Urban Green Space Optimization using Deep Learning and GIS

Upayan Choudhuri, Vedant Vartak, Sandip Shaw
Symbiosis Institute of Geoinformatics
Symbiosis International University, Pune
Maharashtra, India

***Abstract*— Rapid urbanization has significantly reduced urban green spaces, directly impacting millions of citizens and environmental sustainability and subsequently hampering the quality of life. This study combines both Deep Learning and Geographic Information System (GIS) to automatically detect and analyze urban green spaces using satellite images. A Convolutional Neural Network (CNN), specifically a U-Net architecture, was trained on high-resolution Sentinel-2 imagery to perform semantic segmentation of green areas. The model's predictions were post-processed and analyzed in ArcGIS to evaluate the spatial distribution, accessibility, and connectivity of green spaces across the city. The results provide actionable insights into urban planning, highlighting neighborhoods with insufficient green coverage and guiding policymakers toward data-driven environmental interventions.

Keywords—natural language processing, neural networks, geospatial intelligence, artificial intelligence, large language models, Information extraction

I. Introduction

Safety Green spaces, such as parks, gardens, and tree-lined streets play a vital role in urban sustainability by improving air quality, reducing heat islands, and enhancing public well-being. However, with continuous urban expansion, monitoring and managing these spaces at a city-wide scale remains a major challenge.

Traditional methods for assessing green cover rely on manual surveys and coarse land-use maps, which are time-consuming and lack spatial precision. Advances in deep learning and remote sensing now enable automated, pixel-level classification of urban landscapes from satellite imagery. In parallel, Geographic Information Systems (GIS) provide a robust environment to analyze spatial patterns and accessibility metrics.

This research integrates both technologies to develop an automated workflow for urban green space detection and accessibility analysis, using:

- CNN-based image segmentation to identify vegetation areas, and
- GIS-based spatial analytics to measure green space equity and accessibility.

II. Study Area

The study focuses on multiple metropolitan cities like Cairo, Mumbai, Pune, New York, Oslo, Sao Paulo. These are rapidly urbanizing areas with diverse land-use characteristics.

1. Satellite Data
    a. Source: Sentinel-2 imagery from the Copernicus Open Access Hub.
    b. Resolution**:** 10-meter spatial resolution.
    c. Bands Used**:** RGB and Near-Infrared (NIR).
    d. Projection System**:** WGS 84 / UTM Zone
2. Ground Truth and Annotation: The annotations were done using

III. Methodology

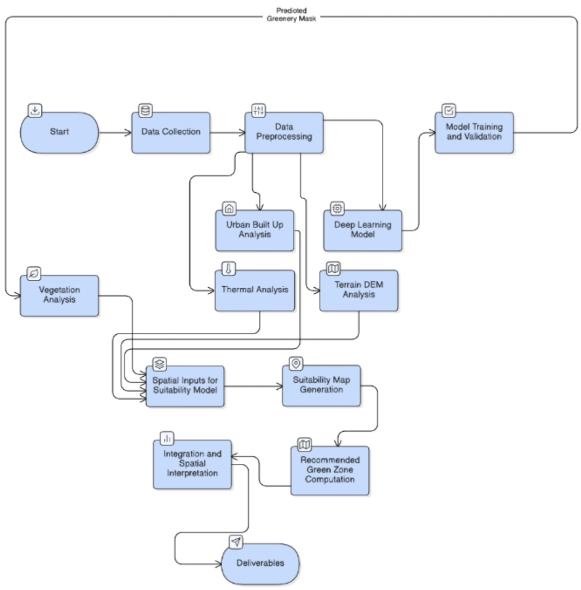

Figure 1: Methodology Diagram

A. Data Collection

This was the first step, which involved extraction city tif files from sentinel-2 having 4 bands. The files were then tiled into 256 x 256-pixel patches followed by normalization for effective model convergence. Multiple data augmentation techniques were used such as rotation, flipping, brightness and contrast variation to improve model robustness and mitigate overfitting.

The corresponding masks were also generated using NDVI thresholding for vegetation (>0.3) which are basically the truth masks. The dataset was split into train (70%), validation (20%) and test (10%) subsets.

B. Deep Learning Model for Green Space Detection

- A **U-Net** CNN architecture was used, known for its encoder–decoder structure suitable for semantic segmentation.
- The encoder captured spatial context, while the decoder reconstructed fine boundary details of vegetation patches.
- Training Parameters

- Framework: Pytorch
- Loss Function: Dice Loss
- Optimizer: Adam
- Learning Rate: 1e-5
- Batch Size = 2
- Epochs = 25

- Evaluation Metrics: Intersection over Union(IOU) and Dice Coefficient were used to evaluate segmentation accuracy.
- The trained model was applied to full-sized Sentinel-2 images and resulting raster predicted the probability (0–1) of each pixel being green space.

