MODERN ALGORITHMS FOR ANCIENT SCRIPTS: A REVIEW OF AI-BASED TECHNIQUES IN INDUS CIVILIZATION RESEARCH

Muzaffar Ali Khan, Syed Khaldoon Khurshid, Muhammad Aslam

Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan.

Abstract: *The Indus script has been one of the most perennial archaeological-epigraphical problems because of the lack of bilingual sources, extreme lack of data, and widespread physical destruction of the artifacts. New computational directions to overcome these limitations are provided by new developments in artificial intelligence, machine learning, and computer vision. The paper is a systematic review of AI-based methods for identifying, analyzing, and interpreting epigraphic texts and graphemes in damaged Indus Valley seal articles. After a systematic review of the literature, publications between 2009 and 2025 were thoroughly reviewed and classified into thematic groups, such as statistical and computational linguistics, deep learning-based visual analysis, allograph identification, pattern modeling, crowdsourced dataset construction and multidisciplinary methodologies. The review assesses methodological designs, data features, preprocessing procedures, model structures and reported performance measures, whereas common pitfalls include artifact distortion, small annotated data, segmentation ambiguity and computational capacity. The study indicates significant advancements in grapheme and motif detection, with ongoing shortcomings in the domain of semantic decipherment and linguistic interpretation. With comparison insights, a multi-phase analytic model is put forward that incorporates preprocessing, sign reduction, visual recognition, pattern recognition, and interpretation modeling. The paper concludes that significant developments should involve more intensive incorporation of computational practices into archaeological and linguistics to foster further research in digital epigraphy and computational archaeology.*

Keywords: *Indus Civilization; Digital Preservation; Artificial Intelligence; Deep Learning; Generative AI.*

Introduction

The Indus Valley Civilization or the Harappan Civilization was one of the earliest large-scale urban communities to be found in the fertile plains of the Indus River that flourished between 3300 and 1300 BCE (Wright, 2010). It had a vast cultural space in the geographical region of South Asia and spread along the major settlement units and probably up to the shores of the long-since-vanished Saraswati River (Petrie et al., 2019). The civilization is well known for great city design, with highly developed drainage systems, integrated baked bricks, and a very stringent weights and measures system (Wright, 2010). The excavations of different sites have shown a huge number of seals and other items carved in the nearly illegible Indus script (Shinde et al., 2018). The Indus script is one of the most significant interpretive mysteries in the archeological analysis because it has not yet been deciphered. The meaning of the world (e.g., Egyptian and Mesopotamian scripts) was deciphered because of the discovery of bilingual materials, such as the Rosetta Stone; however, the Indus script has remained undeciphered for the previous eight decades (Rao et al., 2009b). The fact that it is still undeciphered is one of the largest issues regarding the cultural and language system of the civilization (Parpola, 2009). As of now, about 80% of the Indus inscriptions are on seals (Kenoyer, 2011), (Kenoyer et al.,

2013) and the remainder are found in various objects such as the pieces of pottery, copper, stone objects, and other ornamental pieces. Most of them are broken, damaged, or incomplete due to decades of exposure to negative environmental conditions and a lack of conservation over centuries (Rao, 2010) (e.g., Figure 1 (Frenez & Vidale, 2012)). This destruction renders it difficult to interpret the inscriptions and identify individual graphemes in the right manner (Yadav & Vahia, 2011). Most of the inscriptions were very short, (averaging five signs per text), which further limits linguistic context and complicates analysis.

