

Integrating Knowledge Discovery and Artificial Intelligence in Digital Preservation of Mongolian Medical Ancient Texts

Xuri Kou

Inner Mongolia Medical University, Hohhot 010110, Inner Mongolia, China

Copyright: © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The digitization of ancient Mongolian medical literature is a critical task for preserving traditional medical heritage and enabling its modern transformation. This study analyzes technical adaptability, knowledge mining pathways, and collaborative mechanisms in digitization through multi-regional empirical cases, bibliometric analysis, and web-based surveys. Key findings indicate that standardized metadata development, AI-assisted technologies, and cross-institutional collaboration are essential to overcoming preservation challenges. The research provides theoretical support for building a knowledge service system for Mongolian medical ancient literature, promoting its modern transformation and innovative applications.

Keywords: Mongolian medical ancient texts; Digital preservation; Knowledge discovery; Artificial Intelligence; Collaborative mechanism

Online publication: August 26, 2025

1. Introduction

As the core carrier of traditional Mongolian medicine, ancient Mongolian medical texts record unique medical systems formed since the 13th century, such as the “Three Roots Theory” and the “Seven Elements Theory”^[1]. These ancient texts are not only treasures of Mongolian culture but also an important part of China’s traditional medical heritage. Currently, the preservation of ancient Mongolian medical texts faces multiple challenges. Digital technology has provided a new paradigm for the protection of cultural heritage. In recent years, significant progress has been made in the digitization of ancient texts domestically and internationally. The “China Ancient Books Resource Database” implemented by the National Center for the Preservation and Conservation of Ancient Books has released a total of 106,000 ancient book titles^[2]. The Tibet Autonomous Region Library has enabled full-text search for 47,000 pages of documents through its digitization project of Tibetan biographical ancient texts^[3]. The “Northern Frontier Culture Ancient Books Resource Database” launched in Inner Mongolia has integrated 1.16 million ancient book images and constructed 240,000 knowledge graph triples^[4]. These practices have confirmed the effectiveness of digital technology in the “regenerative preservation” of ancient texts.

This study focuses on the digital preservation of ancient Mongolian medical texts. Through multi-case comparison and technical analysis, it explores the application pathways of knowledge discovery and artificial intelligence technologies

in the preservation of ancient Mongolian medical texts. The significance of this study is as follows: On the one hand, it can provide theoretical guidance and technical solutions for the digital preservation of ancient Mongolian medical texts, promoting the systematization and standardization of preservation work. On the other hand, through the application of knowledge discovery and artificial intelligence technologies, it can deeply explore the value of ancient Mongolian medical texts and facilitate the inheritance and innovation of Mongolian medical knowledge. In addition, the research results can also provide references for the preservation of ancient texts of other ethnic minorities, promoting the inheritance and development of China's excellent traditional culture.

2. Current Status and Challenges

2.1. Current Status of Ancient Text Preservation

The preservation of ancient Mongolian medical texts exhibits a distinct feature of dispersion. Currently, these texts are mainly distributed across three types of institutions:

Public Collection Institutions: These institutions serve as the main force in collecting ancient Mongolian medical texts. As a professional collection institution, the Museum of Mongolian Medicine at Inner Mongolia Medical University houses nearly 100 types of medical works, including the medical volumes of the Beijing woodblock Tibetan version of *Tanjur*, *Essence of Golden Juice*, *Blue Glaze*, and *Ten Million Relics*^[5]. The Museum of Mongolian Medicine Hospital in Fuxin, Liaoning, has a collection of over 4,000 ancient texts, among which 2,195 volumes have completed digital processing^[6]. These ancient texts cover basic theories, classic prescriptions, and manuscripts, holding high academic value.

Religious and Cultural Institutions: Scripture printing houses in monasteries across various regions store a large number of ancient Mongolian medical texts, most of which are preserved together with religious documents. Due to limited storage conditions, some ancient texts are at risk of damage. Particularly in monasteries in remote areas, the lack of professional protection equipment and personnel has led to worrying conditions for the preservation of ancient texts. These institutions usually lack systematic cataloging and organization, resulting in the value of ancient texts not being fully recognized or utilized.

Private and Overseas Collections: Twenty-one types of Mongolian-language unique copies collected by the Harvard-Yenching Library have been repatriated through photolithographic publication. However, more than 25,000 types of ancient texts remain scattered overseas globally^[7], and their digital repatriation needs to be promoted through international cooperation. In addition, some private collectors also hold a large number of ancient Mongolian medical texts, but these texts are often kept in poor conditions and lack professional organization. How to integrate these scattered ancient texts into a unified protection system is an urgent issue to be addressed.

