MANTECH Publications

Journal of AI Ethics and Autonomous Systems
Volume 2, Issue 2, July-December, 2025
ISSN: 3107-8095 (Online)

Time Lens: AI Powered Monument Recognition and Exploration

Prem Javali[1], Ananya J[2], Samrudhi Karadi[3], Ashish Betageri[4], Pradnya Malaganve[5]

Student[1, 2, 3 ,4], Associate Professor[5]

Department of Computer Science and Engineering

Jain College of Engineering and Research, Belagavi

Corresponding Authors' Email: *javaliprem123@gmail.com*

ABSTRACT

India's architectural heritage spans thousands of years and includes temples, forts, palaces, and historic structures that reflect its cultural evolution. However, many users—tourists, students, and even researchers—struggle to identify monuments or access reliable historical information from photographs taken in real-world conditions. To address this gap, TimeLens introduces an AI-powered monument recognition system designed to make heritage exploration simple, accurate, and accessible. The system allows users to upload an image of any monument, after which a multimodal vision-language model analyzes architectural style, structural elements, textures, and contextual cues to identify the monument with high confidence. When the visual input is unclear, a fallback text-based reasoning mechanism ensures robust identification. Once recognized, the system generates a structured historical narrative covering architectural features, cultural significance, and related dynasties. This content is produced in English and automatically translated into Kannada, accompanied by natural audio narration to support diverse users, including those with reading difficulties. By combining computer vision, intelligent text generation, and multilingual accessibility, TimeLens offers a modern and user-friendly approach to cultural interpretation. The project demonstrates how AI can enhance heritage preservation by making historical knowledge more approachable, interactive, and engaging for the public.

KEYWORDS: *Monument Recognition, Multimodal AI, Cultural Heritage, Historical Information Retrieval, Vision-Language Model, Multilingual Narration.*

INTRODUCTION

India's architectural heritage stands as a timeless record of its cultural, religious, and artistic evolution. From ancient temples and forts to palaces and archaeological structures, each monument carries unique historical importance. Yet, identifying these monuments and accessing accurate information remains a challenge for many users. Tourists often rely on incomplete online sources, students struggle to find authenticated historical content, and casual visitors may not know the significance of the structures they encounter. Traditional methods of learning about monuments depend heavily on guidebooks, expert interpretation, or specialized knowledge, which are not always accessible or reliable. To address these limitations, Time Lens introduces an AI-driven solution that simplifies heritage understanding through automated monument recognition. By allowing users to upload a monument image, the system uses a multimodal vision- language model to analyze architectural patterns, structural elements, and contextual cues to identify the monument accurately. When visual clarity is low, a fallback text- based reasoning mechanism ensures dependable results. Once the monument is identified, the system generates a detailed historical explanation covering architectural style, cultural relevance, associated dynasties, and key events. This content is provided in bilingual format—English and Kannada—along with audio narration to make the information inclusive and easier to consume.

Background

India's monumental architecture reflects one of the world's oldest and most diverse cultural histories. Temples, stepwells, mosques, palaces, forts, and archaeological structures across the country showcase centuries of craftsmanship and regional artistic traditions. However, despite their significance, many monuments remain poorly documented or difficult for the general public to identify. The availability of reliable historical information is often scattered across books, tourism portals, or academic archives, making it challenging for visitors and learners to access accurate details instantly. Moreover, photos taken in real-world environments such as crowded surroundings, low lighting, or partial views often complicate manual identification. Advancements in artificial intelligence, particularly in computer vision and multimodal learning, provide an opportunity to bridge this gap. Modern vision-language models can interpret architectural styles, textures, shapes, and contextual cues directly from an image, enabling automated recognition of historical structures. Existing applications typically rely on fixed datasets or simple image-matching techniques, which struggle when

the monument's angle, lighting, or visibility varies. Recent progress in deep-learning architectures, including transformer-based vision models, has significantly improved image understanding and contextual reasoning capabilities. These models can analyze nuanced features like pillar carvings, dome styles, gopuram patterns, and façade geometry elements that are essential for identifying Indian monuments. Similarly, large language models can generate descriptive, context- rich narratives once the monument is recognized. Time Lens leverages these advancements by integrating image-based recognition, fallback reasoning, multilingual translation, and audio narration into a single accessible platform. This approach supports tourists, students, educators, and heritage enthusiasts, making India's rich architectural legacy more discoverable and engaging in a digital-first world.