Validation Accuracy: 84%
Validation Loss: 0.9225

Figure 2: Train and Validation plots

C. GIS Spatial Analysis

This phase integrates the outputs from remote sensing indices and GIS-based terrain analysis to identify and prioritize optimal locations for urban greenery expansion. The analysis combines biophysical, thermal, and urban morphology parameters derived from Sentinel-2 satellite imagery, Digital Elevation Models (DEM), and thermal datasets.

Greenery Detection: To identify areas covered by vegetation. The input data was RGB and Near-Infrared (NIR) bands from Sentinal-2 imagery.
Vegetation was delineated using the **Normalized Difference Vegetation Index (NDVI)**, which quantifies the presence and health of vegetation based on reflectance values:

NDVI = (NIR - RED) / (NIR + RED)

A threshold value of **NDVI > 0.3** was applied to classify pixels as vegetated areas. The resulting raster, **GreeneryMask.tif**, represents existing green cover, with white pixels denoting vegetation and black pixels representing non-green areas.

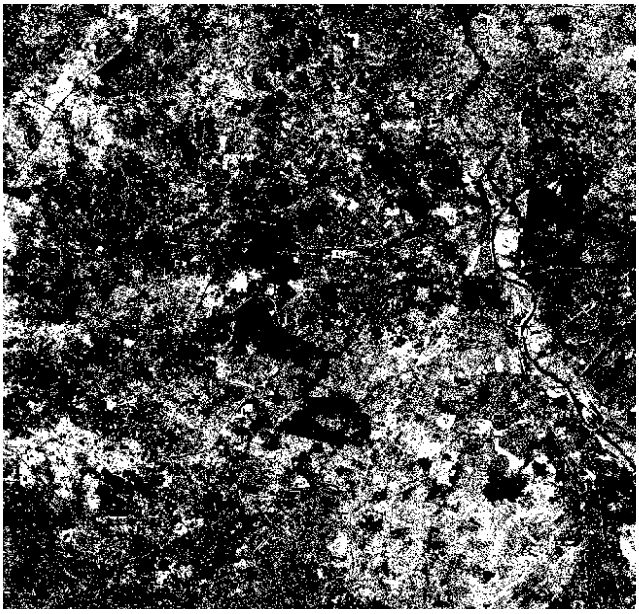

Figure 3: Binary Mask

Urban/Built-Up Area Detection: To delineate built-up and urbanized regions.
The input data was Sentinel-2 SWIR and NIR bands and Urban areas were extracted using the **Normalized Difference Built-up Index (NDBI)**:

NDBI = (SWIR - NIR) / (SWIR + NIR)

The resulting NDBI raster was thresholded and reclassified to distinguish urban (built-up) from non-urban surfaces. This mask was later used to exclude unsuitable urban patches from the greenery expansion analysis.

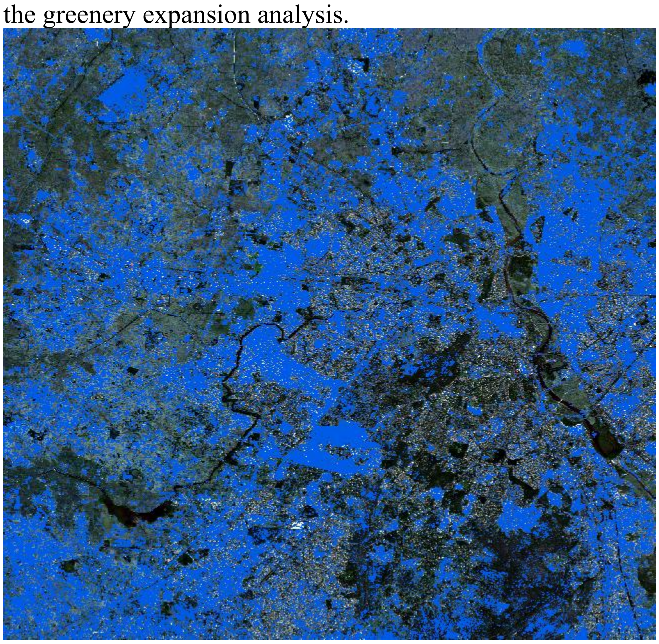

Figure 4: Urban Patch

Terrain and Slope Analysis: To assess topographical suitability for vegetation growth and plantation feasibility.
The input data was Digital Elevation Model (DEM) and using the DEM, a **slope raster** was generated in ArcGIS. The slope values were then **reclassified into suitability classes**, where

flatter terrains received higher suitability scores. Steep slopes were considered less favorable for vegetation due to soil instability and drainage challenges.
Temperature and Thermal Analysis: To identify thermally stressed areas that would benefit most from increased green cover.
The input data was Thermal Infrared (TIR) band or derived **Land Surface Temperature (LST)** layer. And The LST values were normalized using **Z-score standardization** to ensure consistency across varying temperature ranges. The resulting raster, **NormLST.tif**, highlights urban heat island zones, regions where vegetation introduction would yield maximum cooling impact.

