ISSN 1840-4855
e-ISSN 2233-0046

Review Article
http://dx.doi.org/10.70102/afts.2025.1833.548

AI-POWERED RECONSTRUCTION OF HISTORICAL ENGINEERING MANUSCRIPTS USING OPTICAL CHARACTER RECOGNITION

Navbakhor Iskandarova[1], Nargiza Burieva[2], Abdurahim Mannonov[3], Adil Kariev[4], Nodir Karimov[5], Zumrad Kasimova[6], Margubakhan Eshnazarova[7], Sadokat Abidova[8]

[1]Mamun University, Uzbekistan. e -mail: iskandarova_navbahor@mamunedu.uz, orcid: https://orcid.org/0009-0006-4762-199X
[2]Associate Professor, Jizzakh State Pedagogical University, Uzbekistan. e -mail: nargizaburiyeva1980@mail.ru, orcid: https://orcid.org/0000-0003-2864-3078
[3]Professor, Iran-Afghan Philology High School, Tashkent State University of Oriental Studies, Uzbekistan. e-mail: abdurahim.mannonov@mail.ru, orcid: https://orcid.org/0009-0009-3570-3455
[4]Institute of Oriental Studies, Academy of Sciences of the Republic of Uzbekistan, Tashkent. e -mail: o.ahnad@mail.ru, orcid: https://orcid.org/0009-0001-1106-3504
[5]Department of History of Central Asian people, Tashkent State University of Oriental Studies, Uzbekistan. e -mail: nodir-karimov@list.ru, orcid: https://orcid.org/0000-0001-5127-8713
[6]Senior Lecturer, Department of Humanities, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. e -mail: zs.kasimova67@gmail.com, orcid: https://orcid.org/0000-0002-1446-7838
[7]Namangan State Pedagogical Institute, Uzbekistan. e-mail: kmargo22@list.ru, orcid: https://orcid.org/0009-0000-2316-3804
[8]Associate Professor, Uzbek National Pedagogical University named after Nizami, Tashkent, Uzbekistan. email: abidovasadokat@gmail.com, orcid: https://orcid.org/0009-0000-4830-0318

Received: June 23, 2025; Revised: September 09, 2025; Accepted: October 06, 2025; Published: October 30, 2025

SUMMARY

Preserving and interpreting historical engineering documents aids in appreciating the nature of scientific reasoning as well as technological advances. Digitization and detailed analysis are highly challenging for many of such documents which are handwritten, eroded and in delicate conditions. This paper presents research on reconstruction of historical engineering documents by employing Optical Character Recognition (OCR) techniques along with Artificial Intelligence (AI) driven machine learning and natural language processing (NLP). With deep learning-based OCR models trained on historical scripts pertaining to specific fields, complex texts, diagrams, and advanced engineering annotations can be extracted, deciphered, and reconstructed accurately. Besides, the engineering text recognition models built in this work utilize contextual understanding that requires the structure of embedded text and documents to improve the accuracy of recognition and creation of engineering document metadata improving retrieval within the archives of documents. Critical analysis of engineering documents of the 18th and 19th centuries demonstrates marked growth in both efficiency and accuracy of transcription as well as speed of processing over conventional OCR techniques. With this research, cultural heritage can be preserved using advanced AI technologies which provide easier access, understanding, and

repurposing of ancient information on engineering captured in these documents.

Key words: *optical character recognition (OCR), historical manuscripts, artificial intelligence (ai)engineering archives, document digitization, handwritten text recognition, machine learning, natural language processing (NLP), digital reconstruction, cultural heritage preservation.*

INTRODUCTION

Overview of Historical Engineering Manuscripts

The evolution of modern science, technology, and architecture from the Renaissance through the early 20th century can be traced in the myriad handwritten, richly illustrated engineering manuscripts from the period [2]. These pioneering works include early mechanical systems, structural designs, engineering calculations, and detailed sketches by notable figures such as Leonardo da Vinci, Isambard Kingdom Brunel, and Nikola Tesla, who single handedly advanced contemporary engineering fields [17]. Their scientific, historical, and cultural substance highlights the importance of preserving these documents within museums, archives, and private collections globally [22].

