Graduation Project Report

AI & Data Science Track

Land Type Analysis Project

Satellitor

Powered by Alpha V

Team member	Contact
Mohamed Waleed	1141020230362@stud.cu.edu.eg
Sherin Mohamed	s-sherin.kamal@zewailcity.edu.eg
Aly El Din El-Badry	Aibakgaming747@gmail.com
Dina Zahran	dinazahran1718@gmail.com
Amr Yasser	Amr855708@gmail.com

Table of Contents

1. Project Overview

Land use decisions in Egypt are often hampered by fragmented data, outdated soil assessments, and a lack of integrated tools for precision analysis. Farmers struggle to identify optimal crops for their land, and investors lack reliable metrics to evaluate site potential. Traditional methods rely on manual surveys or disjointed datasets, resulting in inefficiencies, higher costs, and missed opportunities for sustainable land management. **Satellitor** addresses these challenges as an advanced web-based platform that leverages high-resolution satellite imagery to analyze land segments in Egypt using high-resolution satellite imagery. Users can select a specific area via an interactive map, and the system processes the satellite data to segment the land type (e.g., agricultural, urban, water) and perform in-depth analyses, including humidity, temperature, pH levels, etc. Based on these insights, Satellitor generates two key recommendations:

1. **Optimal Crop Suggestion** – Recommends the most suitable crops for the analyzed land.
2. **Fertilizer Recommendation** – Suggests appropriate fertilizers to enhance soil quality.

Finally, the system compiles all findings into a detailed PDF report, providing users with actionable insights for agricultural planning, environmental monitoring, or land management.

2. Scope and Objectives

The platform caters to geologists conducting terrain studies, land investors evaluating site potential, and researchers analyzing environmental trends. It also serves urban planners and real estate developers, who require precise land evaluations to determine optimal usage—whether for agriculture or construction. Satellitor aims to achieve a minimum of 90% accuracy in land classification, deliver results within less than 30 seconds for standard requests, and support high-volume usage with seamless scalability. **Future enhancements will prioritize integration.**

3. Methodology

a) Project Evolution and Approach Selection

Our project initially focused on standard land classification but evolved to adopt semantic segmentation after recognizing its superior capability for satellite imagery analysis. Unlike basic classification, this approach provides pixel-level precision, enabling detailed boundary detection and multi-class land cover mapping. This shift was critical to deliver accurate, user-driven land analysis, particularly for applications requiring exact spatial understanding of land features.

b) Data Acquisition and Preprocessing

The data set used for model development was organized into two distinct batches to ensure progressive performance improvement:

I. Initial Batch (Pre-Training)

The primary dataset was sourced from the DeepGlobe Land Cover Classification Dataset (803 satellite images), obtained from Kaggle (https://www.kaggle.com/datasets/balraj98/deepglobe-land-cover-classification-dataset?utm_source=chatgpt.com). Using Roboflow, we annotated these images with five key land cover classes: Urban Land, Agricultural Land, Water Bodies, Forest, and Barren Land.

The preprocessing steps we applied on the images included resizing all the images to be standardized to 512×512 pixels for model compatibility, and augmentation to enhance dataset diversity, where we implemented:

- **Flip**: Horizontal and vertical flipping.
- **Crop**: Random cropping (0% minimum zoom, 20% maximum zoom).
- **Rotation**: Bounded between -10° and +10°.

After augmentation and cleaning, the dataset expanded to 2,818 images, split into 2,586 images for training and 232 images for testing. No validation set was allocated in this phase.

II. **Second Batch (Post-Training Enhancement)**

To add some enhancement to break the limitation of the model capabilities captured directly from the satellite map, integrated into the Satellitor platform. We trained the model on a second patch of 280 images. These images were annotated following the same class taxonomy and preprocessing pipeline, further refining the model's accuracy and adaptability to real-world scenarios.

A preprocessing was applied:

- resize 512x512

An augmentation:

- Flip: Horizontal, Vertical
- Crop: 0% Minimum Zoom, 20% Maximum Zoom
- Saturation: Between -25% and +25%

where the no. of images after augmentation, cleaning, and adding them to the first patch is 3576 images (3303 train, 137 val, 136 test)

III. Third Batch (Post-Training Enhancement)

We identified the need for additional data to address performance limitations during model evaluation. A supplementary batch of 500 high-resolution images was collected, and this dataset is randomly captured from the mapliter map, where this data has been annotated.