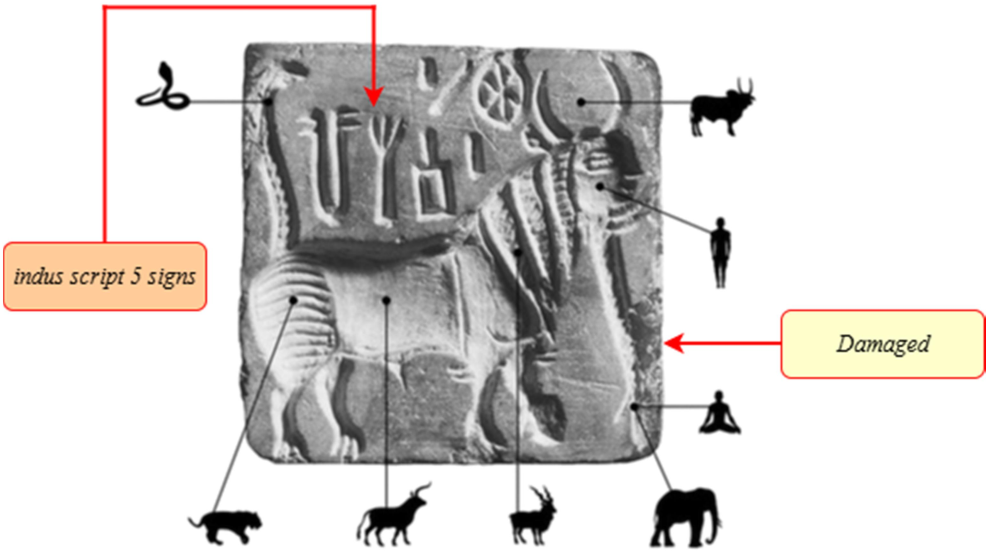

Figure 1. A five-sign inscription Indus Valley seal with clear physical degradation. Signs of preserved Indus script are highlighted, and those areas where the clarity of the grapheme is disturbed.

With the emergence of artificial intelligence and machine learning, new opportunities for dealing with these issues have appeared. In comparison with the conventional epigraphic techniques that are extremely time-consuming and require the work of specialists to conduct the analysis, AI-based methods have the ability to process vast amounts of data, detect some hidden patterns that cannot be perceived by human observers, and test a series of hypotheses in a way that is somewhat systematic and orderly. Nonetheless, the use of these technologies on ancient texts poses special issues that are considerably different from the use of OCR today, such as extreme data sparsity, lack of ground truth labels, large noise levels, and integration of domain expertise. The article presents a review of the current developments in the study of Indus scripts aimed at determining and deciphering the epigraphic text and graphemes in the scope of the damaged seal artifacts. We analyze the approaches used in the past critically, assess their performance, and say how the new technologies, in particular, machine learning and computer vision, can become a major factor in developing the field of archaeology and epigraphy research. Figure 2 provides an overview of our proposed AI-based approach to the methodology of interpretation of degraded Indus seal scripts and identification of motifs.

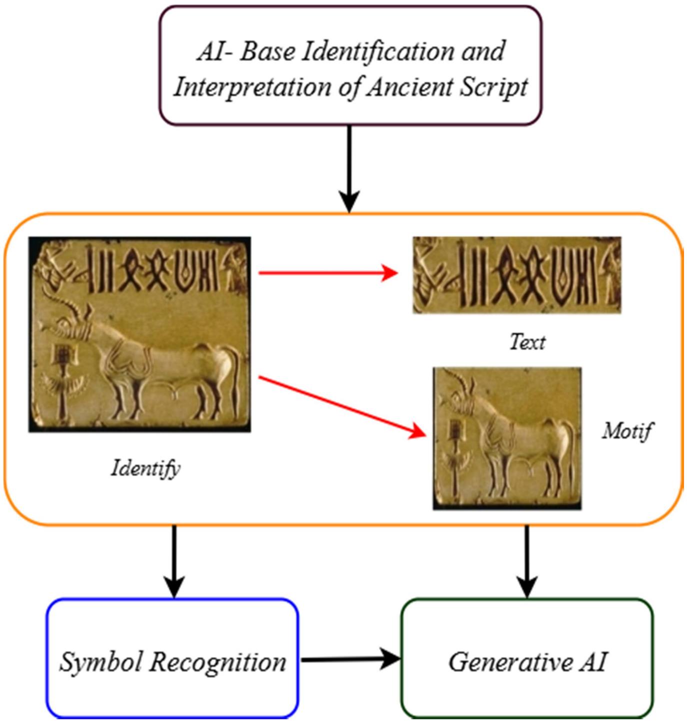

Figure 2. AI-based identification and interpretation of epigraphic scripts and graphemes. The workflow encompasses text segmentation, motif discovery, and symbol recognition, with future prospects directed towards generative AI.

