

Machine Learning in Archaeology for Artefact Classification and Site Analysis

Abstract : Archaeology relies heavily on manual artefact identification and site assessment, which are time-consuming and prone to human error. This project introduces a machine learning-based approach to assist archaeologists in classifying artefacts and analyzing satellite imagery for site detection. Convolutional Neural Networks (CNNs) are applied to classify images of artefacts such as pottery and tools based on their visual features. Additionally, basic image processing techniques are used to analyze satellite images for identifying potential archaeological site patterns. The project is developed using Python and employs libraries such as TensorFlow, Keras, and OpenCV for model training and image analysis. The proposed solution aims to improve the speed and accuracy of artefact classification and provide a supportive tool for preliminary site exploration. By leveraging AI in this domain, the system demonstrates the potential of technology in preserving and understanding historical data more effectively.

Keywords: Machine Learning, CNN, Artefact Classification, Satellite Imagery, python, TensorFlow, Keras, OpenCV.

1. Introduction:

For centuries, the field of **archaeology** has been a patient, meticulous endeavor, relying heavily on the keen eyes and dedicated hands of experts to piece together the narrative of human history. From the smallest shard of pottery to the faint outlines of an ancient settlement, the process of identifying, classifying, and analyzing these **artefacts** and **sites** is fundamental. However, this traditional, manual approach presents significant challenges: it is inherently **time-consuming** and, like all human activities, can be **prone to error**. As the volume of discovered historical data grows, so too does the need for a more efficient, accurate, and scalable method of analysis.

In recent years, the convergence of history and technology—specifically the rise of **Artificial Intelligence (AI)** and **Machine Learning (ML)**—has offered a transformative pathway. This research is driven by the potential of these modern tools to radically assist and enhance the work of archaeologists. We introduce a novel machine learning-based framework designed to tackle two core, labor-intensive tasks in the domain: the **classification of artefacts** and the **preliminary analysis of satellite imagery for site detection**.

Our proposed solution leverages the power of **Convolutional Neural Networks (CNNs)**, a class of deep learning models exceptionally well-suited for image recognition tasks. CNNs are applied to categorize images of archaeological finds, such as **pottery and tools**, by learning and identifying their distinguishing visual features. To complement this, we incorporate **basic image processing techniques** to analyze aerial views from **satellite images**, helping to pinpoint subtle yet telling patterns that may indicate a potential archaeological site.

2.Existing Systems:

S.No	Existing System / Research Work	Approach Used	Advantages	Limitations
1	Traditional Manual Artefact Classification	Human expert visual inspection	Historically accurate, expert judgment	Very slow, highly subjective, limited scalability, requires years of expertise
2	Basic Feature-Based Classification (Pre-CNN Era)	SIFT,SURF, HOG features + SVM	Simple to implement	Poor accuracy for degraded/fragmented artefacts, fails to capture complex textures
3	Deep Learning for Pottery / Artefact Recognition	CNN-based architecture on labelled artefact images	High recognition accuracy, automatic feature extraction	Blind detection without contextual site information, needs large labeled datasets
4	Remote Sensing for Archaeological Site Survey	GIS +Satellite imagery interpretation	Large area coverage	Manual analysis still required, environment noise affects results, not automated
5	ML-based Site Detection (Geospatial Pattern Recognition)	Clustering, PCA, Anomaly detection on satellite images	Detects potential underground settlements	Requires specialized knowledge, lacks artefact-level confirmation
6	Multispectral & LiDAR Based Archaeology	Spectral indices (NDVI), 3D terrain scanning	Useful for buried structure detection	Very costly, requires advanced hardware, limited accessibility
7	Research-Proposed Deep Learning Frameworks (Academic Prototypes)	CNN for image classification only	Improves classification efficiency	Not integrated with satellite-based exploration, not user-friendly for archaeologists
8	Commercial Archaeology Software Tools	Proprietary scanning, mapping solutions	Professional grade analysis	Very expensive, limited ML automation.

3. Proposed System:

The proposed system introduces a fully automated and intelligent framework for archaeological artefact classification and preliminary site detection by utilizing advanced machine learning and computer vision techniques. A **Convolutional Neural Network (CNN)** with **Transfer Learning** is implemented for classifying various artefact categories such as pottery, tools, and coins. By reusing pretrained feature extraction knowledge from large-scale models, the system significantly improves classification accuracy, even when dealing with limited and fragmented archaeological image datasets. The CNN automatically learns high-level visual features including shapes, edges, and textures, enabling robust identification of damaged and weathered artefacts.