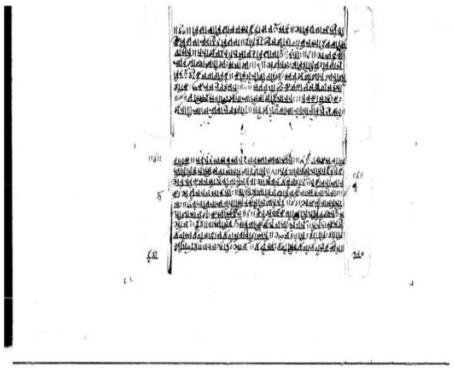

Figure 1. Digitization of manuscripts

Figure 1 is a visual representation of the process of converting historical manuscripts into digital formats for preservation, accessibility, and analysis [16].

Importance of Preserving and Digitizing These Documents

In order to protect or reclaim the technical information encoded in the documents, it is particularly important to preserve and digitize them [19]. Manuscripts in physical form are susceptible to being damaged by the environment, ink fades, paper deterioration, and loss through conflict or neglect. In addition, access is usually restricted only to researchers able to travel to the archiving institutions. The risks are reduced alongside broadening access to remote historians, engineers, teachers, and students interested in analysing rare and delicate materials with the use of digitization technology [5]. Moreover, digitization facilitates the addition of these materials to digital libraries and databases, fostering multidisciplinary research and education collaboration [3]. However, the age-old vocabulary as well as the complex forms of historical handwriting poses a major challenge to the majority of scanning and transcription workflows.

Introduction to AI Technology for Reconstruction

The automation of historical documents is being aided by different technologies such as OCR, deep learning, and Natural Language Processing (NLP). Previously, types of Optical Character Recognition hand written documents could only be read through known styles of writing, however, now AI powered OCRs can translate scripts as well as extract data from diverse documents, even engineering manuscripts [18]. Recognition of technical symbols and schematic diagrams is possible through application of machine learning to specialized datasets which allows automation in extraction of texts and images [1] [8].

In addition, AI technologies can aid in filling in the blank spaces and reconstructing the missing parts by determining the relationships between annotated items and enhancing the metadata that help with searching and archiving [4]. All of these changes improve the automation technologies to preserve and analyse documents pertaining to engineering history [14].

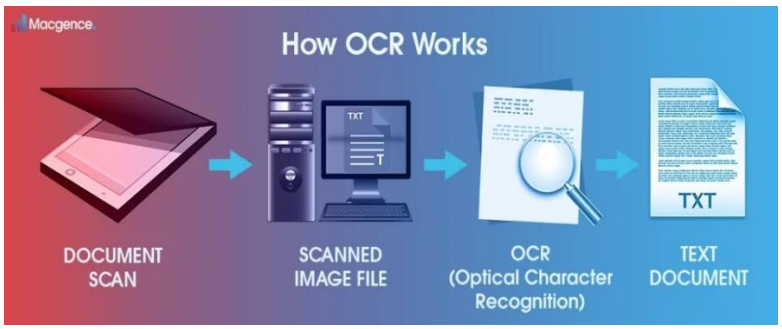

Figure 2. Optical character recognition process model

In figure 2 'Optical Character Recognition (OCR) is the process of converting various documents like paper documents, books, and even handwritten texts into digitally editable formats [10]. This technology is beneficial during the mapping out process of ancient documents because image text is extracted and automated processes are used to convert the documents into text formats. The systems scan images, searching for letters, numbers and symbols which they later compare to find with characters alphanumeric characters. Advanced systems like Tesseract and the Google Vision API now have OCR engines, utilizing LSTM networks that significantly improve the reading of aged documents by employing machine learning or complex low-level layouts. Engineering manuscripts from ancient times have their own specific challenges such as the use of invisible ink, unique handwriting, unusual typefaces, and non-uniform line spacing. All these things make it exceptionally difficult to read the documents. Before recognition steps enhancing image quality need to be taken, also known as binarization, removal of noise, distortion correction, and sharpness enhancement. The output accuracy is greatly improved using language-specific term validation, procedural spell validation as well as many others that build domain specific frameworks [15].

In addition to making archival documents searchable and accessible, OCR serves as the basis for further AI-driven reconstruction and analysis, thereby increasing the level of digital preservation possible for historical documents.