Research Objectives:

The objectives of the review are to:

a. Reorganize and critically analyze the recent studies of AI usage in reading the Indus script.

b. Determine and classify technical issues relating to damaged or absent artifacts in inscriptions.

c. Determine the computer vision methods in grapheme and epigraphic text recognition.

d. Discuss the possibilities and constraints of machine learning to read, decipher, and interpret the Indus script.

e. Suggest a new combined system and new research directions of digital epigraphy in archaeology.

Literature Review

The Indus script still stands as one of the greatest archaeological and epigraphical dilemmas, mainly because of the low state of most of the remains and the lack of bilingual inscriptions to compare them. The script is mainly on seals and is usually viewed as logo-syllabic, and separate signs probably signified words or syllables as in ancient writing systems. The problem of interpreting such scripts has long been noted to be a challenge that cannot be achieved without structural analysis but also the idea of cultural and linguistic context, which is even more complicated by the fact that, due to a great deal of stylistic diversity between inscriptions, such scripts may be rather hard to interpret (Shad et al., 2024). To resolve these problems, scholars have increasingly resorted to computational methods, such as artificial intelligence, natural language processing, optical character recognition, and deep learning, to automate and improve epigraphic analysis in terms of pattern recognition, image processing, semantic modeling, and, more recently, generative AI methods.

Both statistical and linguistic evidence indicate that the Indus script has structural regularities along with the combination of logographic and syllabic features (Mirbahar, 2024). These are in line with the Dravidian hypothesis which has been empirically investigated by scholars like Asko Parpola through rebus-based interpretations. Computational analyses have also been done to analyze regional variation and sequential dependencies in the script. Venkatesh and Farghaly (2023a) used the ICIT corpus to determine the inscriptions of West Asia and the Indian subcontinent using perplexity-based models and thus found out that these two areas had distinct stylistic differences that were caused by the rare sign patterns. The same statistical model results of earlier by Rao et al. (2009a) also showed an equal positional constraint, whereby there were some signs that are often at the start or end of an inscription, indicating a syntactic structure.

The innovation of computer vision and deep learning have enhanced the automated analysis extensively. In the article, Palaniappan and Adhikari (2017) came up with ways of segmenting Indus seals into graphemic, non-textual, and hybrid parts that can be used to identify repetitive symbols in the Mahadevan corpus in a reliable way. The newer end-to-end pipelines, including ASR-Net and MIP-Net by Atturu (2024), are a combination of an object detectors and classification networks that identify graphemes and motifs even in damaged or partially visible seals. These systems use YOLOv3 to localize graphemes, MobileNet to classify, and a specific CNN to identify motifs, which is a significant step in automated OCR of ancient scripts. CNN-based methods in regions with InceptionV3 and ResNet50 architectures have also been shown to be robust to noise, as well as to degradation on epigraphic segmentation tasks (Preethi & Mamatha, 2023). The use of hybrid systems combining CNNs with classical operations of feature extractors, like SIFT with SVMs, also demonstrates the superiority of the deep learning models, with their classification rates reaching up to 98.22% when they are backed with suitable preprocessing tools of contrast-limited adaptive histogram equalization (Sindhu et al., 2023).

Besides visual recognition, allograph identification has also become an important preprocessing step in script analysis. Daggumati and Revesz (2021) used data mining and statistical testing to find unnecessary signs variants and showed that often, mirrored and asymmetric signs are variants of direction, and not variants of semantics. They used their analysis to cut the effective sign inventory to a smaller, easier to compute set of sign pairs, which made further computational modeling significantly easier. The n-gram Markov chain models have been used through complementary statistical methods to predict missing or damaged signs, and have reached prediction accuracies of about 63% on over 100 incomplete inscriptions, providing information on symbol placement and syntactic regularities (Venkatesh & Farghaly, 2023b).