- Applied preprocessing: resize 512x512
- applied augmentation:
- Flip: Horizontal, Vertical
- Crop: 0% Minimum Zoom, 20% Maximum Zoom
- Saturation: Between -25% and +25%

where the no. of images after augmentation, cleaning, and adding them to the first patch and second patch is 4798 images (4446 train, 176 val, 176 test)

NOTE: In this dataset (patch1,patch2,patch3), the forest class has been removed, and any labeled images with class forest have been changed to be Agriculture

c) System Architecture and Model Development

The application was built using React for the frontend (ensuring an interactive map interface) and Flask for backend operations (handling API requests and model inference). We evaluated two segmentation models:

- **YOLOv11-Seg (Nano):** Prioritized for its lightweight efficiency.
- **U-Net:** Selected for its higher accuracy in pixel-wise segmentation.

d) Land Analysis Features

The system provides four analytical outputs designed for practical decision-making:

- **Class Percentage Distribution:** Quantifies land cover composition (e.g., 60% agricultural, 15% urban).
- **Boundary Dynamics:** Tracks changes in land/water edges over time, useful for environmental monitoring.
- **Fragmentation Index:** Measures land parcel cohesion, aiding ecological assessments.
- **Crop Recommendation System:** Suggests optimal crops by integrating humidity and temperature (NASA POWER-LARC dataset) which gets the average annual temperature and humidity, Rainfall which gets the average perception for the year and Soil pH which gets the average ph from 0m to 100m (Google Earth OpenLandMap dataset). Recommendations are based on 35 staple Egyptian crops from the FAO Ecrop database.

e) Data Integration and Workflow

The system follows a streamlined pipeline:

- The user selects a land area via the MapTiler-integrated interface.
- The chosen model generates a segmentation mask.
- Environmental APIs fetch real-time climate/soil data.
- Analysis algorithms compute metrics (e.g., fragmentation, crop suitability).
- Results are visualized interactively and compiled into a PDF report.

4. User Workflow and Features

Satellitor offers an intuitive, step-by-step workflow designed to deliver comprehensive land analysis, crop suitability assessments, and data-driven insights, with minimal user effort. The process is structured to ensure seamless interaction from land selection to detailed report generation.

a) Land Selection:

- Upon accessing the Satellitor platform, the user initiates the process by clicking the "Start" button to activate the interactive map interface.
- The user navigates to their area of interest in Egypt.
- Once the desired location is identified, the user clicks "Start Analysis" to trigger the data processing pipeline.

b) Automated Analysis:

- The system processes the selected area using high-resolution satellite imagery and applies AI-powered segmentation to classify land types (e.g., urban, agricultural, water).
- Simultaneously, it retrieves environmental data (humidity, temperature, rainfall, pH) from integrated APIs (NASA, Google Earth).
- Advanced algorithms compute the fragmentation index, revealing land parcel cohesion or dispersion patterns.

c) Results Delivery:

The platform provides a comprehensive visual overlay of the segmented land, clearly defining boundaries and color-coding land types (e.g., green for agricultural, blue for water).

The detailed sidebar displays quantitative metrics for each land parcel:

- **Climate/Soil Data:** Includes temperature (°C), rainfall (mm), pH levels, and humidity (%).
- **Soil Nutrients:** Provides insights into nitrogen (N), phosphorus (P), and potassium (K) levels, essential for plant growth.
- **Fragmentation Index:** A numerical score (0–1) indicating land connectivity and ecological stability.
- The system generates detailed crop analysis based on the identified land type and environmental conditions, including suitability ratings for staple crops like wheat, barley, corn, soybean, and potato.
- For agricultural land, the system appends crop and fertilizer recommendations based on the analyzed soil nutrient levels and climatic conditions.
- All findings are compiled into a professionally formatted PDF report for further analysis or documentation, including recommendations for irrigation, nutrient management, and temperature control.

d) AI-Assisted Support – Stella

The platform features Stella, an AI-powered chatbot designed to assist users throughout the analysis process.