Suitability Map Generation: To integrate topographic and urban factors into a unified suitability model for potential greenery development.
The input data are slope.tif and UrbanMask.tif. The slope suitability raster was combined with the inverse of the urban mask to exclude built-up areas. The combined raster was computed as:

$$\text{Suitability} = \text{Slope Class} \times (1 - \text{Urban Mask}) \times (1 - \text{Greenery Mask factor}) \times (1 - \text{NDWI factor})$$

This multiplication ensures that only non-urban and topographically feasible regions are marked as suitable for green expansion. The output raster, **SuitabilityMap.tif**, provides a preliminary indication of feasible green space zones.

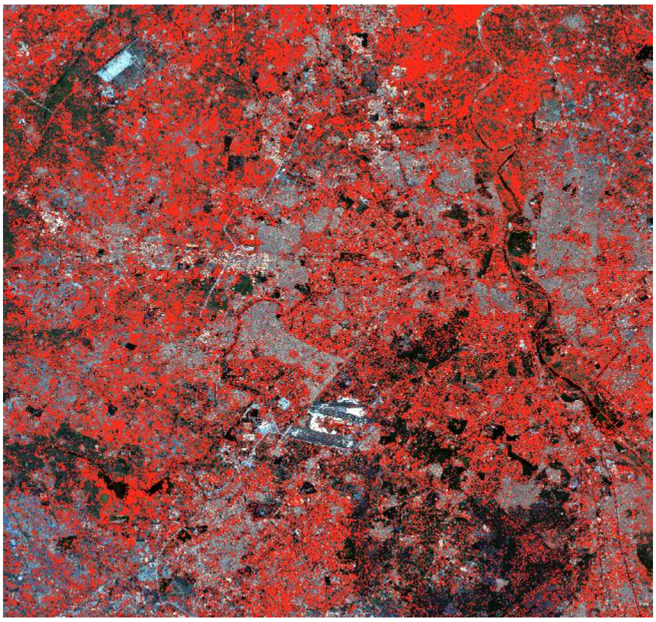

Figure 5: Possible green zones

Recommended Green Zones Identification: To delineate final zones for new greenery development based on environmental, spatial, and thermal parameters.
The input data were SuitabilityMap.tif, GreeneryMask.tif, and NormLST.tif.

Existing green cover was excluded from the suitability map to prevent redundancy using:

$$\text{Green Exclusion} = 1 - \text{GreeneryMask}$$

A composite index was then computed as:

$$\text{Recommended Score} = \text{Suitability} \times \text{Norm LST}$$

The resulting raster was reclassified into **priority zones,** high, medium, and low suitability areas for greenery expansion. Finally, the raster was converted to vector format for spatial visualization and planning in GIS.
Output:

- Raster: RecommendedGreenZones.tif
- Vector: RecommendedGreenZones.shp

Integration and Spatial Interpretation:
The integration of NDVI, NDBI, slope, and LST layers enables a multi-criteria evaluation (MCE) approach for green space suitability analysis. High-scoring zones on the **RecommendedScore** map represent **hot, flat, non-urbanized regions** that currently lack vegetation ideal for tree plantation and park development.
The final **Recommended Green Zones** layer serves as a **spatial decision-support tool** for urban planners, allowing targeted interventions to reduce urban heat islands and improve ecological balance.

Figure 6: Recommendation based on temperature

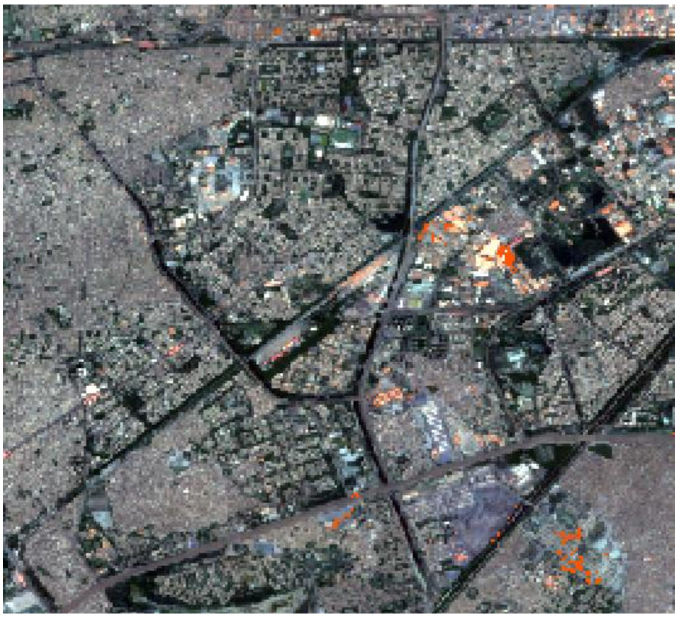
Figure 7: Recommendation based on temperature