To mitigate the continuous dearth of quality training data, scholars have considered other strategies of constructing data sets. Saini et al. (2022, 2024) proposed a crowdsourcing methodology (the IndusDraw platform) using which volunteers created handwritten examples of Indus symbols. Although the dataset was limited, classification models trained on this dataset obtained the level of approximately 70% accuracy, which proved that social data collection can reduce the problem of data scarcity and enhance the robustness of models to stylistic differences across regions and over time.

The recent multidisciplinary activities are additional examples of the widening of the scope of computational epigraphy. A combined framework developed by Neukart (2025), combining computer vision and machine learning, clustering, symbol frequency analysis, comparative mythology, and proto-linguistic reconstruction, was proposed to study inscriptions on Indus copper plates. Even though it is a small sample with disputable conclusions, the study offers symbolic and thematic similarities between the Indus script and other early writing systems and poses significant inquiries regarding the cultural and philosophical purposes of the Indus script. All these studies together show that although computational approaches have greatly improved the analysis of structure and visual recognition, the successful interpretation of the Indus script depends on the combination of AI procedures with archaeological background and linguistic analysis.

Table 1. Comparative summary of AI, machine learning, and statistically driven approaches for Indus script analysis

Reference	Data Type	Data Quantity	Achievement	Limitations
Rao et al. (2009a)	Indus inscriptions	Not specified	Establishes positional constraints of signs; identifies statistical regularities	Only statistically encoded; controls statistical patterns only; no semantic decoding
Palaniappan & Adhikari (2017)	Seal images; graphemic regions	Mahadevan corpus seal images	Automated segmentation of text/non-text regions; successfully identified repetitive symbols (sign 342)	Only successfully done for segmentation; deciphering not addressed; needs human-assisted annotation
Daggumati & Revesz (2021)	Sign lists	50 pairs of allographs (23 mirrored, 27 non-mirrored)	Sign reduction established; multi-directionality analysis	Multi-directionality can only be proved by validation of groupings
Saini et al. (2022, 2024)	Handwritten Indus symbols	10 symbols × 36 volunteers = 360 samples	70 percent classification; handwritten dataset created reproducibly	Old-fashioned; only 10 symbols; limited samples
Preethi & Mamatha (2023)	Ancient epigraphic inscriptions	Not specified	Noise-resistant and degradation-resistant segmentation; successful digitization	No specifications to Indus script; no deciphering capability; large training dataset needed
Sindhu et al. (2023)	Ancient inscription images	Not diverse; not tested on Indus script	98.22% accuracy in classification	Not tested on Indus script; limited dataset diversity
Venkatesh & Farghaly (2023a)	Indus inscriptions of West Asia and Indian subcontinent	ICIT corpus (varying regional datasets)	Anomaly detection identified	No meaning extraction; reliant on corpus quality

Reference	Data Type	Data Quantity	Achievement	Limitations
Ali Farghaly et al. (2023b)	Inscriptions with missing signs	100+ texts	63% missing sign prediction; understanding symbol placement; syntactic hints	Uncertainty inherent in probabilistic approach; moderate accuracy; no semantic prediction
Atturu (2024)	High-resolution annotated seal images	High-resolution annotated datasets	Effective on damaged and partial images; automated grapheme and motif recognition	Computationally expensive; visual recognition only; no decoding
Mirbahar (2024)	Theoretical linguistic framework	Not applicable	Evidence in Indus script signs; evidence in linguistic patterns	Hypothesis requires empirical testing
Shad, Egon, and Potter (2024)	Bilingual inscriptions; historical texts	Not applicable	Emphasized problems of text interpretation; importance of cultural context	No available Indus bilingual texts; abstract framework; minimal practical use
Neukart (2025)	Sequences of symbols	9 copper plate inscriptions	Proposed meanings of 9 inscriptions; parallels with ancient languages; suggested theological content	Very small sample size; interpretations controversial; lacks independent verification