- **Stella** provides real-time guidance, answering user questions related to land selection, analysis settings, data interpretation, and report generation.
- Users can interact with Stella for personal recommendations or deeper insights into the land analysis results.

This conversational support ensures a smooth user experience, reducing the learning curve and enhancing decision-making confidence.

e) Report Generation

The final step includes generating a comprehensive PDF report of the analysis using a pretrained LLM model that captures all analyzed data, including soil type, nutrient content, crop suitability, and environmental factors.

- This report is designed to support data-driven decision-making, facilitating better land use planning, agricultural management, and investment analysis.

together.ai Website APIs were used to get any free model from the website until it reached this model version, which is: ***mistralai/Mixtral-8x7B-Instruct-v0.1***

With this structured workflow and AI-assisted support, Satellitor empowers users to make informed decisions about land utilization, crop selection, and long-term sustainability with ease.

5. Results and Validation

The Satellitor platform's deep learning models were rigorously tested to ensure high accuracy and reliability for land cover segmentation. This section details the performance metrics and experimental setups for each training phase, including data augmentation and model evaluation.

1. **Initial Model Testing (First Patch)**

- **Model Tested:** YOLOv11-Seg (Nano)
- **Dataset:** Initial training set containing diverse land types.
- **Results:**
    - YOLOv11-Seg (Nano) Performance:
        - mAP:50 = **0.328**
        - mAP:50-95 = **0.189**
- **Conclusion:** YOLOv11-Seg significantly outperformed initial benchmarks, establishing itself as the preferred model for further testing.

Results of first yolov11 model
(https://drive.google.com/file/d/11OageRKbEwaKZuBQtwHnzXyDtRklW5Ai/view?usp=sharing)

2. **Expanded Training (Second Patch)**

Dataset: Second Batch (Post-Training Enhancement)

- **Results:**
  - **mAP:50 = 0.567**
  - **mAP:50-95 = 0.364**

The results after training on this dataset (patch1+patch2) :
(https://drive.google.com/file/d/1hr4IneA0MEsOSfFD-OwFv-qwNbxISRIK/view?usp=sharing)

3. **Comprehensive Training (Third Patch)**

- **Dataset:** Third Batch (Post-Training Enhancement)
- **Results:**
  - **mAP:50 = 0.461**
  - **mAP:50-95 = 0.268**

The results after training on this dataset (patch1+patch2+patch3) :
(https://drive.google.com/file/d/1rXAvUykM31Zt21VDmdIHZrZNRUL57ZK-/view?usp=sharing)

The training results indicate significant performance improvements with the second patch, followed by a slight decline in mAP scores after integrating the third patch, potentially due to the removal of the forest class. Further optimization and targeted augmentation may be required to address this performance variability.

6. Challenges and Solutions

Developing a robust deep learning model for satellite image segmentation presented several technical and data-related challenges. This section outlines the major challenges encountered and the strategies implemented to address them.

1. Small Dataset Size

Problem:

The initial dataset was relatively small, limiting the model's ability to generalize across diverse land cover types. A small dataset can lead to overfitting, where the model performs well on the training data but fails to generalize to new, unseen images. This is particularly problematic in satellite image analysis, where high variability in land cover can significantly affect model accuracy.

Impact:

- Poor generalization on validation and test sets.
- High variance in performance across different patches.
- Reduced stability in model training, leading to potential overfitting.

Solution:

To overcome this limitation, we progressively expanded the dataset by adding two additional patches, each containing hundreds of new, annotated images. This approach included extensive data augmentation to simulate various real-world conditions and increase the effective size of the training set. Key augmentation techniques included horizontal and vertical flips, random crops, and saturation adjustments, which collectively increased the diversity of the training samples and improved model robustness.

Outcome:

- Significant improvement in mAP scores from **0.328** (first patch) to **0.567** (second patch).
- Enhanced model generalization, reducing overfitting and improving real-world performance.

2. Image Color Variability

Problem:

Satellite imagery captured from different sources or at varying times can exhibit significant color variability, including differences in brightness, contrast, and hue. This inconsistency can confuse the model during training, as it learns to associate certain colors with specific land cover types. These color shifts can arise due to differences in sensor calibration, atmospheric conditions, or lighting variations, leading to increased classification errors.

