bits to define clear pixels. Pixels are masked if they are flagged as: - Fill - Dilated cloud - Cirrus - Cloud - Cloud shadow - Snow

Only pixels classified as clear are used in index composites.

2.2.6 Yearly median compositing (robust aggregation)

For each year, the model computes **yearly median composites** (per pixel): - Median NDVI - Median NDMI

Median compositing is robust to outliers and reduces noise from remaining atmospheric effects or partial contamination.

Additionally, the pipeline stores a per-pixel **valid observation count** (valid_count) to ensure yearly composites are based on sufficient clear observations.

2.3 Methods Used

2.3.1 Indices used

Two indices were used to capture both greenness and moisture changes:

NDVI (Normalized Difference Vegetation Index):

[NDVI =]

NDMI (Normalized Difference Moisture Index):

[NDMI =]

NDVI is sensitive to vegetation greenness and canopy density, while NDMI is sensitive to vegetation moisture and disturbance. Using both helps reduce confusion between temporary greenness changes and real canopy removal.

2.3.2 Baseline forest definition (2013–2015)

The baseline forest mask is created using the baseline median NDVI composite:

- Baseline years: **2013–2015**
- Forest threshold:
 - `forest_ndvi_threshold = 0.55`

Pixels are considered baseline forest if:

- baseline NDVI is valid and supported by enough observations, and
- baseline NDVI ≥ 0.55 .

This produces:

- `baseline_ndvi.tif`
- `baseline_ndmi.tif`
- `baseline_forest_mask.tif`

2.3.3 Candidate deforestation detection (per year)

For each analysis year (2016–2025), the method computes drops relative to baseline:

$$[\text{NDVI} = \text{NDVI}_{\{\text{baseline}\}} - \text{NDVI}_{\{\text{year}\}}] [\text{NDMI} = \text{NDMI}_{\{\text{baseline}\}} - \text{NDMI}_{\{\text{year}\}}]$$

A pixel becomes a **candidate loss** in year y if:

- It was baseline forest, and
- It has sufficient valid observations (`min_valid_obs_per_year`), and
- NDVI drop is large and NDVI becomes non-forest-like, and - NDMI drop is also large.

Key thresholds from the run configuration:

- `ndvi_drop_threshold = 0.20` - `ndmi_drop_threshold = 0.10`
- NDVI must drop below forest threshold (year NDVI < 0.55)
- `min_valid_obs_per_year = 4` (for Nuwara Eliya)

This makes detection more conservative and reduces false positives due to transient effects.

2.3.4 Persistence rule (confirmation)

To confirm deforestation, the project uses a **persistence rule**:

- `min_persistence_years = 2`

A pixel is assigned a deforestation year only if it is flagged as candidate loss for **two consecutive years**. The assigned year is the **first year** of that consecutive sequence.

This produces:

- deforestation_year.tif
 - Value 0 = no detected loss
 - Value = year of first persistent loss (e.g., 2018)

 - deforestation_mask.tif
 - 1 = deforestation detected (since baseline)
 - 0 = no detected loss
-

3. Results and Discussion

3.1 Main outputs produced by the workflow

The notebook generates both raster and tabular products:

Raster outputs

- Yearly composites: yearly/ndvi_YYYY.tif, yearly/ndmi_YYYY.tif
- Baseline layers: baseline_ndvi.tif, baseline_ndmi.tif, baseline_forest_mask.tif
- Deforestation results: - deforestation_year.tif
- deforestation_mask.tif

Tabular and reporting outputs

- Deforestation_Yearly_Stats.csv
- Plots in figures/
- Run metadata and logs (run_config.json, scenes_index.csv, etc.)