Methodology

The research takes a systematic literature review technique mixed with critical and comparative analysis of the computational strategies towards deciphering Indus script. The multi-phase framework is utilized with the combination of qualitative content analysis and quantitative analysis of the artificial intelligence and machine learning methods. This study is designed based on five objectives: (1) critical reorganization of recent AI-based research, (2) taxonomy of technical issues related to damaged artifacts, (3) computer vision approaches to grapheme recognition, (4) machine learning strengths and weaknesses, and (5) development of a unified system and future study objectives.

Search Methodology

A detailed systematic search was conducted across various academic databases such as IEEE Xplore, ACM Digital Library, Google Scholar, arXiv, ScienceDirect, and Springer Link, among others. The search strategy was to use the relevant keywords; Indus script, Indus Valley Script, Harappan script together with the terms artificial intelligence, machine learning, deep learning, computer vision, optical character recognition, natural language processing, pattern recognition, and decipherment. The search temporal range was restricted mainly to the period between 2009 and 2025, with closer focus on the literature published after 2017, with the aim of getting the latest technological advances.

Inclusion and Exclusion Criteria

The selection of the studies was also performed with the predefined inclusion and exclusion criteria to guarantee relevancy and rigor of the methodology. Works that specifically promote the analysis of Indus scripts by computational means, with the help of AI, machine learning, deep learning, or computer vision, all of which are peer-reviewed journal articles, conference proceedings, or credible preprints. Moreover, the chosen studies contain enough methodological description and measurable findings, and studies on similar ancient scripts were taken into consideration in case of transferability of methodologies.

The publications were filtered out when they had no computational or technological aspect, were exclusively an archaeological or historical analysis without any AI or machine learning, were not in English but had no reliable translation, and had insufficient methodological transparency or reproducibility.

Analytical Framework

To meet the first aim, the analyzed literature was systematized into thematic groups in accordance with their major methodological orientation: statistical and computational linguistics; deep learning and computer vision; allograph identification and sign reduction; pattern recognition and Markov models; dataset construction and crowdsourcing; and multidisciplinary approaches. The studies were critically assessed based on theoretical basis, methodological rigor, reproducibility, novelty and scholarly contribution. The comparison analysis Table 1 was created to present the main characteristics, such as methodological strategy, the nature of data, the size of the dataset, documented performance, and identified limitations.

Systematic identification and classification of reported technical challenges in the literature (grouped into four categories: artifact degradation (physical damage, erosion, partial visibility), data scarcity (limited corpora, absence of bilingual texts, lack of annotations), recognition and segmentation challenges (symbol variation, overlap, and image noise), and computational constraints (scalability, resource requirements, and model generalization)) served the second objective. The categories were studied to determine prevalence, solutions offered, and effectiveness reported.

To address the third goal, computer vision-based solutions to grapheme and epigraphic text recognition were compared and contrasted on a variety of dimensions, such as the model architecture (e.g., ResNet50, InceptionV3, YOLOv3, MobileNet, and custom CNNs), preprocessing techniques (image enhancement, noise reduction, CLAHE), feature extractors (handcrafted, SIFT, and learned representations), and evaluation measures (accuracy etc.). The focus of comparative analysis was placed on robustness to noise, damage, and incomplete inscriptions, in which the transfer learning and domain adaptation were considered in particular.

The fourth goal was to evaluate machine learning methods along three dimensions of analysis, namely, reading performance (segmentation and symbol recognition), decipherment potential (pattern discovery, sequence modeling, allograph detection, and sign reconstruction), and interpretative limitations (semantic inference and linguistic structure detection). The so-called strong sides, like high classification accuracy, were opposed to the remaining constraints, like data dependency, poor semantic grounding, and limited linguistic context.

