

# AI-DRIVEN ARCHAEOLOGICAL SITE MAPPING AND PRESERVATION SYSTEM

A Final Project Report

**Submitted by:**

Prabhakar A

B.E - CSE (AIML)

*Project Duration: 8 Weeks*

## Chapter 1: Abstract

This project presents a comprehensive AI-driven solution for the automated discovery and risk assessment of archaeological sites. Traditional survey methods are labor-intensive and often fail to detect subterranean structures or predict environmental threats effectively. This system integrates three core artificial intelligence modules: Semantic Segmentation using U-Net to identify ruins and vegetation patterns from satellite imagery, Object Detection using YOLOv5 to locate small-scale artifacts from drone feeds, and Erosion Risk Prediction using XGBoost to analyze topographical data.

The final output is a unified Streamlit dashboard that allows archaeologists to visualize site maps, detect artifacts in real-time, and assess soil erosion risks, thereby facilitating data-driven conservation strategies.

## Chapter 2: Introduction

### 2.1 Problem Statement

Archaeological sites are rapidly disappearing due to urbanization and natural erosion. Manual documentation is slow and cannot easily scale to cover large regions. Furthermore, many sites remain buried under vegetation, invisible to the naked eye but detectable through multi-spectral imaging.

### 2.2 Objectives

- To implement a U-Net model for segmenting ruins and vegetation from Sentinel-2 satellite imagery.
- To deploy a YOLOv5 model for the precise detection of small artifacts (pottery, tools) using drone imagery.
- To predict soil erosion rates using machine learning (XGBoost) on Digital Elevation Models (DEM).
- To integrate all models into a user-friendly Streamlit web dashboard.

## Chapter 3: Data Acquisition & Management

Data was collected from multiple remote sensing sources during Week 1, ensuring a diverse dataset for all three modules.

### 3.1 Data Sources

- Sentinel-1 (Radar): Used for ruins segmentation. The VV and VH polarizations help detect surface roughness and buried walls, even through cloud cover.
- Sentinel-2 (Optical): Bands B4 (Red) and B8 (NIR) were used to calculate the Normalized Difference Vegetation Index (NDVI), revealing crop marks caused by buried structures.
- Drone Imagery: High-resolution (<30 cm) RGB images were sourced via OpenAerialMap to detect small artifacts invisible to satellites.
- SRTM DEM: The Shuttle Radar Topography Mission (30m resolution) provided elevation and slope data for erosion modeling.

### 3.2 Dataset Organization

A structured folder hierarchy was established to prevent data leakage and ensure compatibility with training scripts:

- dataset/segmentation/: Contains 'images' and 'masks' (PNG) for U-Net.
- dataset/detection/: Contains 'images' and 'labels' (TXT) for YOLOv5.
- dataset/erosion/: Contains geospatial TIFFs (DEM, Slope, NDVI).

## Chapter 4: Module 1: Ruins Segmentation (U-Net)

Semantic segmentation was performed to classify every pixel in an image as 'Ruins', 'Vegetation', or 'Background'.

### 4.1 Model Architecture

The U-Net architecture was selected for its ability to work with small datasets and preserve spatial details through skip connections. A ResNet-34 backbone pretrained on ImageNet was used as the encoder to extract features such as edges and textures efficiently.

### 4.2 Training & Results

The model was trained on 512x512 images using the Adam optimizer and Dice Loss. Data augmentation (Horizontal Flip, Rotation) was applied using the Albumentations library.

#### Performance Metrics:

- Intersection over Union (IoU): 0.69
- Dice Coefficient: 0.82

The high Dice score confirms the model effectively identifies vegetation patterns that often hide archaeological ruins.

## Chapter 5: Module 2: Artifact Detection (YOLOv5)

Object detection was implemented to locate specific small artifacts such as pottery, stone tools, and structure pieces.

### 5.1 Implementation Details

YOLOv5 (You Only Look Once) was chosen for its real-time inference speed. Annotations were created using CVAT and exported in YOLO format (class\_id, x\_center, y\_center, width, height).

### 5.2 Training Results

The model was trained for 50 epochs with a batch size of 16. Training logs showed a consistent decrease in Box Loss and Objectness Loss.

#### Validation Metrics:

- mAP (Mean Average Precision): 0.488
- Precision: 0.999
- Recall: 0.42

The high precision (0.999) indicates that when the model predicts an artifact, it is highly likely to be correct, minimizing false positives.

## Chapter 6: Module 3: Erosion Risk Prediction

A machine learning approach was used to predict soil erosion rates based on environmental factors.

### 6.1 Feature Engineering

A dataset of 9,864 samples was analyzed. Key features included Slope Angle, Rainfall (mm), Soil Moisture, and Vegetation Cover. Redundant features were removed to prevent data leakage.

### 6.2 Model Comparison

Two models were trained and compared during Week 5 and 6:

1. Random Forest Regressor (Baseline)
2. XGBoost Regressor (Advanced)

XGBoost demonstrated superior performance with a lower RMSE and higher R-squared score, proving it could better handle the non-linear relationships in topographical data.

## Chapter 7: System Integration (Streamlit)

In Week 7, a unified web dashboard was developed using Streamlit to make the AI models accessible to non-technical users.

### 7.1 Dashboard Features

- Site Mapping Tab: Users upload satellite images; the U-Net model overlays segmentation masks for ruins/vegetation.
- Artifact Detection Tab: Users upload drone images; YOLOv5 draws bounding boxes around detected tools/pottery.
- Risk Assessment Tab: Interactive sliders allow users to input terrain data (e.g., slope, rainfall) and receive an instant erosion risk score.

## Chapter 8: Conclusion & Future Scope

This project successfully delivered a functional, multi-modal AI system for archaeology. The integration of U-Net, YOLOv5, and XGBoost covers the full spectrum of site analysis - from macro-scale mapping to micro-scale artifact detection and environmental risk assessment.

Future work will focus on integrating 3D terrain analysis, expanding the dataset to include multi-regional sites, and incorporating live satellite feeds for real-time monitoring.