3.2 Deforestation statistics (2016–2025)

From your provided Deforestation_Yearly_Stats.csv summary:

- **Baseline forest area (2013–2015 baseline):** 1352.1879 km²
- **Total detected forest loss by 2025:** 5.3982 km²
- **Percent loss of baseline by 2025:** ~0.399%

Annual new loss (km²) during the analysis period:

Year	New loss (km ²)	Cumulative (km ²)	% baseline lost
2016	0.7137	0.7137	0.0528%
2017	0.4977	1.2114	0.0896%
2018	0.7803	1.9917	0.1473%
2019	0.9891	2.9808	0.2204%
2020	0.6606	3.6414	0.2693%
2021	0.2763	3.9177	0.2897%
2022	0.8136	4.7313	0.3499%
2023	0.3132	5.0445	0.3731%
2024	0.3537	5.3982	0.3992%
2025	0.0000	5.3982	0.3992%

Interpretation of trend - The largest annual detected loss is in **2019 (~0.99 km²)**. - Loss continues through **2024**, with moderate annual increments. - **2025 shows 0.0 km²**, which may indicate either:

- genuinely low or no detected persistent loss in that year, or
- reduced ability to detect loss due to observation availability or incomplete-year coverage. Because the method requires persistence across consecutive years, the final year in a time series can sometimes show artificially low “new loss” if the next year is not available to confirm persistence. This is an important methodological consideration when interpreting the last year.

3.3 Strengths of the approach

1. **No labels required:** The workflow is suitable when ground truth is limited.
2. **QA-based masking:** Reduces cloud/shadow false positives.
3. **Median composites:** Increase robustness under variable cloud cover.
4. **Persistence rule:** Helps distinguish real canopy conversion from temporary vegetation stress or noise.
5. **Clear documentation and reproducibility:** Saved configuration files and scene indexes support transparency.

3.4 Limitations and uncertainty

1. Index-threshold dependence

NDVI and NDMI thresholds can vary by ecosystem type and season. Montane forest, tea plantations, and mixed vegetation in Nuwara Eliya may have different baseline index ranges. A single threshold ($NDVI \geq 0.55$) may misclassify some land cover types.

2. Forest definition vs plantations

NDVI-based baseline forest masks can include tree plantations or exclude sparse natural forest edges. If the project’s definition of “forest” aims to match legal/protected forest, additional land cover masks or reference datasets may be needed.

3. 30 m resolution constraints

Landsat may miss small clearings or narrow linear disturbances smaller than a pixel. Mixed pixels in heterogeneous landscapes can reduce sensitivity.

4. Dependence on observation density

The requirement `min_valid_obs_per_year = 4` improves reliability but may exclude some pixels/years with insufficient clear observations.

5. Persistence rule and end-of-series effect

The 2-year persistence requirement is beneficial but can cause undercounting in the final year if future confirmation is not available (e.g., 2025 loss may not be confirmed without 2026 data).

4. Prediction Model (Forecasting Future Deforestation)

4.1 Why include a prediction component?

Monitoring past loss answers “what happened.” A prediction/forecasting extension supports planning by estimating “what might happen next” if recent trends continue. For district-level management, a forecast can guide:

- resource allocation for monitoring and enforcement,
- prioritization of field inspections,
- awareness of potential cumulative impacts.

4.2 What can be predicted using the available outputs?

Using only `Deforestation_Yearly_Stats.csv`, the feasible prediction is a **non-spatial forecast** of:

- `new_loss_km2` (annual forest loss in km^2/year), and derived: - `cumulative_loss_km2`, - `percent_loss_of_baseline`.

This does **not** predict *where* deforestation will occur. Spatial forecasting would require historical per-pixel loss labels plus driver variables (roads, slope, proximity to settlements, etc.).

4.3 Data constraints for forecasting

Your time series has only **10 annual points** (2016–2025). With such short series:

- complex models can easily overfit,
- uncertainty is high,
- the most defensible approach is to compare simple baselines and simple time-series models with proper backtesting.

Also, 2025 has `new_loss_km2 = 0.0`. This could be real, but could also be influenced by the method’s **2-year persistence rule** (loss starting in 2025 may not yet be confirmed). Therefore, forecasts should either:

- (a) exclude 2025 from training, or
- (b) treat 2025 as “lower confidence” and compare results both with and without it.

4.4 Forecasting approach

Target: forecast `new_loss_km2` for 2026–2030.

Models to include (minimal but defensible):

1. **Naïve baseline:** moving average of the last 3–5 years (e.g., 2020–2024).

This is often a strong baseline for short time series. 2. **ARIMA / Exponential Smoothing (ETS):** lightweight time-series models that can capture gradual trend changes. 3. **(Optional) Prophet (trend-only):** can be used, but with annual data and short series, it should be configured without seasonality.

Post-processing constraints:

- enforce non-negativity: $\widehat{\text{new_loss}} \geq 0$

Derive cumulative and percent loss: $\widehat{\text{cumulative_loss}}_t = \text{cumulative_loss}_{2025} + \sum_{y=2026}^t \widehat{\text{new_loss}}_y$ $\widehat{\text{percent_loss}}_t = 100 \times \frac{\widehat{\text{cumulative_loss}}_t}{\text{baseline_forest_area}}$

4.5 Model evaluation (backtesting)

Instead of one train/test split, use **rolling-origin backtesting**: - train on 2016–2020 → predict 2021

- train on 2016–2021 → predict 2022

- ...and so on.

Evaluate with: - MAE, RMSE

This is more reliable when the dataset is small.

4.6 Interpretation and limitations of forecasts

A univariate forecast (using only past annual loss) assumes that recent behavior continues. In reality, deforestation depends on:

- policy enforcement,
- commodity prices and agricultural expansion,
- development projects,
- climate extremes and disasters.

Therefore, forecasts should be presented as **scenario projections** with uncertainty ranges, not as guaranteed outcomes.

5. Conclusion

This project developed and applied a **time-series, unlabeled deforestation detection model** for **Nuwara Eliya District, Sri Lanka** using **Landsat Collection 2 Level-2 Surface Reflectance imagery (2013–2025)**. The workflow used rigorous preprocessing including AOI clipping, SR scaling, and QA-based cloud/shadow masking, followed by yearly median compositing of NDVI and NDMI.

Deforestation was detected as a **persistent decrease** in vegetation greenness and moisture relative to a baseline forest state (2013–2015). Based on the produced yearly statistics, the baseline forest area was estimated at **~1352.19 km²**, and the cumulative detected forest loss by 2025 was **~5.40 km²**, representing **~0.40%** of the baseline.

Implications for environmental management in Sri Lanka

- The district-scale forest-loss estimates can support **local planning, conservation prioritization, and monitoring**.
- The produced **deforestation year map** is valuable for identifying *when* forest loss occurred and for focusing field inspections or policy interventions on key periods and locations.

- For stronger policy relevance, future work should include:
 - validation using high-resolution imagery and/or field observations,
 - refinement of forest definitions (distinguishing natural forest from plantations),
 - extension with additional sensors (Sentinel-2, higher resolution) to detect smaller clearings and improve confidence.

A forecasting extension is feasible using the annual loss time series (`new_loss_km2`) to project possible district-wide deforestation totals for 2026–2030. However, because the series is short and the final year may be affected by the persistence rule, forecasts should be interpreted cautiously and validated using rolling backtesting and uncertainty intervals.

Overall, the study demonstrates how satellite imagery and careful time-series processing can provide repeatable, transparent evidence for monitoring forest-cover change in Sri Lanka, even in the absence of detailed labeled training data.

Appendix — Key configuration parameters

- Baseline years: 2013–2015
- Analysis years: 2016–2025
- Forest threshold: $\text{NDVI} \geq 0.55$
- NDVI drop threshold: ≥ 0.20
- NDMI drop threshold: ≥ 0.10
- Minimum persistence: 2 consecutive years
- Minimum valid observations per year: 4