The last goal that suggested a synthesized approach to the analytical framework was obtained through the synthesis of high-performing and complementary approaches that were found in the review. The suggested framework is based on a multi-stage pipeline with a modular design that includes preprocessing, sign reduction, recognition, pattern analysis, and interpretation. It was designed based on gap analysis, modular architecture development, evidence-based method selection, optimized integration and data flow, and well-defined validation protocols to ensure robustness and scalability as well as compatibility with future extensions by generative AI-based modular extensions.

Future Research

The systematic analysis of the following questions identified the direction of the future research: (1) the countless issues that cannot be resolved in the course of numerous studies, (2) the innovations that could be offered by new technologies that have not been used to study the Indus script so far, (3) the limitation of the methods that may be addressed by innovation, (4) the opportunities available in the interdisciplinary context of archaeology, linguistics, and computer science, and (5) the possibility of their extension and generalization on an even greater scale to implement them in digital epigraphy. All the suggested directions were assessed in terms of their feasibility, possible impact, resource needs, and technological aspects of the up-to-date trends. Special care was taken to determine high-profile research areas that could deliver breakthrough findings in script decipherment or make a big step forward in the area of computational archaeology.

Conclusions and Future Work

The review has discussed how artificial intelligence, machine learning, and computer vision have come to play a role in solving the longstanding problem of Indus script analysis. By employing a systematic reorganization and the critical examination of the available literature, the paper has brought out how computational techniques have significantly enhanced the tasks of grapheme, motif recognition and segmentation, as well as the reconstruction of damaged inscriptions. Visual models built with deep learning, statistical methods of pattern recognition, and training efforts to build datasets have shown quantifiable improvements in the ability to process degraded and incomplete artifacts. Nevertheless, the analysis also shows that there are still limitations, such as the absence of a semantic background, extreme scarcity of data, bilingual reference, and limited linguistic contextualization, which still prevent complete decipherment.

The suggested hybrid analytical model combines the most effective methodologies into a pipeline of modules to match the visual recognition with the pattern analysis and interpretative modeling. This framework emphasizes the need for comparisons between disciplines, computational efficiency, and archaeological expertise, as well as linguistic theory. All in all, the paper is able to conclude that, although there is a small possibility that AI-based approaches will be able to decipher the Indus script in the near future; indeed they represent a formidable suite of tools in terms of eliminating hypotheses, rebuilding lost data, and aiding in human-centered interpretation. Further development of generative AI, transfer learning, and collaborative dataset development, in turn, will be a crucial factor in defining the future of digital epigraphy and computational archaeology.

References

Atturu, D. M. (2024). Deep learning in Indus Valley script digitization [Doctoral dissertation, Florida Institute of Technology]. Florida Tech Repository. Available at: https://repository.fit.edu/etd/1416/ (last view: 29.05.2025)

Daggumati, S., & Revesz, P. Z. (2021). Data mining ancient script image data using convolutional neural networks. In Proceedings of the 25th International Database Applications & Engineering Symposium (pp. 136-145). ACM.

Frenez, D., & Vidale, M. (2012). Harappan chimaeras as “symbolic hypertexts”: Some thoughts on Plato, chimaera, and the Indus civilization. South Asian Studies, 28(2), 107–130. DOI: https://doi.org/10.1080/02666030.2012.725578

Kenoyer, J. M. (2011). Changing perspectives of the Indus civilization: New discoveries and challenges. Purātattva: Bulletin of the Indian Archaeological Society, 41, 1-18.

Kenoyer, J. M., Price, T. D., & Burton, J. H. (2013). A new approach to tracking connections between the Indus Valley and Mesopotamia: Initial results of strontium isotope analyses from Harappa and Ur. Journal of Archaeological Science, 40(5), 2286-2297. DOI: https://doi.org/10.1016/j.jas.2012.12.040

Mirbahar, Q. (2024). Indus script: Complexities and deciphering challenges. Journal of History, Art and Archaeology, 19-27.