Impact:

- Reduced model confidence in class predictions.
- Inconsistent segmentation results, particularly for similar land types with different spectral characteristics.
- Increased false positives and false negatives due to misclassification of regions with varying color profiles.

Solution:

To address this issue, we implemented color normalization and histogram equalization techniques during preprocessing to standardize the appearance of satellite images. These methods adjust the color distribution of the images, reducing the impact of lighting and contrast variations, and ensuring more consistent feature extraction during training.

Outcome:

- Improved model robustness to varying lighting conditions.
- More stable training process with reduced overfitting.
- Higher overall mAP scores, reflecting better feature consistency across different data patches.

7. Future Work

1. **Improving Model Generalization**
   - Train the model with a larger, more diverse dataset that covers different seasons, weather conditions, and geographic regions to improve robustness and reduce overfitting.
   - Incorporate domain adaptation techniques to enhance model transferability across different satellite sources.
2. **Integration of Multispectral and SAR Data**
      - Combine optical satellite imagery with multispectral or synthetic aperture radar (SAR) data to improve segmentation accuracy,

especially in challenging environments like dense forests or urban areas.

3. **Refinement of Data Augmentation Strategies**
   - Implement more advanced augmentation methods to enhance data diversity and model performance.

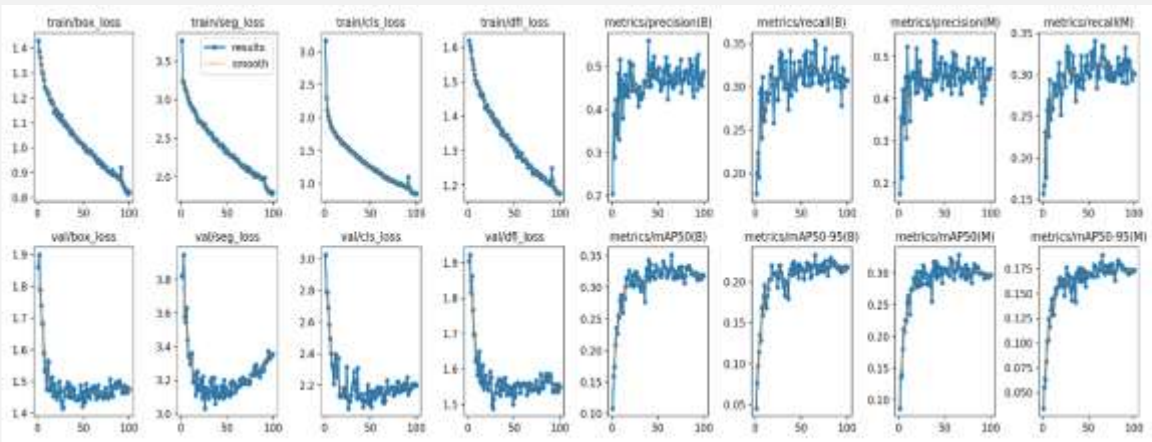
8. Conclusion

The development of the *Satellitor* platform represents a significant advancement in precision land analysis for Egypt. By integrating high-resolution satellite imagery, AI-driven segmentation models, and real-time environmental data, the system offers a robust solution to long-standing challenges in planning and agricultural decision-making. It enables users, from geologists and researchers to investors and planners, to make data-informed decisions with greater speed and accuracy. The modular architecture and scalable design ensure the platform's adaptability for future enhancements, including deeper soil diagnostics, more crop variety support, and broader regional coverage. Overall, *Satellitor* sets a strong foundation for smart, sustainable land management powered by modern geospatial and deep learning technologies.

