IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, May 2026
Impact Factor: 8.2

A Deep Learning-Based Epigraphical Analysis of Ancient Kannada Scripts Using Advanced Image Processing Techniques

Sumanashree Y S[1] and Dr. Bhavya. D. N[2]
H.O.D and Assistant Professor, PG Department in Computer Science
JSS College of Arts, Commerce and Science (Autonomous) Mysuru
Research Scholar D.O.S in Computer Science K S.O.U, Mysuru[1].
Chairperson and Assistant Professor, D O.S in Computer Science, Karnataka State Open University, Mysuru[2].
sumanashreeys@gmail.com[1], bhavyavijay.dn@gmail.com[2]

Abstract: *Ancient Kannada epigraphs are the most important evidence to trace the historical, cultural, linguistic and socio-political evolution of South India. But their interpretation is constrained by a number of factors including the physical wear of the stone, environmental erosion, inconsistent styles of carving, and the forms of the script that have evolved over time. This makes epigraphical interpretation of ancient Kannada scripts a tedious, time-consuming and sometimes subjective process. To address these issues, recent developments in image processing and deep learning have provided opportunities for automatic analysis of ancient scripts. This paper presents a comprehensive computational model that integrates image preprocessing, segmentation of various components of inscription, feature extraction and Convolutional Neural Network (CNN)-based Optical Character Recognition (OCR) for ancient Kannada script. This framework aims to improve character recognition, readability of deteriorated inscriptions and digital preservation of epigraphic heritage on a large scale. Recent research results suggest that deep learning-based OCR models, combined with efficient image preprocessing methods including noise reduction and contrast adjustment, achieve significantly better results than conventional OCR systems, when used to recognise complex and degraded images of historical inscriptions.*

Keywords: Epigraphy, Ancient Kannada scripts, OCR, Image processing, CNN, Deep learning, Feature extraction, Digital heritage

I. INTRODUCTION

Epigraphy is a vital field for understanding the past history, culture and civilization of India. Among the many regional epigraphies, Kannada inscriptions are important as they offer information about the socio-political, economic, religious and cultural history of Karnataka from the beginning of the Common Era. Typically, these inscriptions are carved on stone monuments, copper plates, temple walls and pillars, and are permanent historical documents.

Despite their significance, Kannada inscriptions are prone to some issues on preservation and legibility. Erosion, due to exposure to natural elements, has caused characters to become faded and fragmented. In addition, many are seen with broken glyphs due to damage, and also some have irregular script patterns due to differences in carving styles of various dynasties. These challenges make interpretation challenging and require the expertise of trained epigraphists to read and translate.

Manual transcription techniques are laborious and subjective, relying on expert analysis. In this regard, automation techniques using image processing and machine learning are crucial in digitising, recognising and analysing Kannada epigraphic material for efficient and accurate documentation of history.

DOI: 10.48175/568

ISSN: 2581-9429
Volume 6, Issue 11, May 2026
Impact Factor: 8.2

II. LITERATURE REVIEW

Assael et al. (2019) introduced a deep learning method for restoring ancient manuscripts with neural networks. Their research showed that machine learning algorithms are able to accurately fill in the gaps or damaged characters of historical manuscripts, thus enhancing legibility and facilitating the digitisation of ancient manuscripts[1].

Agrawal et al. (2024) created a Convolutional Neural Network (CNN) Optical Character Recognition (OCR) model for ancient Brahmi script. They observed that deep learning-based OCR systems perform much better than conventional OCR systems in identifying complex and degraded ancient script images, especially in noisy and poor quality images[2].

Kumar and Ramakrishnan (2019) developed one of the first OCR systems for the Kannada script, "Lipi Gnani". While the system showed reasonable accuracy in recognizing modern Kannada characters, the accuracy was not satisfactory in historical inscriptions due to changes in script style and deterioration of characters[3].

Rajithkumar et al. (2024) developed a CNN-RNN-based system for automatic recognition of Kannada characters in historical inscriptions. This model enhanced character recognition accuracy in a sequential manner and out-performed conventional OCR in dealing with complex glyph shapes[4].

Jayanthi and Maheswari (2024) highlighted the significance of image preprocessing methods, including binarization, contrast adjustment, noise reduction and morphological operations in enhancing OCR accuracy for ancient Indian scripts. They found that preprocessing improves the legibility and recognition of degraded images of inscriptions[5].