Neukart, F. (2025). Cracking the code of the Indus Valley civilization: A computational approach to lost knowledge. SSRN Electronic Journal. DOI: https://doi.org/10.2139/ssrn.5141753

Palaniappan, S., & Adhikari, R. (2017). Deep learning the Indus script. arXiv. Available at: https://arxiv.org/abs/1702.00523 (last view: 29.05.2025)

Parpola, A. (2009). Deciphering the Indus script. Journal of the Royal Asiatic Society, 19(1), 59-65. DOI: https://doi.org/10.1017/S1356186308008845

Petrie, C. A., Parikh, D., Smith, M. L., & Green, A. S. (2019). 'We are inheritors of a rural civilisation': Rural complexity and the ceramic economy in the Indus Civilisation in northwest India. World Archaeology, 51(2), 240-263. DOI: https://doi.org/10.1080/00438243.2019.1601463

Preethi, P., & Mamatha, H. R. (2023). Region-based convolutional neural network for segmenting text in epigraphical images. Artificial Intelligence and Applications, 119-127. DOI: https://doi.org/10.47852/bonviewAIA2202293

Rao, R. P. N. (2010). Statistical analysis of the Indus script using n-grams. PLOS ONE, 5(3), e9506. DOI: https://doi.org/10.1371/journal.pone.0009506

Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2009a). A Markov model of the Indus script. Proceedings of the National Academy of Sciences, 106(33), 13685-13690. DOI: https://doi.org/10.1073/pnas.0906237106

Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2009b). Entropic evidence for linguistic structure in the Indus script. Science, 324(5931), 1165. DOI: https://doi.org/10.1126/science.1170391

Saini, S., Shibata, H., & Takama, Y. (2022). Toward construction of handwritten Indus signs dataset. In 10th International Symposium on Computational Intelligence and Industrial Applications (ISCIIA 2022). ISCIIA.

Saini, S., Shibata, H., & Takama, Y. (2024). Construction of handwritten Indus signs dataset employing social approach. Journal of Advanced Computational Intelligence and Intelligent Informatics, 28(1), 122-128. DOI: https://doi.org/10.20965/jaciii.2024.p0122

Shad, R., Egon, A., & Potter, K. (2024). Application of artificial intelligence in deciphering ancient scripts and languages. EasyChair Preprint. Available at: https://easychair.org/publications/preprint/jNQt (last view: 29.05.2025)

Shinde, V. S., Kim, Y. J., Woo, E. J., Jadhav, N., Waghmare, P., Yadav, Y., Munshi, A., Chatterjee, M., Panyam, A., Hong, J. H., Oh, C. S., & Shin, D. H. (2018). Archaeological and anthropological studies on the Harappan cemetery of Rakhigarhi, India. PLOS ONE, 13(2), e0192299. DOI: https://doi.org/10.1371/journal.pone.0192299

Sindhu, C., Vasanth, R., & Dalara, A. (2023). Epigraphy analysis using image processing. AIP Conference Proceedings, 2782(1), Article 020045. DOI: https://doi.org/10.1063/5.0149878

Venkatesh, V., & Farghaly, A. (2023a). Deciphering the Indus script: Decoding missing and unclear Indus signs and identifying anomalous Indus texts from West Asia using Markov chain language models. In 2023 IEEE Integrated STEM Education Conference (ISEC) (pp. 233-233). IEEE. DOI: https://doi.org/10.1109/ISEC57411.2023.10402308

Venkatesh, V., & Farghaly, A. (2023b). Identifying anomalous Indus texts from West Asia using Markov chain language models. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE. DOI: https://doi.org/10.1109/ICCCNT58801.2023.10451706

Wright, R. P. (2010). The ancient Indus: Urbanism, economy, and society. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511778186

Yadav, N., & Vahia, M. N. (2011). Indus script: A study of its sign design. SCRIPTA: International Journal of Writing Systems, 3, 133-172. DOI: https://doi.org/10.1145/3472163.3472172