In addition to artefact analysis, the proposed system employs **OpenCV-based image processing** techniques such as edge detection, thresholding, anomaly detection, and contour extraction to analyze satellite imagery and identify potential archaeological sites. The complete solution is developed using **Python, TensorFlow, and Keras**, ensuring lightweight implementation and efficient computation suitable for real-world field deployment. By integrating both artefact classification and satellite image-based site detection into a single decision-support system, it reduces human dependency, improves interpretation speed, minimizes cost, and enhances accessibility in remote or resource-constrained excavation environments. This unified automated framework thus supports archaeologists in heritage preservation, excavation planning, and large-scale cultural documentation with improved accuracy and efficiency.

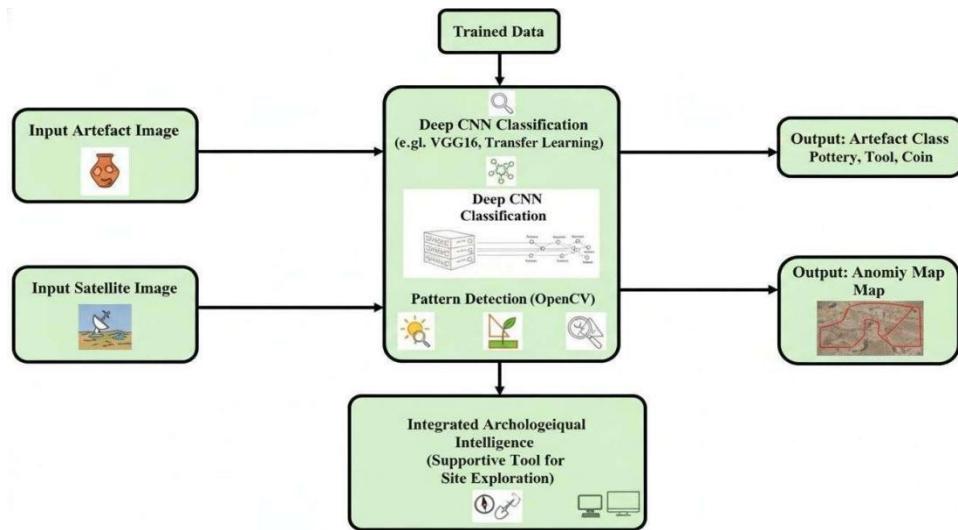


Fig : Workflow of the proposed archaeological decision-support system integrating CNN-based artefact classification with satellite imagery processing for improved detection accuracy and efficiency.

Table 1: Workflow of the implementation methodology

Step	Technique Used	Purpose
1	Artefact Image Acquisition	Collect archaeological, artefact images
2	Preprocessing (Resizing, Noise removal, Normalization)	Improve image quality and input consistency
3	CNN with Transfer Learning	Extract deep texture/shape features and classify artefact types
4	Satellite Image Acquisition	Collect satellite or aerial imagery of archaeological regions
5	OpenCV-Based Image Processing (Edge detection, Thresholding, Contour extraction)	Highlight potential archaeological sites and ground anomalies
6	Unified Decision-Support System	Integrate findings from artefact and satellite analysis
7	Output Results	Identify artefact category and probable site region

3.1 Artefact Image Acquisition

This is the initial stage where artefact images such as pottery, tools, and coins are collected from archaeological datasets or digital archives. These images serve as model input for classification.

3.2 Preprocessing of Artefact Images

The artefact images undergo operations such as resizing, background removal, noise reduction, and normalization to enhance quality.

To suppress noise, **Gaussian Smoothing** is applied:

$$G(x, y) = 1 / (2\pi\sigma^2) \cdot e^{-(x^2 + y^2) / (2\sigma^2)}$$

Purpose: Removes high-frequency noise so the CNN extracts clear structural features.

3.3 Feature Extraction and Classification – CNN with Transfer Learning

A **Convolutional Neural Network (CNN)** with transfer learning extracts hierarchical texture and shape patterns from artefact images.

The **convolution** operation detects edges and visual motifs:

$$(F * K)(i, j) = \sum_m \sum_n F(i + m, j + n) K(m, n)$$

The final classification layer uses **Softmax** to generate class probabilities:

$$\sigma(z)_i = e^{\{z_i\}} / \sum_{j=1}^c e^{\{z_j\}}$$

Training optimization uses **Categorical Cross-Entropy Loss**:

$$L = - (1 / N) \cdot \sum_{i=1}^n \sum_{l=1}^c y_{i,l} \cdot \log(\hat{y}_{i,l})$$

Purpose: CNN learns to classify artefacts such as Pottery, Tool, or Coin.

3.4 Satellite Image Acquisition

Satellite images are collected using sources like **Google Earth**, **Sentinel-2**, etc. These images contain large-scale terrain information that may indicate archaeological sites.