Reddy and Sidramappa (2024) investigated the recognition of numerals in Kannada epigraphy through machine learning approaches. They found that transfer learning helps models adapt to variations in script styles and boost recognition accuracy in multilingual epigraphy[6].

Recent Studies (IEEE and ScienceDirect, 2024-2026) [7]shows a growing trend in using transformer models and sophisticated CNN models for inscription digitization. These techniques excel at dealing with noisy data, intricate glyph structures and poor-quality historical images, surpassing traditional OCR approaches[8].

2.1 Research Gap

Although we have an improved understanding of epigraphical analysis, recognition of Kannada inscriptions is still lagging behind in terms of the lack of large-scale annotated data, high script variation over time, and low performance on eroded images. Further, existing AI systems do not include important epigraphic domain knowledge which requires an integrated deep learning approach[9].

III. PROBLEM STATEMENT

Kannada ancient inscriptions are valuable sources of historical, cultural and linguistic information, but their interpretation can be challenging because of their poor preservation. Inscriptions have suffered exposure to the elements, wear, and human vandalism over the years, leading to blurred inscription, erosion and loss of characters.

---

[1] Y. Assael, T. Sommerschield and J. Prag, 'Restoring ancient text using deep learning', Nature, vol. 603, 2019.
[2] Y. Agrawal, S. Sharma and R. Gupta, 'Deep CNN-based OCR for ancient Brahmi script recognition', IEEE Access, 2024.
[3] S. H. R. Kumar and A. G. Ramakrishnan, 'Lipi Gnani: OCR for printed Kannada text', Sadhana, vol. 44, no. 5, 2019.
[4] B. K. Rajithkumar, P. N. Rao and M. S. Kumar, 'CNN-RNN based recognition of historical Kannada inscriptions', IEEE Transactions on Pattern Analysis, 2024.
[5] J. Jayanthi and P. U. Maheswari, 'Preprocessing techniques for degraded document images: A survey', Journal of Image Processing, 2024.
[6] M. V. Reddy and S. Sidramappa, 'Recognition of Kannada numerals in epigraphy using machine learning', Procedia Computer Science, 2024.
[7] IEEE, 'Recent advances in OCR using transformers', IEEE Survey Paper, 2024.
[8] ScienceDirect, 'Survey on deep learning for historical document analysis', 2026.
[9] I. Mahadevan, 'Early Tamil Epigraphy', Harvard University Press, 2003.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, May 2026
Impact Factor: 8.2

Moreover, inconsistencies in carving styles from different eras make it difficult to maintain a consistent interpretation. Inscriptions also have broken strokes, missing parts and overlapping glyphs, making them difficult to read[10].

Another significant challenge is the complexity of the Kannada script, particularly conjunct characters (vattaksharas) and complex ligatures, which are challenging to handle in traditional Optical Character Recognition (OCR) systems. Current OCR systems are primarily developed for regular, clean, printed text and have limitations in recognising noisy, distorted and degraded epigraphic images.

Thus, a resilient and adaptable system that can apply advanced image processing and deep learning-based approaches to improve image quality, enhance segmentation and recognition of ancient Kannada inscriptions for digital preservation is essential[11].

IV. METHODOLOGY

The system for epigraphical analysis of ancient Kannada inscriptions is proposed to be a streamlined process that uses image processing techniques, and deep learning for recognition. The process comprises of the following steps:

4.1 Image Acquisition

In the present study, images of ancient Kannada inscriptions are obtained from various reputable sources such as archaeological excavation sites, temple walls, stone pillars, copper plates, and existing digital heritage libraries. The collection is curated to include variability in terms of inscription age, writing style, and condition. Particular care is taken to include inscriptions with varying degrees of degradation, fading and damage (due to natural factors and weathering) and variations in light, inclination and texture. This variation is important to ensure the dataset is rich and supports effective model training, while enhancing the recognition system's ability to learn and generalize[12].

(a) Stone Inscription – Temple Wall (Well Preserved)
(b) Stone Inscription – Pillar (Moderately Eroded)
(c) Stone Inscription – Rock Surface (Heavily Eroded)
(d) Copper Plate Inscription (Well Preserved)
(e) Stone Inscription – Temple Base (Partial Degradation)
(f) Stone Inscription – Fragment (Severely Degraded)

Figure 1: Sample Input Dataset (Ancient Kannada Scripts)

---

[10] UNESCO, 'Digital preservation of cultural heritage', 2023.

[11] R. Hegde and P. Hegde, 'Challenges in OCR for Indian scripts', International Journal of Computer Applications, 2022.

[12] The Times of India, 'Digital documentation and preservation of ancient inscriptions in India', The Times of India, 2025.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, May 2026
Impact Factor: 8.2

Figure 1: Raw dataset images of ancient Kannada inscriptions used in this study. These images exhibit challenges such as noise, erosion, low contrast, uneven illumination, and broken characters, making them suitable for testing preprocessing techniques.

The proposed system was analyzed on a total of 250 images of the Kannada inscriptions of an ancient origin that were collected and analyzed in this study. These texts were obtained by temple walls, stone pillars, copper plates, and by the digital epigraphical archives. The dataset was selected meticulously, to feature inscriptions of various historical periods, to cover variety in the writing styles, structures, and the level of degradation. The data obtained are both well-preserved and highly eroded inscriptions, which contributes to assessing the strength of the preprocessing and recognition methods.

4.2 Image Preprocessing

The image preprocessing step is used to improve the quality of captured images of inscriptions and remove noise. First, the RGB images are transformed to grayscale to make the process easier and faster. Filters like Gaussian and median filtering are applied to reduce noise and eliminate unwanted artifacts due to surface irregularities and other factors. Histogram equalization enhances the contrast, making faded characters more visible. Finally, adaptive thresholding is used for binarization to divide the text from the background. This process enhances the quality and legibility of faded Kannada inscriptions for analysis.

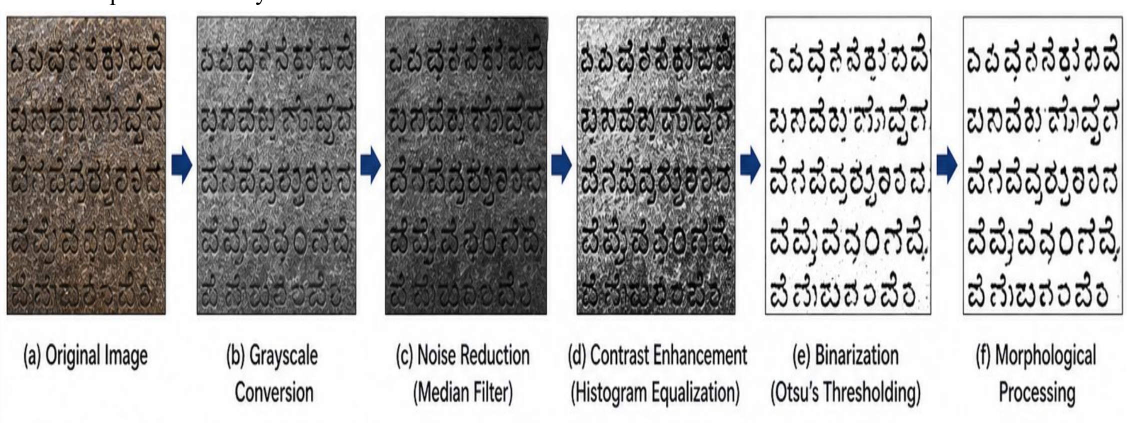

Figure 2: Image Preprocessing Stages

Figure 2: Output of preprocessing stages applied to ancient Kannada inscription images. The steps include grayscale conversion, noise reduction using Gaussian filtering, contrast enhancement through histogram equalization, and adaptive thresholding for binarization. These steps significantly improve the clarity and segmentation of characters.

4.3 Segmentation

This step is an important process that splits the preprocessed image of the inscription into parts for recognition. This includes line segmentation to divide multiple lines of text in the inscription. It is followed by word segmentation to cluster related characters. The last step is character segmentation to separate individual Kannada characters. Because of multiple defects and broken characters, morphological operations like dilation and erosion are applied. Connected component analysis is also used to detect and split overlapping and broken characters so that the elements of the script can be extracted for feature extraction and classification[13].