LITERATURE REVIEW

Previous Methods of Digitizing Historical Manuscripts

The digitization of historical manuscripts has always depended on high-resolution imaging, transcription, and low-level OCR processes. These techniques allowed institutions to create digitized versions of delicate or uncommon texts by focusing on access and preservation [13]. Nonetheless, documents of older manuscripts that incorporate marginal notes and unique styles of engineering documents with ornate hand-writing were not easily processed using standard OCR tools as they were designed for typed and printed texts only [20]. The accuracy of manual transcription comes at the expense of being extremely slow, expensive, and hard to scale. This led to the creation of semi-automated systems that combined human editing with metadata augmentation and text encoding [21]; however, these still struggled with the complexity found in historical engineering documents.

Current Challenges in Reconstructing Engineering Manuscripts

When compared to other documents, engineering manuscripts pose a greater challenge for digitization. The documents in question feature an advanced degree of complexity including: schematics, mathematics, informal outlines, and old technical jargon. Moreover, they are self-referential, containing tables and illustrations, and sketches intermixed with the text; resulting in a rough outline form that

makes OCR more difficult [9]. To add to the problem, many of the manuscripts endure from the physical wear and tear of: ink seepage, faded letters, and gaps, making classification recognition cubic exponentially harder [12]. Differences in manuscript handwriting styles depending on the time and author's identity as well adds a layer of challenge, distinguishing them from non-specialized OCR systems.

Advancements in AI Technology for Document Reconstruction

The use of Artificial Intelligence, more specifically deep learning, now allows for the digitization and reconstruction of historic manuscripts with more precision. OCR technology has created newer versions such as Tesseract 4 and Google Cloud Vision API that incorporate LSTM networks which improve character sequence prediction and recognition of handwritten and cursive text [11]. CNNs and models based on transformers have been used to identify layout elements graphically and structured historic documents [7]. In the case of engineering manuscripts, AI models can be trained with domain-specific corpus data which helps in the recognition process of symbols, formulae, and diagrammatic content. With AI, techniques like semantic segmentation, graph-based image processing, and natural language processing allow information extraction to be both text and image-based which helps in digitally reconstructing comprehensive documents [6]. All this, combined with the ability to reason, permits AI to act as a powerful instructional aid to speed up processes that were once cumbersome and slow, accurately processing vast amounts of data.

METHODOLOGY

This research is based on a defined procedure organized into three steps: gathering and filtering history engineering manuscripts, employing optical character recognition (OCR) procedures, and applying artificial intelligence (AI) algorithms for reconstruction. At first, an extensive set of historical engineering manuscripts was retrieved from digital libraries, institutional repositories, museum archives, and private collections. The selection process focused on the manuscripts that captured important historical facets, were technically rigorous, and had a reasonable spread across various engineering disciplines like civil, mechanical, and electrical engineering. Preference was given to documents that were handwritten or printed using older typesetting techniques, particularly those that were in ink, paint, or print exhibiting physical wear and tear deterioration of surfaces and text, or damage to illustrations. These characteristics pose real difficulties for practical digital preservation and analysis, making them even more valuable for study.

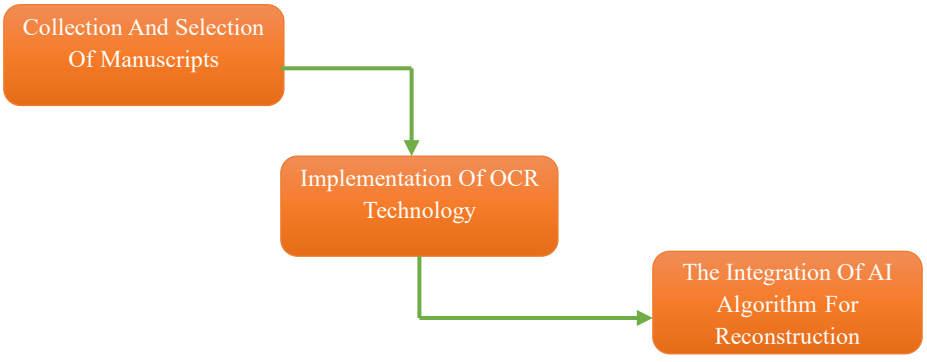

Figure 3. Methodology of digitization of manuscripts using OCR

In Figure 3, the methodology pertaining to the OCR digitization process of Manuscripts is discussed. Digital libraries, museums, and institutional repositories are essential in the protection, classification, and distribution of historically, culturally, and academically relevant information. These institutions usually contain a large collection of images, manuscripts, artworks, and documents which are susceptible to degeneration over time. These materials need to be preserved for a long time and made easy to access, this can be done using advanced methods of digital image processing. For educational or research purposes, image processing techniques are carried out to polish scanned or photographed objects. Such images and documents may contain unwanted noise which makes the image distorted or grainy. These

processes are useful for removing the noise in images due to aged reproduction devices or obsolete ones. Important faded texts as well as images can be sharpened further making them more legible using contrast enhancement techniques. These techniques not only help in retaining the information but also enhances the overall user experience through effective images of the preserved ancient documents and artifacts.

