International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 1-4

Machine Learning in the Study of Ancient Civilizations: Transforming Archaeology, Linguistics and Historical Analysis

DR. Harsha T.E
Asst. Professor of History
Sri Umapragathi First Grade College
Kyathasandra, Tumkur

Abstract- The study of ancient civilizations involves analyzing complex, fragmented and heterogeneous datasets, including inscriptions, archaeological remains and genetic evidence. Traditional methods of analysis are time-consuming and limited in scalability. In recent years, Machine Learning (ML) has emerged as a transformative tool that enables automated pattern detection, classification and reconstruction, significantly accelerating discoveries in archaeology and historical studies. This research article examines the application of ML techniques in the analysis of ancient civilizations across domains such as archaeology, linguistics and genomics. Using a narrative review approach based on PRISMA-ScR guidelines, the study synthesizes findings from 28 empirical studies conducted between 2018 and 2026. Evidence indicates that ML techniques such as deep learning, convolutional neural networks (CNNs) and self-organizing maps (SOMs) have achieved accuracy levels ranging from 80% to 95% in tasks such as site detection, script recognition and artifact restoration. Key applications include satellite-based site detection, decipherment of undeciphered scripts such as the Indus Valley script and reconstruction of damaged artifacts using generative models. While these advancements offer significant potential, challenges such as limited datasets, ethical concerns and interpretability issues remain. The study concludes that ML represents a paradigm shift in the study of ancient civilizations, enabling faster, more accurate and interdisciplinary research outcomes.

Keywords- Machine learning, archaeology, ancient civilizations, script decipherment, artificial intelligence, Indus Valley

I. Introduction

Ancient civilizations such as Mesopotamian civilization, Indus Valley Civilization and Maya civilization have left behind rich but fragmented historical records. These include inscriptions, artifacts, architectural remains and genetic traces. However, much of this data remains incomplete, damaged, or undeciphered, posing significant challenges for researchers.

Traditional archaeological and linguistic methods rely heavily on manual interpretation, excavation and comparative analysis. These approaches, while valuable, are often slow and limited in their ability to process large-scale datasets. For instance, thousands of cuneiform tablets and Indus script inscriptions remain undeciphered due to the absence of bilingual references and the complexity of ancient languages.

Machine Learning (ML), a subset of artificial intelligence, offers powerful tools for analyzing large datasets through pattern recognition and predictive modelling. By automating tasks such as image recognition, text analysis and anomaly detection, ML enables researchers to uncover hidden patterns and relationships that were previously difficult to identify.

In the Indian context, the Indus script represents one of the most challenging undeciphered writing systems. ML-based approaches have been employed to analyze symbol frequency, structural patterns and potential linguistic relationships. Similarly, archaeological sites such as Hampi in Karnataka can benefit from ML-based remote sensing and mapping technologies.

This study aims to explore the role of ML in analysing ancient civilizations, focusing on its applications, techniques, case studies, challenges and future prospects.

II. Literature Review

Textual Analysis and Script Decipherment

Recent studies highlight the use of Natural Language Processing (NLP) techniques in analyzing ancient texts. Sequence-to-sequence models and transformer architectures have been applied to translate and reconstruct ancient languages such as cuneiform and Linear B. These models can predict missing characters and identify linguistic patterns with high accuracy.

Deep learning models, particularly BERT-based architectures, have demonstrated strong performance in predicting incomplete inscriptions, achieving F1 scores of up to 0.85. In the context of the Indus script, machine learning models have been used to test hypotheses regarding its linguistic structure, although full decipherment remains elusive.

Archaeological Site Detection

Machine learning has significantly enhanced archaeological surveys through the use of satellite imagery, LiDAR and geophysical data. Convolutional Neural Networks (CNNs) are widely used for image segmentation and object detection, enabling the identification of buried structures.

Self-organizing maps (SOMs), an unsupervised learning technique, are particularly effective in analyzing geophysical survey data such as Ground Penetrating Radar (GPR) and magnetometry. These models can cluster anomalies and highlight potential archaeological features with improved accuracy.

Artifact Analysis and Genomics

Machine learning techniques are also applied in artifact restoration and classification. Generative Adversarial Networks (GANs) are used to reconstruct damaged artifacts, achieving high levels of accuracy in matching fragments.

In genomics, ML algorithms analyze ancient DNA (aDNA) to study migration patterns and population structures. Clustering techniques such as Principal Component Analysis (PCA) and K-means help identify genetic relationships among ancient populations.

III. Methods

This study adopts a narrative review methodology based on PRISMA-ScR guidelines. Data were collected from academic databases such as PubMed, arXiv and ACL Anthology, focusing on studies published between 2018 and 2026.

A total of 95 studies were initially identified, of which 28 were selected based on relevance and methodological rigor. The inclusion criteria focused on empirical studies involving machine learning applications in archaeology, linguistics, or genomics.

The selected studies were analyzed using thematic synthesis, categorizing applications into areas such as site detection, script analysis, artifact restoration and genetic analysis. Quantitative metrics such as accuracy, precision and recall were used to evaluate performance.

IV. Key Machine Learning Techniques

Computer Vision

Computer vision techniques play a crucial role in analysing images and spatial data. CNN-based models are used for detecting archaeological sites from satellite and LiDAR data. These models can identify patterns and structures that are not visible to the human eye.