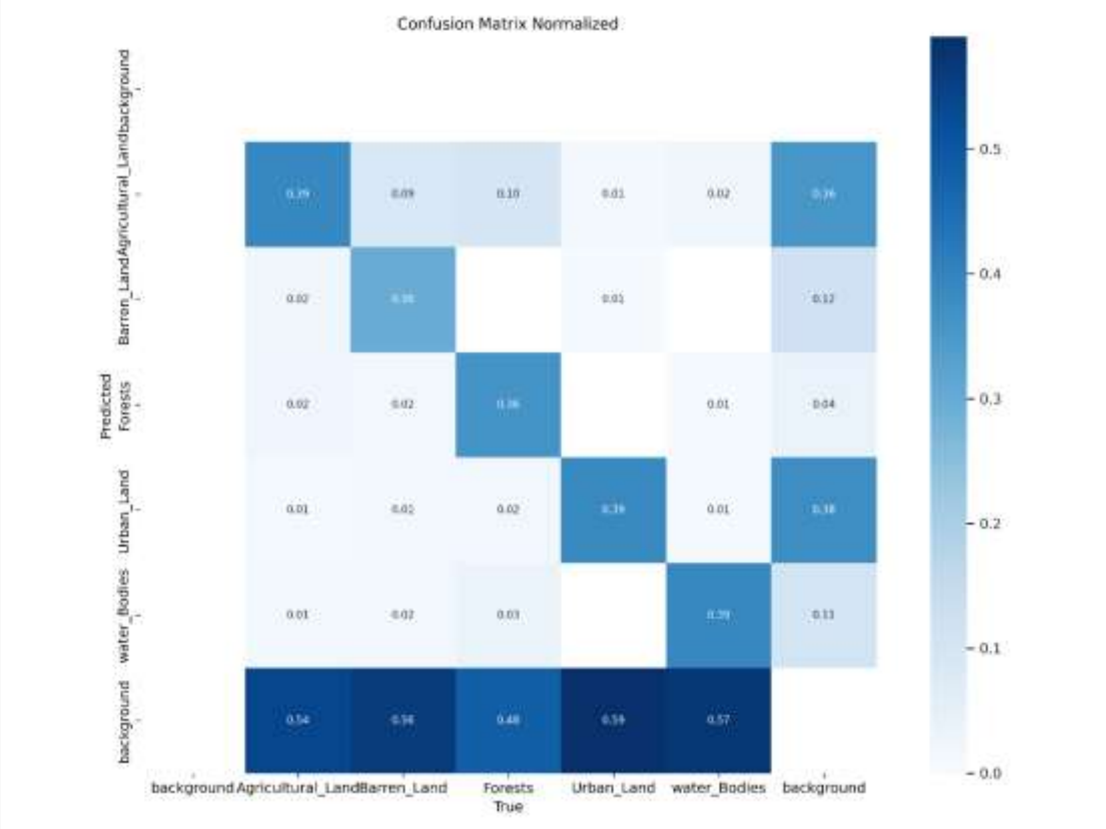
9. Appendix

I. *Model_01_Graphs:*

a. <u>Results:</u>

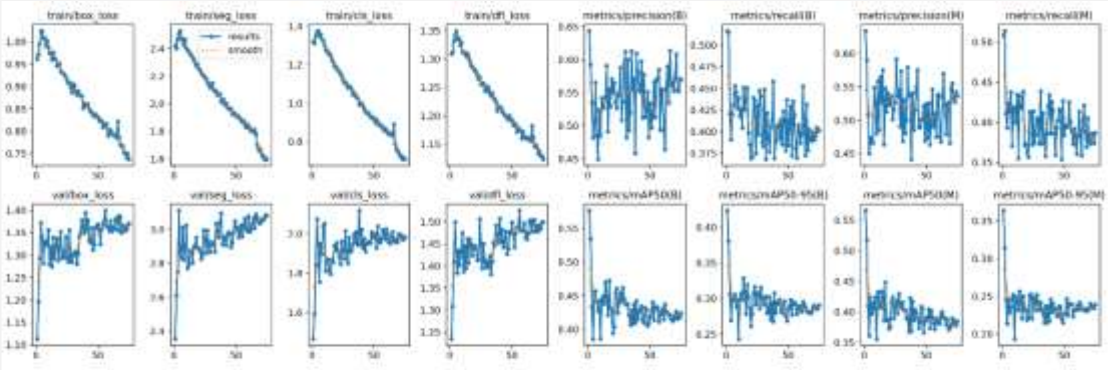


b. <u>Confusion Matrix normalized:</u>

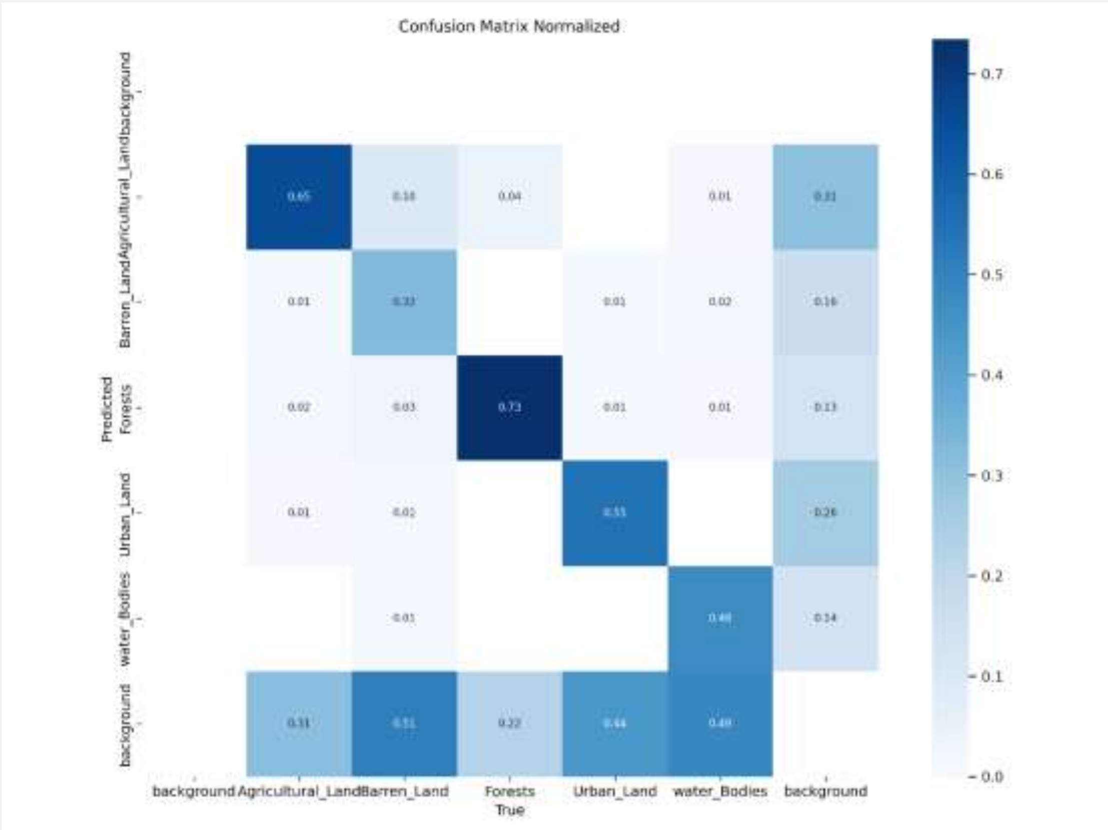