LITERATURE REVIEW

The preservation and digitization of ancient manuscripts have been widely explored inrecent studies. Several authors emphasize the need for automated systems capable of handling degraded scripts, low-quality images, and linguistic diversity found in historical documents [1], [2]. Several OCR- based approaches have been introduced, with Tesseract OCR and CNN-enhanced models improving recognition accuracy for both printed and handwritten scripts [3]. Deep learning based preprocessing methods, such as noise reduction and character enhancement, further support the extraction of text from damaged or blurred sources [4].

Researchers have also examined multilingual translation systems using transformer models, BERT variants, and regional language embedding for converting extracted text into languages like English, Hindi, and Kannada [5]. Related studies on monument identification use CNN architectures to automate image-based recognition and retrieve cultural or historical information [6], [7]. Despite these advancements, existing systems still perform poorly on extremely degraded stone inscriptions and palm-leaf manuscripts [8].

Moreover, current literature lacks an integrated workflow that combines OCR, translation, and monument-based historical context generation within a single unified pipeline. Therefore, this study aims to address these gaps by proposing an AI- driven framework capable of text extraction, multilingual translation, and automated historical information generation to support cultural preservation.

The Problems

The investigation revealed several issues in the current system, including the following:

- Historical information remains scattered across unreliable sources, preventing quick access for users today.

- Most platforms lack multilingual support and accessibility features, limiting inclusivity for diverse audiences.

- No unified system combines recognition, reasoning, translation, and narration into one seamless workflow.

- Architecturally similar monuments confuse conventional systems, leading to frequent misidentification and errors overall.

Methodology

The Time Lens Monument Recognition system follows a structured, multi-stage workflow designed to ensure accurate identification, robust reasoning, and accessible historical narration. The complete methodology consists of image acquisition, preprocessing, AI-based analysis, fallback reasoning, post-processing, and multilingual output generation. Therefore, monument recognition system designed to make heritage exploration simple, accurate, and accessible.

Figure 1, shows the overall architecture of the monument recognition model which includes image input and validation, vision based monument analysis, fallback text reasoning, historical information generation and translation and audio narration modules.