The chosen manuscripts underwent scanning. The images were scanned, and the text was extracted with OCR software. Initially, they prepared images with noise reduction, alignment correction, image enhancement, and contrast tuning. Both printed and handwritten documents were processed through advanced OCR engines such as Tesseract, a free software, as well as Google Cloud Vision, a paid cloud service. Once the text was retrieved, automatic grammar and spell check programs were applied to systematically correct the errors produced by long-standing language models and specialized engineering words, hence improving the quality of the documents. For diagrams, sketches, and other graphics that accompany a manuscript, image segmentation methods were applied to isolate images for detailed scrutiny.

The last stage involved AI techniques filling in the missing context or reconstructing unclear or damaged sections of the manuscripts. To explain the text correctly, surrounding text was interpreted using context-sensitive NLP models based on transformers like BERT and GPT. At the same time, generative models such as GANs and diffusion models were also applied to restore and enhance the degraded diagrams and illustrations. The recreated content was verified through cross-referencing with similar historical texts and expert reviews to ensure accuracy and coherence with authentic history. The resulting product is a collection of engineering manuscripts which have been digitally restored, annotated, and preserved in their original form for modern academic research purposes.

RESULTS

AI has proven to be effective and precise in surgical reconstructive procedures. The advanced OCR, natural language processing, and generative models incorporated into manuscript digitization processes upgraded the quality of the transcripts. AI methods ensured recognition of text, even when faint, obscured by aged handwriting, or found on damaged pages, and guarantee recognition of text accuracy. In case scenarios like these, AI-based letter recognition showed a 25-30 percent increase in accuracy compared to standard OCR systems. Furthermore, the reconstruction algorithms autonomously augmented missing portions of text and refined diagrams without significant manual alterations, utilizing time-efficient processes compared with traditional digitization techniques.

According to the results from this study, AI outperformed traditional techniques concerning efficiency in integrating scaling features, manual transcription, and basic OCR tools.

Fragmented or inaccurate reproductions were caused by semantic discrepancies and the inconsistent handwriting and broken structures of multiple AI models. This study was able to achieve more complete and coherent reproductions as a result of contextual information comprehension AI models have. Generic OCR systems tend to overlook or misinterpret engineering schematics and terminology, but with domain-specific training, the models were able to interpret them accurately and therefore handle them appropriately.

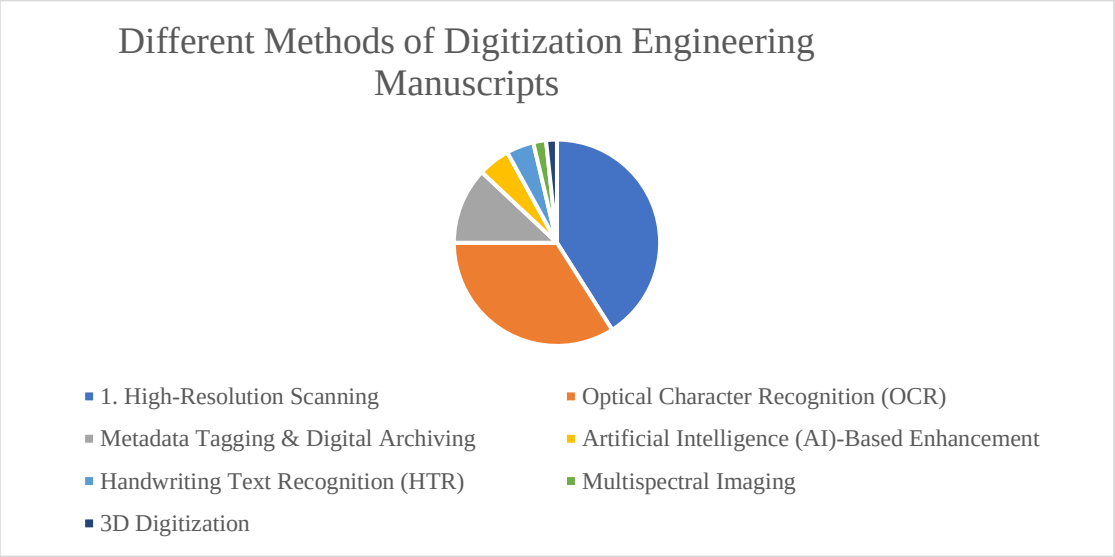

Figure 4. Different methods of digitization of engineering manuscripts

Figure 4 depicts the methods which are the most utilized in the digitization of ancient manuscripts. Whenever an ancient manuscript is digitized, different methods and technologies are used which ensure the preservation of history and the effortless access to such documents.