3.5 Satellite Image Processing OpenCV Techniques

To detect buried structures or *Edge Detection (Canny)*

Image gradients:

$$M = \sqrt{(G_x^2 + G_y^2)}$$
$$\theta = \arctan(G_y / G_x)$$

Where G_x & G_y are derivatives from Sobel filters.

Purpose: Identifies walls, pits, and soil disruption boundaries.

A) NDVI for Vegetation Stress (If multispectral data used)

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

Purpose: Highlights vegetation anomaly areas caused by buried archaeological remains.

3.6 Unified Decision-Support Framework:

The system integrates:

CNN artefact classification

Pattern detection from satellite images

Purpose: To provide stronger archaeological interpretation and improve excavation accuracy.

3.7 Output and Interpretation:

Final output includes:

Detected Artefact Class

Highlighted Potential Site Locations Visual interpretation maps

This allows faster site validation, better field planning, and preservation of Cultural heritage



Ceramic Vessels



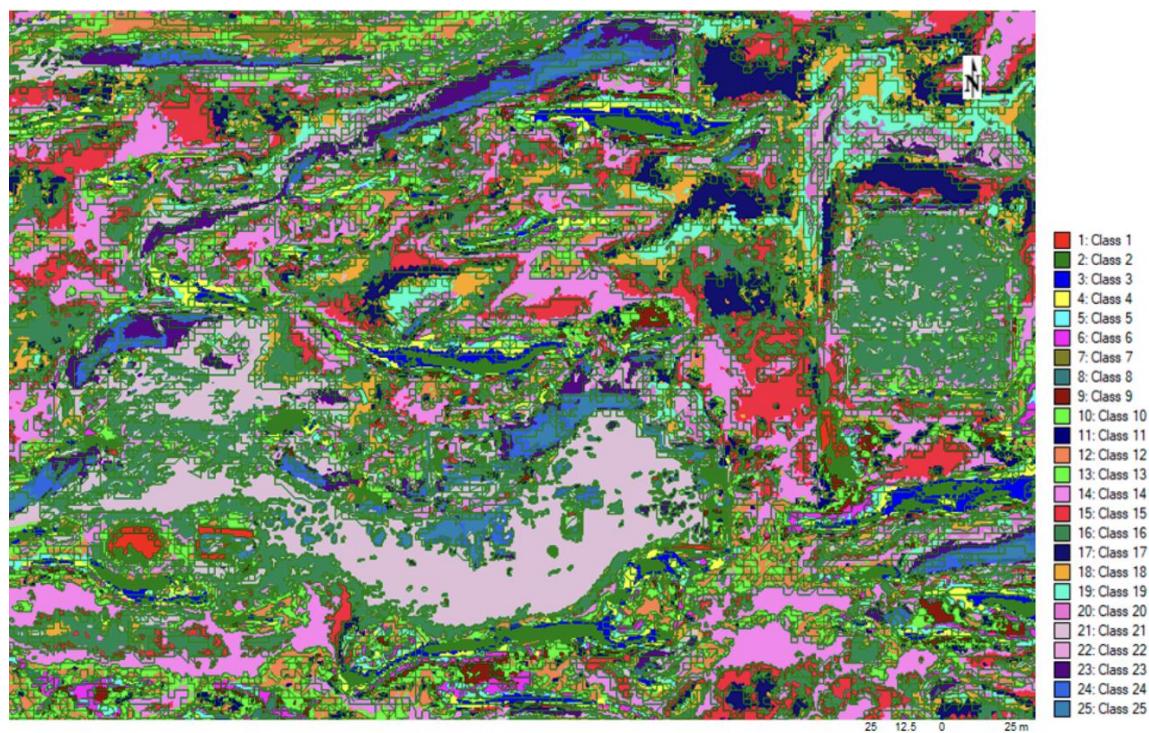
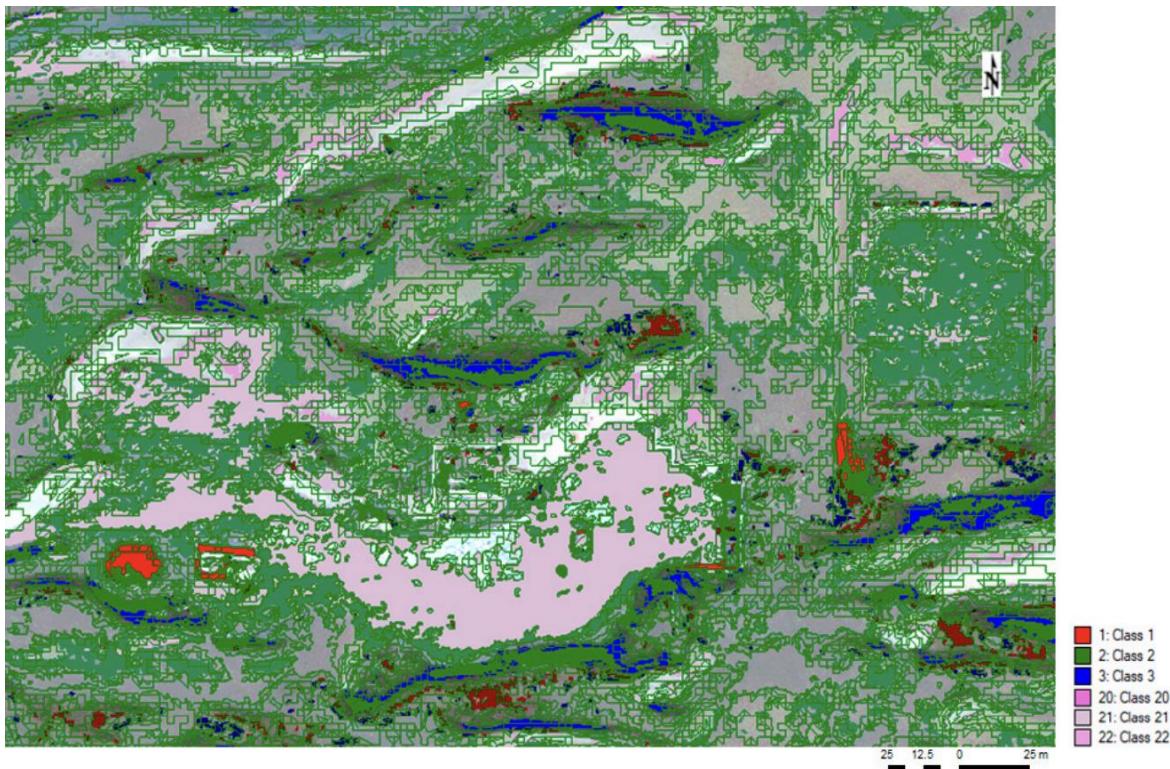
Penny/Coin