Muzaffar Ali Khan

Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan

ORCID https://orcid.org/0009-0000-7927-0684

muzaffar.khanji@outlook.com

*Corresponding author

Dr. Syed Khaldoon Khurshid

Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan

ORCID https://orcid.org/0000-0003-1818-9115

khaldoon@uet.edu.pk

Pro. Dr. Muhammad Aslam

Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan

ORCID https://orcid.org/0000-0002-8977-9457

maslam@uet.edu.pk

Received: 04 June 2025
Accepted: 11 December 2025
Published: 27 December 2025
DOI: https://www.doi.org/10.55630/KINJ.2025.110201

Abstract and Keywords in Bulgarian

Резюме: *Индската писменост е един от най-трайните археологическо-епиграфски проблеми поради липсата на двуезични източници, изключителната липса на данни и широко разпространеното физическо унищожаване на артефактите. Новите разработки в областта на изкуствения интелект, машинното обучение и компютърното зрение предоставят нови изчислителни насоки за преодоляване на тези ограничения. Статията представлява систематичен преглед на методи, базирани на изкуствен интелект, за идентифициране, анализ и интерпретиране на епиграфски текстове и графеми в повредени печати от долината на Инд. След систематичен преглед на литературата, публикациите, публикувани в периода между 2009 и 2025 г., бяха щателно прегледани и класифицирани в тематични групи, като например статистическа и компютърна лингвистика, визуален анализ, базиран на дълбоко обучение, идентификация на алографи, моделиране на шаблони, изграждане на набори от данни чрез краудсорсинг и мултидисциплинарни методологии. Прегледът оценява методологичните дизайни, характеристиките на данните, процедурите за предварителна обработка, структурите на моделите и отчетените показатели за ефективност, като често срещаните недостатъци включват изкривяване на артефакти, малък размер на анотираните данни, неяснота на сегментацията и изчислителен капацитет. Проучването показва значителен напредък в откриването на графеми и мотиви, като същевременно съществуват и недостатъци в областта на семантичното дешифриране и езиковата интерпретация. С помощта на сравнителни прозрения се предлага многофазен аналитичен модел, който включва предварителна обработка, редуциране на знаци, визуално разпознаване, разпознаване на образи и моделиране на интерпретацията. В статията се заключава, че значителните развития трябва да включват по-интензивно включване на компютърни практики в археологията и лингвистиката, за да се доведат до повече бъдещи изследвания в областта на дигиталната епиграфика и компютърната археология.*

Ключови думи: *Индската цивилизация; дигитално съхранение; изкуствен интелект; дълбоко обучение; генеративен изкуствен интелект.*

KIN Journal, 2025, Volume 11, Issue 2

Science Series Cultural and Historical Heritage: Preservation, Presentation, Digitalization

Научна поредица Културно-историческо наследство: опазване, представяне, дигитализация

Научная серия Культурное и историческое наследие: сохранение, презентация, оцифровка

Editors	**Редактори/съставители**
Prof. PhD. Petko St. Petkov	проф. д-р Петко Ст. Петков
Prof. PhD. Galina Bogdanova	проф. д-р Галина Богданова
Copy editors	**Технически редактори**
Assist. prof. PhD. Nikolay Noev	гл. ас. д-р Николай Ноев
Assist. prof. PhD. Mirena Todorova-Ekmekci	гл. ас. д-р Мирена Тодорова-Екмекджи
Assist. prof. PhD. Negoslav Sabev	гл. ас. д-р Негослав Събев
Assist. prof. PhD. Paskal Piperkov	гл. ас. д-р Паскал Пиперков
© Editors, Authors of Papers, 2025	**© Редколегия, Авторски колектив, 2025**
Published by	**Издание на**
Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, Sofia, Bulgaria	Институт по математика и информатика при Българска академия на науките, София, България

http://www.math.bas.bg/vt/kin/

ISSN: 2367-8038