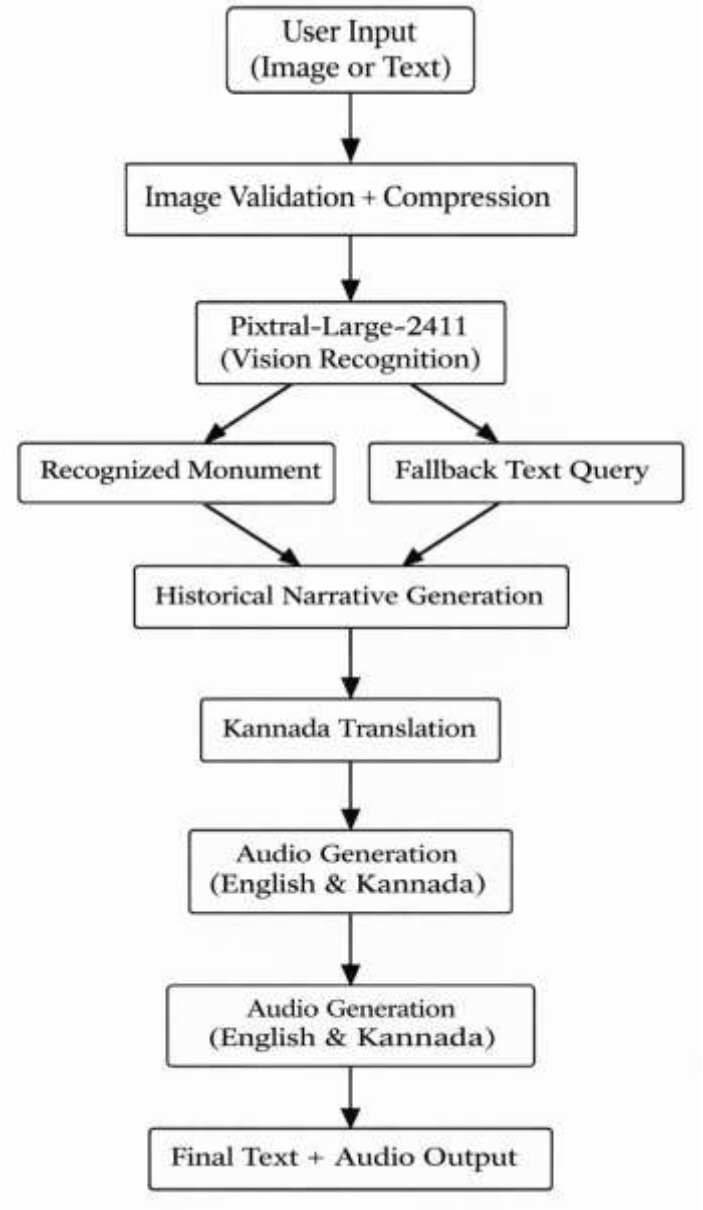

Figure 1: System architecture

1. **Image Input and Validation:** Users begin by uploading a monument image or providing the monument name as text. The system validates the input by checking format, resolution, and size. Oversized images are compressed using Pillow to maintain performance without losing essential visual details.

2. **Vision-Based Monument Analysis:** The validated image is processed by the Pixtral-Large-2411 multimodal vision model. This model analyzes architectural elements such as domes, pillars, carvings, arches, gopurams, façade geometry, and material texture. It uses these cues to predict the most probable monument with high confidence.

3. **Fallback Text Reasoning Module:** If the vision model is uncertain due to obstructions, low lighting, or partial visibility, the system activates a fallback reasoning step. Here, the model uses user-provided text or inferred visual hints to re-evaluate and refine the prediction, ensuring reliable monument identification.

4. **Historical Information Generation:** After confirming the monument, a specialized Mistral Agent generates a structured historical narrative. The output includes architectural style, cultural significance, associated dynasty, construction period, and notable historical events.

5. **Translation and Audio Narration:** The generated English content is translated into Kannada using automated translation tools. Both versions undergo validation for clarity. The system then uses gTTS to produce natural audio narration in English and Kannada, making the information accessible to a wider audience.

RESULT

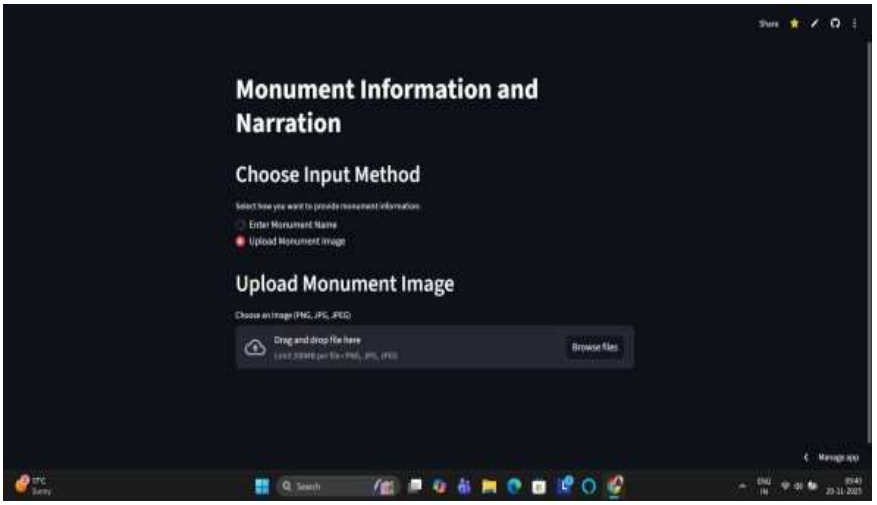

Figure 2: Monument Recognition Module

Figure 2, shows that the user is given flexibility to add image or the text i.e name of the monument as input to the module.