II. Model_02_Graphs:

a. <u>Results:</u>

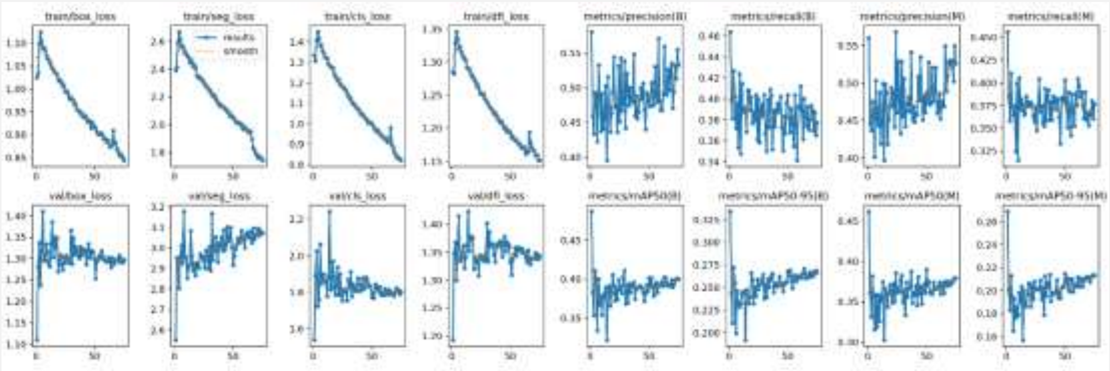


b. Confusion Matrix normalized:



III. *Model_03_Graphs*

a. Results:



b. Confusion Matrix normalized: