Tree Level Change Detection Across Maharashtra State using NDVI and Machine Learning Techniques

G30 GROUP
Rajarambapu Institute of Technology, Islampur

August 2025

1 Introduction

Monitoring vegetation health over time is critical for climate studies, forest preservation, agriculture planning, and urban development. By analyzing the greenness of an area via satellite imagery and using image processing and machine learning techniques, we can classify the severity of vegetation loss or gain.

2 Why is NDVI Important?

What is NDVI?

The **Normalized Difference Vegetation Index (NDVI)** is a widely used remote sensing index that quantifies vegetation health by analyzing the difference between the green and red reflectance from vegetation surfaces.

$$\text{NDVI} = \frac{G - R}{G + R + \varepsilon}$$

Where:

- G = Green channel intensity
- R = Red channel intensity
- $\varepsilon = 10^{-5}$ is a small constant to avoid division by zero

NDVI values lie in the range $[-1, 1]$, and their meaning is interpreted as follows:

- NDVI ≈ 1: Dense, healthy vegetation
- NDVI ≈ 0: Sparse vegetation or barren land
- NDVI < 0: Water, snow, or non-vegetated surfaces

Why is NDVI Important?

NDVI plays a central role in this project and has wide-ranging applications, including:

- **Vegetation Monitoring:** NDVI is a robust indicator of the presence and health of vegetation.
- **Change Detection:** NDVI trends over time reveal areas undergoing deforestation, afforestation, or urban expansion.
- **Climate Studies:** Seasonal and long-term NDVI analysis supports environmental and climate research.
- **Agriculture:** NDVI is used to monitor crop health, detect stress, and manage irrigation.
- **Remote Sensing Standard:** NDVI is used by NASA, ISRO, and local governments for ecological monitoring.

Role in This Project

In this system, NDVI serves as the foundational metric for zone classification:

- NDVI is computed for every image per Taluka and Year.
- Linear Regression is applied on yearly NDVI values to detect vegetation trends.

- The slope of the NDVI trend is used to classify each Taluka into:
  - **Red Zone:** Significant vegetation loss
  - **Orange Zone:** Slight or no change
  - **Green Zone:** Vegetation improvement

Summary Table

NDVI Use	Outcome
Per Image	Captures the greenness or health of vegetation for a specific year
Over Time (Regression)	Detects increasing or decreasing vegetation trends
Final Output	Enables zone classification for environmental awareness and action

Table 1: Summary of NDVI's Role

GLCM – Texture Feature Extraction

GLCM (Gray Level Co-occurrence Matrix) is a method to extract **texture features** by analyzing how often pairs of pixel intensities occur together in an image at a certain distance and direction.

We extract the following features:

- **Contrast:** Measures intensity variation between neighboring pixels.
  $\Rightarrow$ High contrast = sharp texture or edges.
- **Homogeneity:** Measures closeness of pixel pairs to the diagonal in GLCM.
  $\Rightarrow$ High homogeneity = smoother texture.
- **Entropy:** Measures randomness or complexity in texture.
  $\Rightarrow$ Higher entropy = more complex, irregular texture.

Importance in Our Project

- Captures spatial texture patterns in satellite/tree images.
- Helps classify zones (green, orange, red) based on surface texture, not just color.
- Improves machine learning model performance with richer, more meaningful features.
- Works even when shape or color is similar but texture differs.

3 Feature Extraction Pipeline

3.1 Input Data

Images are organized in folders by District and Taluka. Each image filename contains the year.

3.2 Steps

1. Traverse all images using `os.listdir`.
2. Resize to 256×256 using OpenCV.
3. Split into R, G, B channels.
4. Calculate NDVI and Green/Red Ratio.
5. Convert to grayscale and calculate GLCM.
6. Extract GLCM features: Contrast, Homogeneity, and Energy.
7. Save features in `features.csv`.

3.3 Green/Red Ratio

$$\text{Green/Red Ratio} = \frac{\mu_G}{\mu_R + \varepsilon}$$

3.4 GLCM Features (Mathematical Definitions)

Using `graycomatrix` and `graycoprops`, we calculate:

- **Contrast**:
$$\text{Contrast} = \sum_{i,j} |i - j|^2 \cdot P(i, j)$$
- **Homogeneity**:
$$\text{Homogeneity} = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|}$$
- **Energy** (Angular Second Moment):
$$\text{Energy} = \sum_{i,j} P(i, j)^2$$

4 Code Snippet (Feature Extraction)

Listing 1: Feature Extraction from Image

```
img = cv2.imread(img_path)
img = cv2.resize(img, (256, 256))
b, g, r = cv2.split(img)
ndvi = (g.astype(float) - r.astype(float)) / (g + r + 1e-5)
mean_ndvi = np.nanmean(ndvi)
gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
glcm = graycomatrix(gray, distances=[1], angles=[0], levels=256,
                    symmetric=True, normed=True)
contrast = graycoprops(glcm, 'contrast')[0, 0]
```

5 Trend Analysis using Linear Regression

5.1 Model

We use **Linear Regression** to find NDVI trends over years per taluka.

$$NDVI = m \cdot \text{Year} + c$$

5.2 Interpretation of Slope m

- $m < -0.02 \Rightarrow$ Red Zone (Severe decline)
- $-0.02 \leq m < 0.01 \Rightarrow$ Orange Zone (Moderate or stable)
- $m \geq 0.01 \Rightarrow$ Green Zone (Improvement in vegetation)

6 Code Snippet (Zone Classification)

Listing 2: Linear Regression for Zone Classification

```
model = LinearRegression()
model.fit(X, y)
slope = model.coef_[0]
if slope < -0.02:
    zone = 'Red Zone'
elif -0.02 <= slope < 0.01:
    zone = 'Orange Zone'
else:
    zone = 'Green Zone'
```

7 Output Files

- `features.csv`: Contains extracted features per image.
- `zone_classification.csv`: Contains NDVI trend and zone per taluka.

8 Sample Output Table

District	**Taluka**	**NDVI Trend**	**Zone**
Pune	Haveli	-0.0243	Red Zone
Nashik	Igatpuri	0.0131	Green Zone
Kolhapur	Karvir	-0.0062	Orange Zone

Table 2: Zone Classification Output

9 Applications

- Forest monitoring and conservation
- Agricultural planning
- Urban green cover management
- Smart city environmental assessment

10 Conclusion

This system successfully combines remote sensing, image processing, and machine learning to assess vegetation change. The zone classification helps prioritize areas for awareness, afforestation, or environmental interventions.