Figure 3: Uploaded Image

Figure 3, shows the image that is uploaded in Figure 2 is displayed to the user.

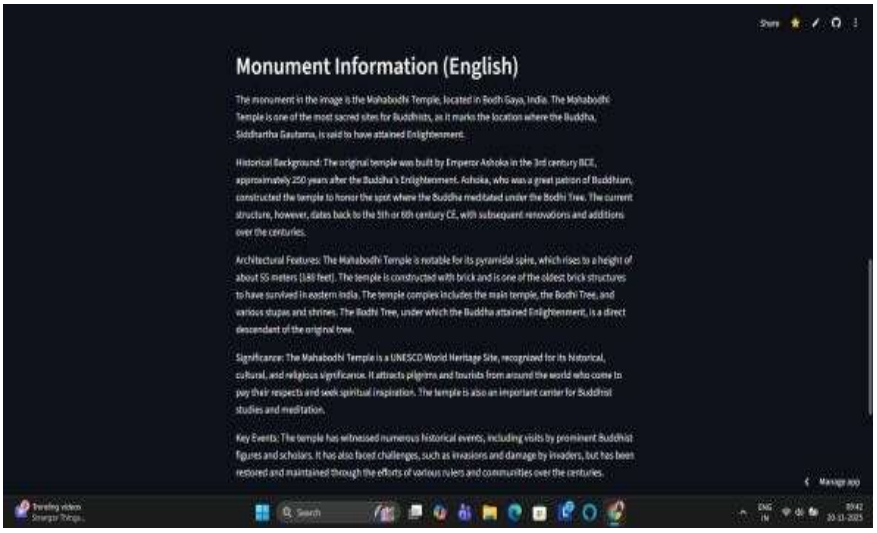

Figure 4: Monument History in English

Once the image or the text is accepted and is processed by the model the history of the monument is generated in English, which is shown in Figure 4.

Figure 5: Monument history audio narration and Kannada translation

Figure 5, shows that the information generated in Figure 3 is audio narrated to the users which, increases the flexibility and also the figure shows the information generated in English is translated into Kannada and also the audio narration of it is provided to the users in Kannada.

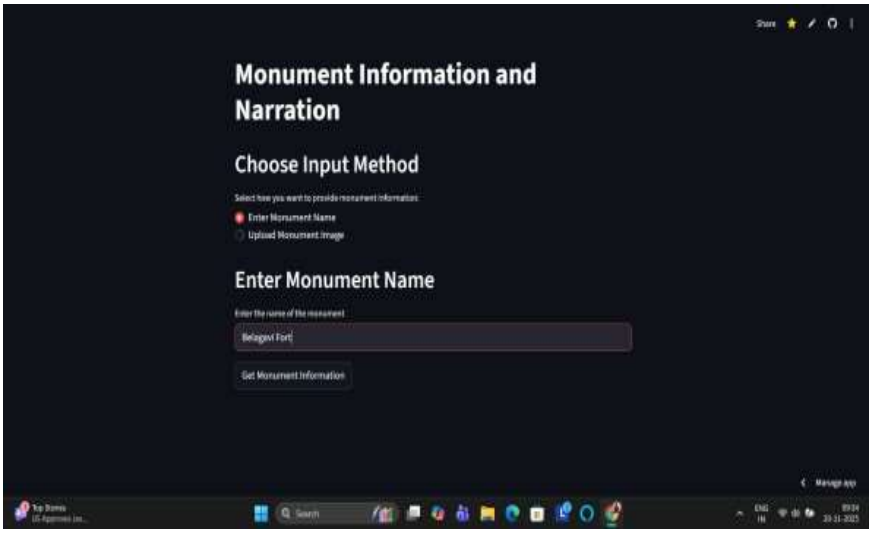

Figure 6: Text input

Figure 6, shows the text input where the user gives monument name as an input and the remaining procedure remains same as the above which includes recognizing the model and history narration in Kannada and English along with audio narration.