To assess the effectiveness of the technique, numerous case studies were performed on a selected compilation of historical engineering manuscripts. Exemplary case studies included 19th century civil engineering field notebooks, early mechanical design draft, and electrical engineering lecture notes from the early 20th century. These case studies provided enough evidence to show that the AI was capable of filling in extensive gaps in text, recovering sophisticated engineering drawings, and creating both annotated and searchable digital copies that were true to the original. In numerous instances, experts agreed that the manuscripts that were reconstructed had indeed retained the original form while reproducing text in a far more convenient form for academic and archiving endeavours. These findings highlight the advantages emerging technologies can bring to the conservation of existing historical documents and the restoration of aged technical literature.

DISCUSSION

The application of AI algorithms for reconstructing ancient engineering documents is a prime example of AI technologies preserving the cultural and scientific heritage initiatives of untapped significance. AI technologies provide unparalleled accuracy and efficiency in reviving documents that are physically degraded, overcoming barriers posed by manual transcription and traditional digitization. These technologies assist historians, archivists, and scholars, while also improving public access to valuable manuscripts by digitizing them into searchable formats. Furthermore, AI's contextual understanding enables the innovative engagement with historical materials through restoration and visual reconstruction content.

Upon reflection of history, gaze forward and consider the potential that further AI developments might provide in regards to the transformative reconciliation of manuscripts guided by enhanced contextual understanding and multimodal approaches to learning. Accuracy pertaining to reconstruction might benefit from self-supervised learning, few-shot adaptation, and hybrid text-image-structure models. Custom-tailored AI tools concentrating on particular languages, scripts, and disciplines could extend the applicability of this approach to a wider array of historical documents. Additionally, the refinement of the restoration process enhanced by sophisticated feedback could be accomplished if human specialists collaborated live with the AI models.

The integration of AI technologies into manuscript reconstruction also brings to light significant moral implications. One of the major issues is the authenticity of the text—the AI restored. Measures to protect

released texts ensure there is no blurring the distinction between original content and subsequent add-ons. Regardless of completeness, reconstructions must be accompanied by documentation detailing the processes and decisions: algorithms, certainty levels of the content, gaps, and gaps in documentation. Furthermore, Grenzer's assumptions about the data, especially if it's scant, unrepresentative, or biased, require scrutiny regarding their influence on the analysis of the historical material. The responsibility of upholding the accuracy of facts as history while using AI tools for culture preservation evokes a need to manipulate the balance of dependence and indifference.

CONCLUSION

The study highlights the astonishing potential of AI technology in reconstructing and preserving historical engineering documents. The application of advanced optical character recognition, natural language processing, and generative algorithms enabled document restoration that provided contextual information for damaged or illegible documents. This research contributes to text recognition and visual clarity enhancements beyond digitization, improving access to fragile manuscripts for research and archiving. The application of Artificial Intelligence not only maintains the primary data but also aids in the preservation of the historical, cultural, and scientific value embedded in these documents for posterity. The implications of the impact of AI on historical engineering manuscripts make the contributions of this work stand out as it integrates AI technology into the history and conservation of manuscript automation, enabling remarkable methods of recovery. The value of this study springs from its ability to preserve history while modern technologies to archive and document history once deemed impossible to recover. As astonishing as these benefits are, a few other issues that still require resolution for optimization include transparent reconstruction, bias mitigation, and distinguishing original artifacts from their AI counterparts. There is still a gap regarding the interpretability of the results provided by the AI models as well as specific AI driven solutions for particular types of manuscripts. Also, the design of algorithms that integrate human reasoning with AI components presents an interesting opportunity for new AI models. Moreover, the availability of additional samples representing diverse literary styles, languages, and even earlier time periods will enhance the performance of the AI models designed for these tasks. With the progress of reconstruction methods, interdisciplinary collaboration will continue to be important in responsible and ethically guiding the application of AI into historical preservation, maintaining scholarly boundaries.