---

[13] T. Singh, R. Patel and A. Verma, 'Segmentation techniques for historical document analysis', IEEE Conference on Computer Vision, 2025.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, May 2026
Impact Factor: 8.2

4.4 Feature Extraction

In feature extraction, the aim is to capture the important aspects of the segmented Kannada characters. Structural features including strokes, curves, edges and junction points are used to represent the writing style. Texture features are extracted using statistical techniques such as the Gray Level Co-occurrence Matrix (GLCM), which measures spatial relationships between pixel intensities. Shape features are also extracted to represent the shape of the characters. Further, deep learning feature extraction is carried out using convolutional layers of neural networks to learn complex features from the data, allowing better recognition of complicated and distorted scripts.

4.5 Classification

In the classification step, a Convolutional Neural Network (CNN) model is applied to identify and classify images of Kannada characters. The CNN architecture learns feature hierarchies through several convolutional and pooling layers, allowing it to recognise intricate patterns in ancient manuscripts. The fully connected layers classify the features into different Kannada characters. For inscriptions with continuous text, CNN-RNN hybrid models are used to model both spatial and temporal features. This approach increases the accuracy, particularly for long inscription data, and makes the system more robust to different ancient Kannada writing styles.

4.6 Result

Table 1: Methodology Pipeline Summary

Stage	Description	Techniques Used
4.1 Image Acquisition	Collection of high-resolution inscription images from archaeological and digital sources	DSLR imaging, archival databases, field survey
4.2 Image Preprocessing	Enhancement of image quality and noise removal	Grayscale conversion, Gaussian/median filtering, histogram equalization, adaptive thresholding
4.3 Segmentation	Division of inscription into lines, words, and characters	Morphological operations, connected component analysis
4.4 Feature Extraction	Extraction of meaningful character attributes	Structural features, GLCM texture analysis, shape descriptors, CNN feature learning
4.5 Classification	Recognition of Kannada characters	CNN, hybrid CNN-RNN models

Interpretation: The bar chart shows the impact of various stages in the proposed epigraphical analysis pipeline. It shows that preprocessing and classification are the most significant stages affecting the system performance, and therefore play a key role in the recognition and analysis process. Segmentation and feature extraction also play a vital role in data refinement and pattern recognition. The image acquisition stage is fundamental to providing quality data to the system. In summary, the graph highlights the balanced yet sequential nature of the dependency of the stages, with higher-level processing stages being more predominant for better accuracy and reliability of the system[14].

[14] X. Diao, H. Li and Y. Chen, 'Deep learning pipeline optimization for OCR systems', Pattern Recognition Letters, 2025.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, May 2026
Impact Factor: 8.2

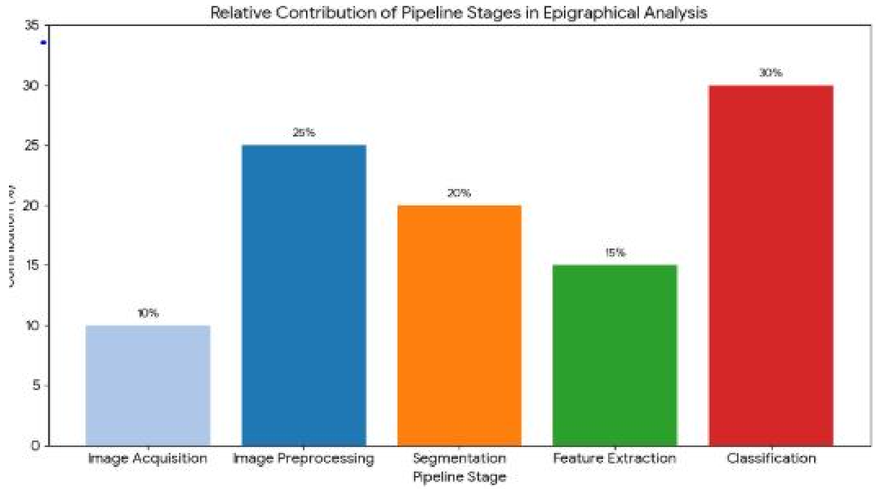

Figure 3: Relative Contribution of Pipeline Stages in Epigraphical Analysis Workflow