CONCLUSION

The Time Lens Monument Recognition system demonstrates how modern artificial intelligence can meaningfully enhance cultural exploration by making historical knowledge more accessible, reliable, and engaging. India's monumental heritage spans countless architectural styles, dynasties, and cultural influences, yet identifying these structures and understanding their significance has traditionally required specialized knowledge or fragmented online research. By integrating multimodal computer vision, intelligent text reasoning, and multilingual communication into a single unified pipeline, Time Lens provides a practical and user-friendly solution to this challenge. The system's strength lies in its ability to analyze real-world monument images, even when captured from imperfect angles or under varying conditions. When visual clarity is insufficient, the fallback reasoning mechanism adds an extra layer of reliability, ensuring consistent identification. Once a monument is recognized, the generated historical narrative offers users accurate and contextual information, helping them appreciate the architectural style, cultural background, and historical importance of each structure. The bilingual output, paired with natural audio narration, further broadens accessibility, supporting tourists, students, educators, and users with reading or visual limitations.

Time Lens thus bridges the gap between advanced AI technologies and heritage interpretation, turning complex architectural information into a seamless, interactive experience. Future enhancements may include expanding the model's monument database, improving regional language support, and enabling real-time mobile recognition. With continued development, Time Lens has the potential to become a comprehensive digital gateway to India's rich architectural legacy.

REFERENCES

1. Ayush Maheshwari, Nikhil Singh, Amrith Krishna, and Ganesh Ramakrishnan, "A benchmark and dataset for post-OCR text correction in Sanskrit," Findings of the Association for Computational Linguistics: EMNLP 2022, pp. 6258–6265, Dec. 2022.
2. K. R. Ingole and V. K. Shandilya, "Image restoration of historical manuscripts," International Journal of Computer Science and Engineering Technology (IJCSET), vol. 2, no. 4, pp. 102–105, 2011.
3. A. Tomar, M. Choudhary, and A. Yerpude, "Ancient Indian scripts image pre-

processing and dimensionality reduction for feature extraction and classification: A survey," International Journal of Computer Trends and Technology (IJCTT), vol. 21, no. 2, pp. 85– 90, Mar. 2015.

4. S. Guan, M. Lin, C. Xu, X. Liu, J. Zhao,J. Fan, Q. Xu & D. Greene, "PreP-OCR: A Complete Pipeline for Document Image Restoration and Enhanced OCR Accuracy" arXiv preprint arXiv:2505.20429v1, May 2025.
5. Shivraj Gaikwad, Renu Kachhoria, and Gitanjali Yadav, "AI-Based OCR for Digitizing Ancient Indian Texts: Preserving Linguistic Heritage and Overcoming Script Challenges," International Journal of Linguistics Applied Psychology and Technology (IJLAPT), vol. 2, no. 03 (Mar), pp. 1–12, Apr. 2025.
6. S. V. Khedaskar, M. A. Rokade, B. R. Patil, and T. P. N., "A survey of image processing and identification techniques," VIVA-Tech International Journal for Research and Innovation, vol. 1, no. 1, pp. 1–10, 2018.
7. A. Kunchukuttan, D. Kakwani, S. Golla, G. N.C., A. Bhattacharyya, M. M. Khapra & P. Kumar, "AI4Bharat- IndicNLP Corpus: Monolingual Corpora and Word Embeddings for Indic Languages," arXiv preprint arXiv:2005.00085v1, April 2020.
8. M. Wadhwani, D. Kundu, D. Chakraborty & B. Chanda, "Text Extraction and Restoration of Old Handwritten Documents," arXiv preprint arXiv:2001.08742v1, January 2020.