2.2. Technical Bottlenecks and Barriers to Utilization

The current digitalization process of ancient Mongolian medical texts is confronted with multiple challenges:

Inadequate Adaptation of Recognition Technology: The low accuracy of OCR (Optical Character Recognition) for ancient Mongolian scripts (such as Uyghur-Mongolian script and Todo script) is attributed to variant characters and connected-writing variations. This is mainly caused by three factors: first, the diverse fonts of ancient Mongolian texts, with significant differences in writing styles across different periods and regions; second, the extensive use of abbreviations and variant characters in ancient texts, which increases the difficulty of recognition; third, the insufficient existing training data, which limits the performance of machine learning algorithms.

Weak Knowledge Connection: Most existing databases only remain at the level of image scanning, lacking structured analysis of the content in ancient texts. Compared with Tibet, which has completed metadata cataloging for 24,441 folios of ancient Tibetan texts, the work of semantic indexing for ancient Mongolian medical texts lags significantly behind. This leads to three consequences: first, the content of ancient texts is difficult to be understood and processed by machines; second, the connection between different pieces of knowledge cannot be established; third, in-depth knowledge mining

and analysis are hard to carry out.

Lack of Sharing Mechanisms: The digital platforms built by various institutions are independent of each other, with no unified data standards or sharing mechanisms in place.

Shortage of Talents: The digital preservation of ancient Mongolian medical texts requires interdisciplinary talents with expertise in both Mongolian medicine and digital technologies. At present, there is a severe shortage of such talents, which has become a bottleneck restricting the progress of related work. This shortage is mainly reflected in three aspects: first, the insufficient establishment of relevant majors in colleges and universities; second, the imperfect training mechanisms for existing personnel; third, a serious problem of brain drain^[8].

Inadequate Fund Investment: The digital preservation of ancient texts requires substantial financial support, covering equipment procurement, system development, personnel training, and other aspects. Currently, there are three problems regarding fund investment: first, the single source of funds, which mainly relies on government appropriations; second, the low efficiency of fund utilization; third, the low participation of social capital.

3. Key Technology Pathways

3.1. Standardized Organization and Digital Collection

Salvage conservation of documents is a prerequisite for digitization. In practical work, a systematic protection process needs to be established:

Restoration and Conservation Technologies: Modern technologies such as enzymatic consolidation are adopted to restore endangered ancient texts. The restoration process strictly adheres to the principle of “restoring as per the original state”, in order to preserve the historical original appearance of ancient texts to the greatest extent.

Digital Collection Standards: Image collection is conducted using non-contact scanners (with a resolution of ≥ 600 dpi) to ensure digital quality. Meanwhile, unified technical standards and work specifications are established, including three aspects: first, image collection standards, which specify parameters such as resolution, color mode, and file format; second, metadata standards, which define mandatory fields and optional fields; third, quality inspection standards, which clarify the requirements for qualified images and inspection methods.

Quality Control System: A quality control process covering links such as image quality inspection, metadata indexing, and content proofreading is established. Detailed operating procedures and recording systems are formulated for each link to ensure that problems are traceable and rectifiable.

3.2. Knowledge Base Construction and Intelligent Retrieval

The interlinking of knowledge from ancient Mongolian medical texts is the key to breaking the pattern of “valuing preservation over utilization”. The specific implementation pathways include:

Knowledge Graph Construction: Drawing on the experience of the “Northern Frontier Culture Ancient Books Resource Database”, the content of ancient texts is deconstructed into a multi-dimensional association system of “theory - prescription - medicine - disease”. The construction process involves four steps: first, entity recognition, which extracts entities such as personal names, place names, and medicine names from the text; second, relationship extraction, which identifies the relationships between entities; third, knowledge fusion, which integrates knowledge from different sources; fourth, knowledge storage, which uses a graph database to store the knowledge graph. This approach can establish a knowledge network containing tens of thousands of entities and relationships, laying a foundation for in-depth knowledge discovery.

Intelligent Retrieval System: A system supporting multi-modal retrieval is developed to achieve the following functions: first, the “image-to-text search” function, allowing users to upload images of handwritten Mongolian scripts, which the system recognizes and uses to retrieve relevant content; second, semantic expansion retrieval, enabling the system to understand the user’s query intent and automatically expand related terms; third, prescription composition

analysis, providing detailed information such as prescription components, effects, and indications; fourth, personalized recommendation, which recommends relevant resources based on the user's historical behavior. These functions greatly improve retrieval efficiency and user experience.