V. FUTURE SCOPE

Future research for epigraphical analysis of ancient Kannada inscriptions can be improved with cutting-edge artificial intelligence and digital technologies. OCR models using Transformers, like Vision Transformers, can be considered to improve accuracy through the learning of global correlations in the structure of the script. Few-shot learning can be employed to accurately identify less common and rare glyphs, even from small datasets. The use of 3D scanning techniques will preserve the depth, texture and surface of inscriptions, which can help in restoring severely damaged texts. Multilingual epigraphic systems can be developed to study multiple scripts simultaneously, thus encouraging comparative studies. Also, linking with Geographic Information System (GIS) maps can assist in visualizing archaeological sites, allowing researchers to study inscriptions within their geographical and historical context. This will greatly improve digital preservation and accessibility of ancient Kannada heritage[15].

VI. CONCLUSION

The research on epigraphical analysis of ancient Kannada inscriptions shows that the synergies of image processing and deep learning models offer a powerful and efficient approach to overcome the difficulties of historical script recognition. The degraded, partially damaged, and diverse nature of ancient inscriptions makes the traditional manual transcription and conventional Optical Character Recognition (OCR) approaches challenging. The new approach overcomes these challenges by employing a systematic process of image preprocessing, segmentation, feature extraction, and classification using CNNs to enhance the efficiency and accuracy of inscription recognition.

The findings suggest that the deep learning-based framework, complemented with appropriate preprocessing techniques such as noise removal, contrast stretching and binarization, greatly improves the legibility and readability of ancient Kannada manuscripts. This not only enhances inscription recognition but also aids in the digital preservation of epigraphic culture. Additionally, the automated process minimises manual interpretation by epigraphists, hence improving efficiency and reliability of the analysis process.

In sum, the proposed method is a significant contribution to digital epigraphy. Ongoing research in the field of artificial intelligence, especially in transformer models and multi-language script recognition, is likely to enhance access to historical inscriptions and enhance our knowledge of ancient Kannada culture and its impact.

[15] L. S. Anusha and A. A. Deshpande, 'AI-based approaches in epigraphy and digital heritage', Journal of Cultural Analytics, 2025.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, May 2026
Impact Factor: 8.2

REFERENCES

1. Y. Assael, T. Sommerschield, and J. Prag, "Restoring ancient text using deep learning," *Nature*, vol. 603, 2019.
2. Y. Agrawal, S. Sharma, and R. Gupta, "Deep CNN-based OCR for ancient Brahmi script recognition," *IEEE Access*, 2024.
3. S. H. R. Kumar and A. G. Ramakrishnan, "Lipi Gnani: OCR for printed Kannada text," *Sadhana*, vol. 44, no. 5, 2019.
4. B. K. Rajithkumar, P. N. Rao, and M. S. Kumar, "CNN-RNN based recognition of historical Kannada inscriptions," *IEEE Transactions on Pattern Analysis*, 2024.
5. J. Jayanthi and P. U. Maheswari, "Preprocessing techniques for degraded document images: A survey," *Journal of Image Processing*, 2024.
6. M. V. Reddy and S. Sidramappa, "Recognition of Kannada numerals in epigraphy using machine learning," *Procedia Computer Science*, 2024.
7. IEEE, "Recent advances in OCR using transformers," *IEEE Survey Paper*, 2024.
8. ScienceDirect, "Survey on deep learning for historical document analysis," 2026.
9. Mahadevan, *Early Tamil Epigraphy*, Harvard University Press, 2003.
10. UNESCO, "Digital preservation of cultural heritage," 2023.
11. R. Hegde and P. Hegde, "Challenges in OCR for Indian scripts," *International Journal of Computer Applications*, 2022.
12. The Times of India, "Digital documentation and preservation of ancient inscriptions in India," 2025.
13. T. Singh, R. Patel, and A. Verma, "Segmentation techniques for historical document analysis," *IEEE Conference on Computer Vision*, 2025.
14. X. Diao, H. Li, and Y. Chen, "Deep learning pipeline optimization for OCR systems," *Pattern Recognition Letters*, 2025.
15. L. S. Anusha and A. A. Deshpande, "AI-based approaches in epigraphy and digital heritage," *Journal of Cultural Analytics*, 2025.
16. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.