Sculpted Pieces



Barbed Tools



References:

1. Detecting and Predicting Archaeological Sites Using Remote Sensing and Machine Learning—Application to the Saruq Al-Hadid Site, Dubai, UAE (2023)
<https://www.mdpi.com/2076-3263/13/6/179>
2. A Review on Deep Learning Application for Detection of Archaeological Structures (2022)
https://www.researchgate.net/publication/359844590_A_Review_on_Deep_Learning_Application_for_Detection_of_Archaeological_Structures
3. Deep learning-based detection of qanat underground systems using satellite imagery (2024)
<https://www.sciencedirect.com/science/article/pii/S0305440324001213>
4. Convolutional neural networks for archaeological site detection – Finding princely tombs (2019)
<https://www.sciencedirect.com/science/article/abs/pii/S0305440319300846>
5. A Review of Artificial Intelligence and Remote Sensing for Archaeological Research (2022)
https://www.researchgate.net/publication/365790779_A_Review_of_Artificial_Intelligence_and_Remote_Sensing_for_Archaeological
6. The Use of Machine Learning and Satellite Imagery to Detect Roman Fortified Sites: The Case Study of Blad Talh (Tunisia Section) (2023)
https://www.researchgate.net/publication/368678366_The_Use_of_Machine_Learning_and_Satellite_Imagery_to_Detect_Roman_Fortified_Sites_The_Case_Stud
Blad Talh Tunisia Section
7. Advanced Image Processing for Archaeological Site Identification, Management, and Conservation (2024)
<https://journal.hmjournals.com/index.php/JIPIRS/article/view/4707>

Conclusion:

The proposed system introduces a reliable and intelligent deep learning-based approach for archaeological artefact classification and site detection. By using a **CNN with transfer learning**, the system can automatically learn visual features such as shapes, textures, and material patterns from artefact images, allowing accurate classification even when the objects are partially damaged or weathered. In addition, **OpenCV-based satellite image processing** helps identify meaningful land anomalies and potential excavation zones, reducing the need for extensive manual surveying.

This dual-stage design significantly enhances both the efficiency and quality of archaeological research. Since the model operates with minimal expert involvement, it serves as a practical solution in field environments where professional analysis tools are limited. The integrated output, which links artefacts to suspected site locations, provides strong decision support for archaeologists during exploration and heritage preservation. Overall, the system demonstrates a promising and scalable approach to modernizing archaeological investigation through AI driven automation.