Clinical Decision Support: The digital achievements are integrated with clinical practice to develop the following applications: first, diagnosis and treatment plan recommendation, which suggests appropriate Mongolian medicine treatment plans based on patients' symptoms; second, drug compatibility analysis, which provides information on drug interactions and contraindications; third, efficacy evaluation, which assesses the effectiveness of treatment plans based on records in ancient texts and modern clinical data; fourth, knowledge push, which regularly pushes the latest research results and clinical experience to physicians.

4. Knowledge Mining and In - depth Utilization

4.1. Mining of Academic Research Value

Digital technology provides new methods and approaches for the research of Mongolian medicine ancient books.

Theoretical System Research: Through tools such as knowledge graphs, the development context and internal connections of Mongolian medical theories are visualized. For example, a knowledge graph of the "Three-Root Theory" can be constructed to show the interrelationships among the three basic elements of Heriyun, Xier, and Badagan, as well as their applications in physiology, pathology, and diagnosis-treatment. This visual analysis helps researchers more intuitively understand the Mongolian medical theoretical system and discover its laws and characteristics.

Cross-cultural Comparative Research: Using digital platforms, comparative research on Mongolian medicine and other medical systems is carried out. For example, the similarities and differences between Mongolian medicine and Chinese medicine in theoretical systems, diagnosis-treatment methods, and drug applications can be compared; the origin relationship between Mongolian medicine and Tibetan medicine can be analyzed; and the exchange and influence between Mongolian traditional medicine and Central Asian and West Asian medicines can be explored. These studies help to deepen the understanding of Mongolian medicine and promote the exchange and mutual learning among different medical systems.

4.2. Clinical Application Transformation

Digital achievements are highly valuable in clinical practice.

Development of Classic Famous Prescriptions: Based on the prescription records in ancient books, modern drug research and development are carried out. For example, an in - depth study is conducted on "Wen Guan Mu Si Wei Tang"^[9]. Through modern extraction processes, the content of its effective components is increased, and pharmacological and clinical trials are carried out to verify its curative effect and safety in the treatment of rheumatoid arthritis. This research and development model not only maintains the characteristics of traditional prescriptions but also meets modern drug standards, helping to promote the modernization of Mongolian medicine.

Standardization of Diagnosis-and-Treatment Protocols: Through data mining technology, diagnosis - treatment rules are extracted from ancient books to form standardized diagnosis - treatment programs. This standardization work helps to improve the standardization and repeatability of Mongolian medical diagnosis - treatment.

Application in Preventive Health Care: The health-care knowledge in ancient books is mined to develop health - management programs suitable for modern people. For example, the contents of seasonal health-care, dietary conditioning, and sports health-care in ancient books are sorted out. Combined with modern health-management concepts, a series of health products and services with Mongolian medical characteristics are developed. This not only helps to promote the preventive health-care concepts of Mongolian medicine but also meets the needs of modern people for a healthy life.

5. Construction of Collaborative Mechanisms

5.1. Multi-level Cooperation System

Establishing an effective collaborative mechanism is crucial for advancing the digital preservation of ancient Mongolian medical texts:

Institutional Collaboration: Develop a cross-institutional cooperation mechanism to integrate resources from libraries, museums, research institutions, and other parties. Specific measures include: first, establishing a joint catalog of ancient Mongolian medical texts to realize resource sharing; second, launching joint restoration projects to pool expert resources and technical capabilities; third, organizing academic exchange activities to facilitate experience sharing. For instance, an “Alliance for the Preservation of Ancient Mongolian Medical Texts” can be established, with regular working meetings held to coordinate and resolve major issues.

Regional Collaboration: Drawing on the model of the “Tibetan Medicine Joint Data Center across Four Provinces/Autonomous Regions”, build a regional cooperation platform. Key regions for Mongolian medicine—such as Inner Mongolia, Liaoning, Jilin, and Heilongjiang—can collaborate on the following initiatives: first, formulating unified digital standards and specifications; second, constructing a regional database of ancient Mongolian medical texts; third, launching joint talent training programs; fourth, organizing cross-regional academic research activities. Such regional collaboration helps break down administrative barriers and achieve optimal allocation of resources.

International Collaboration: Proactively engage in international exchanges and cooperation, with a focus on the following efforts: first, collaborating with countries such as Mongolia and Russia to conduct joint surveys and organization of ancient Mongolian medical texts; second, partnering with international organizations to apply for funding support for international cooperation projects; third, participating in international academic conferences to showcase China’s achievements in preserving ancient Mongolian medical texts; fourth, launching talent exchange programs to learn advanced international experiences. Through such international collaboration, global sharing and preservation of ancient Mongolian medical texts can be promoted.