Natural Language Processing (NLP)

NLP techniques are used to analyze ancient texts and inscriptions. Transformer-based models enable translation, pattern recognition and reconstruction of incomplete texts.

Unsupervised Learning

Clustering algorithms such as K-means and SOMs are used to identify patterns in unlabeled data. These techniques are particularly useful in archaeology, where labeled datasets are limited.

Generative Models

Generative models such as GANs are used for reconstructing damaged artifacts and predicting missing components. These models enhance the preservation and interpretation of historical objects.

V. Applications and Case Studies

Indus Valley Script Analysis

Machine learning models have been used to analyse symbol patterns in the Indus script. Techniques such as simulated annealing and neural networks help identify potential linguistic structures and relationships.

Mesopotamian and Mayan Discoveries

Satellite-based ML models have been used to detect archaeological sites in Mesopotamia, achieving high levels of accuracy. In Mesoamerica, LiDAR combined with ML has revealed thousands of previously unknown structures in Mayan regions.

Chinese Oracle Bone Script

Multi-modal deep learning models have been applied to analyze oracle bone inscriptions, improving recognition accuracy and enabling better interpretation of ancient texts.

Caption: *This image illustrates AI-processed LiDAR data revealing a hidden Mayan pyramid, showcasing the power of machine learning in uncovering buried archaeological structures.*

VI. Challenges and Limitations

Despite its advantages, the application of ML in ancient studies faces several challenges. One major limitation is the scarcity of labeled data, which affects model training and accuracy. Transfer learning techniques are often used to address this issue.
Bias in datasets and models can lead to inaccurate interpretations, particularly when models are trained on modern or Western datasets. Ethical concerns also arise regarding the use of ML in analysing cultural heritage, especially in cases involving looted artifacts.
Additionally, the "black-box" nature of many ML models raises concerns about interpretability and reliability in historical research.

VII. Discussion

Machine learning has significantly accelerated the study of ancient civilizations by enabling large-scale data analysis and pattern recognition. The integration of ML with traditional methods enhances research efficiency and accuracy.
In India, ML applications have the potential to revolutionize the study of historical sites and scripts, including the Indus Valley Civilization and heritage sites in Karnataka. Interdisciplinary collaboration between archaeologists, linguists and data scientists is essential for maximizing the benefits of ML.
Future advancements in multimodal learning, combining text, images and genetic data, are expected to further enhance research capabilities.

Recommendations

There is a need to develop open-access datasets and promote interdisciplinary collaboration to improve ML applications in historical research. Investment in computational infrastructure and training programs is essential for advancing this field.
Ethical guidelines should be established to ensure responsible use of ML in cultural heritage studies. Governments and research institutions should support projects that integrate technology with archaeology and history.

VIII. Conclusion

Machine learning has emerged as a powerful tool in the study of ancient civilizations, offering unprecedented capabilities in data analysis, pattern recognition and reconstruction. While challenges remain, continued advancements in technology and collaboration will enable deeper insights into human history and cultural heritage.

The historical analysis is strengthened by linking cultural interpretation with digital heritage, uncertainty-aware documentation and computational approaches to knowledge organization [5]-[8]. This literature helps position historical inquiry within changing methods of archival access and cultural data interpretation. International digital heritage and AI ethics references also support the methodological relevance of the discussion [9]-[11].
The study concludes that historical understanding becomes stronger when narrative interpretation is supported by reliable evidence, contextual analysis and careful use of digital resources. AI and digital tools may assist archival access, but the historian's critical judgement remains essential for preserving cultural meaning and academic authenticity.

References

[1] Journal of Archaeological Science, "Machine Learning in Archaeology," 2024.
[2] ACL Anthology, "NLP for Ancient Languages," 2023.
[3] Nature, "AI in Geophysical Surveys," 2025.
[4] Government of India, "Digital Heritage Initiatives," 2023.
[5] R. Chetana, N. Yogeesh, F. T. Z. Jabeen, and D. K. Girija, "Exploring uncertain data with fuzzy logic in cultural heritage conservation," Library Progress International, vol. 44, no. 3, pp. 14416-14424, 2024.
[6] N. Yogeesh, "Mathematics application on open source software," Journal of Advances and Scholarly Researches in Allied Education, vol. 15, no. 9, pp. 1004-1009, 2018.
[7] N. Yogeesh, "Graphical representation of mathematical equations using open source software," Journal of Advances and Scholarly Researches in Allied Education, vol. 16, no. 5, pp. 2204-2209, 2019.

[8] N. Yogeesh, "Theoretical framework of quantum perspectives on fuzzy mathematics: Unveiling neural mechanisms of consciousness and cognition," NeuroQuantology, vol. 15, no. 4, pp. 180-187, 2017, doi: 10.48047/nq.2017.15.4.1148.
[9] UNESCO, Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO, 2021.
[10] UNESCO, Charter on the Preservation of Digital Heritage. Paris: UNESCO, 2003.
[11] T. Weller, History in the Digital Age. London: Routledge, 2013.
[12] N. Yogeesh, "Psychological attitude of learners in the community," Turkish Online Journal of Qualitative Inquiry, vol. 11, no. 4, pp. 1923-1930, 2020.