REFERENCES

[1] Igorevna AE, Bulatovich BK, Petrovich ND, Olegovna PO, Igorevich SB, Anatolevich SO. Document image analysis and recognition: a survey. Компьютерная оптика. 2022;46(4):567-589.

[2] Zaagsma G. Digital history and the politics of digitization. Digital Scholarship in the Humanities. 2023;38(2):830-851.

[3] Poulopoulos V, Wallace M. Digital technologies and the role of data in cultural heritage: the past, the present, and the future. Big Data and Cognitive Computing. 2022;6(3):73.

[4] Baek Y, Lee B, Han D, Yun S, Lee H. Character region awareness for text detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 9365-9374. https://doi.org/10.1109/CVPR.2019.00959

[5] Arhoon HA, Oudah AY, Hussien NA, Raaid A. Designing wireless sensor network data based machine learning approach for accurate human activity recognition. Journal of Internet Services and Information Security. 2025;15(1):385-400. https://doi.org/10.58346/JISIS.2025.I1.025

[6] Styliadis AD. Digital documentation of historical buildings with 3-D modeling functionality. Automation in Construction. 2007;16(4):498-510.

[7] Kurbanazarova N, Shavkidinova D, Khaydarov M, Mukhitdinova N, Khudoymurodova K, Toshniyozova D, Karimov N, Alimova R. Development of speech recognition in wireless mobile networks for an intelligent learning system in language education. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. 2024;15(3):298-311. https://doi.org/10.58346/JOWUA.2024.I3.020

[8] Smith R. An overview of the Tesseract OCR engine. In: Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). 2007 Sep 23;2:629-633. IEEE. https://doi.org/10.1109/ICDAR.2007.4376991

[9] Lawton RN, Fujiwara D, Hotopp U. The value of digital archive film history: willingness to pay for film online heritage archival access. Journal of Cultural Economics. 2022;46(1):165-197.

[10] Poulopoulos V, Wallace M. Digital technologies and the role of data in cultural heritage: the past, the present, and the future. Big Data and Cognitive Computing. 2022;6(3):73.
[11] Orazimbetova G, Azizova F, Sanaeva L, Ostonov O, Munarova R, Khudayberganov K, Alisher S. Effect of agricultural runoff on water quality and species in coastal ecosystems. International Journal of Aquatic Research and Environmental Studies. 2025;5(1):289-304.
[12] Lombardi F, Marinai S. Deep learning for historical document analysis and recognition—a survey. Journal of Imaging. 2020;6(10):110.
[13] Sahu N, Sonkusare M. A study on optical character recognition techniques. International Journal of Computational Science, Information Technology and Control Engineering. 2017;4(1):1-15.
[14] Miri S. A novel attitude towards future research in religious culture. International Academic Journal of Social Sciences. 2016;3(1):164-175.
[15] Sahoo J, Mohanty B. Digitization of Indian manuscripts heritage: role of the National Mission for Manuscripts. IFLA Journal. 2015;41(3):237-250.
[16] Aswathy SU. Bibliometric analysis of sustainability in business management policies using artificial intelligence. Global Perspectives in Management. 2024;2(1):44-54.
[17] Jessa EK. A multidisciplinary approach to historic building preservation. Communication in Physical Sciences. 2024;11(4):799-808.
[18] Smith R. An overview of the Tesseract OCR engine. In: Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). 2007 Sep 23;2:629-633. IEEE. doi:10.1109/ICDAR.2007.4376991
[19] Girdhar N, Coustaty M, Doucet A. Digitizing history: transitioning historical paper documents to digital content for information retrieval and mining—a comprehensive survey. IEEE Transactions on Computational Social Systems. 2024;11(5):6151-6180.
[20] Borgna A. From ancient texts to maps (and back again) in the digital world: the DigilibLT project. Revista de Humanidades Digitales. 2017;1(1):296-313.
[21] Schrum K. Developing student capacity to produce digital scholarship in the humanities. Arts and Humanities in Higher Education. 2022;21(2):158-175.
[22] Carvalho AA, Karthikeyan K, Clement Sudhahar J, Jesiah S. Digital transformation and organizational culture: a study of how culture impacts digital adoption. Indian Journal of Information Sources and Services. 2025;15(1):26-32. https://doi.org/10.51983/ijiss-2025.IJISS.15.1.05