5.2. Standardization and Regulation Development

Standardization is the foundation for ensuring digital quality:

Technical Standards: Formulate unified technical standards and work specifications. These standards should reference international standards^[10] while taking into account the unique characteristics of ancient Mongolian medical texts.

Metadata Standards: Establish a comprehensive metadata standard system. The metadata standards should support multilingual and multi-script environments to meet the needs of both Mongolian and Chinese bilingual applications.

Quality Specifications: Develop quality evaluation standards for digital achievements and establish a full-process quality control system.

6. Conclusions and Prospects

Through in-depth analysis of the current situation, this study identifies the main problems and challenges faced in the preservation of ancient Mongolian medical texts. On this basis, it proposes key technical pathways and implementation strategies to provide guidance for practical work. In the future, it is recommended to further strengthen efforts in the following aspects:

Improving the Standard System: Accelerate the formulation of relevant standards and technical specifications for the digital preservation of ancient Mongolian medical texts, covering standards for all stages, including data collection, storage, and service provision. It is suggested to refer to international standards while taking into account the uniqueness of Mongolian medicine, so as to develop practical and feasible standards and specifications.

Strengthening Talent Cultivation: Cultivate interdisciplinary talents who possess both knowledge of Mongolian medicine and proficiency in digital technologies. It is recommended to set up relevant professional directions in colleges

and universities to carry out academic education; at the same time, conduct on-the-job training to improve the capabilities of existing personnel; and introduce high-level talents to optimize the structure of the talent team.

Promoting Innovative Applications: Explore innovative applications of digital achievements in fields such as scientific research, clinical practice, and education. For example, develop intelligent diagnosis and treatment systems to support clinical decision-making; build virtual museums to enhance educational experiences; and conduct digital humanities research to innovate research methods.

Facilitating International Cooperation: Strengthen cooperation with international institutions to promote the global sharing and utilization of ancient Mongolian medical texts. Joint research projects can be carried out to share technical resources; international conferences can be held to exchange work experience; and an international cooperation network can be established to facilitate continuous communication.

The digital preservation of ancient Mongolian medical texts is a systematic project that requires the coordinated advancement of technology, talent, standards, and policies. Through the integrated application of knowledge discovery and artificial intelligence technologies, it is possible not only to improve the efficiency and quality of digital preservation but also to deeply explore the value of ancient Mongolian medical texts, thereby promoting the inheritance and innovation of Mongolian medical culture.

Funding

Youth Project of Inner Mongolia Medical University (Project No.: YKD2024QN008).

Disclosure statement

The author declares no conflict of interest.

References

- [1] Ba J, 2004, A Study on the History and Literature of Mongolian Medicine. Shenyang: Liaoning Ethnic Publishing House.
- [2] National Center for the Preservation and Conservation of Ancient Books. Digitalization and Activation of Ancient Books: Illuminating the Inheritance of China's Excellent Traditional Culture. China Cyberspace, 2025 (03): 18-22.
- [3] Ethnic Affairs Commission of the Tibet Autonomous Region. The Digitalization Project of Precious Ancient Books in the Tibet Autonomous Region Library Passes Acceptance. https://mw.xizang.gov.cn/xwzx/ztzl/mzwh/201912/t20191218_126505.html, 2019-12-18.
- [4] Ethnic Minority Ancient Books Collection and Research Office. The Northern Frontier Culture Ancient Books Resource Database Platform is Officially Launched. https://mw.nmg.gov.cn/zt/mzgjzl/mzgj/202404/t20240407_2490438.html, 2024-04-08.
- [5] Literature Exhibition Hall of the Museum of Mongolian Medicine, Inner Mongolia Medical University. <https://myyjy.immu.edu.cn/myybwg/gjwxzt.htm>, 2021-10-08.
- [6] Xinhua News Agency. Liaoning: 2,195 Volumes of Ancient Mongolian Medical Texts Realize Digital Maintenance. <https://www.163.com/dy/article/JMHHCQC705346RC6.html>, 2025-01-22.
- [7] Bi A W, 2015, A Collection of Mongolian Literature in the Harvard-Yenching Library of Harvard University, USA. Guilin: Guangxi Normal University Press.
- [8] Research on the Path of Preservation and Activation of Ethnic Minority Ancient Books—Taking the Southwest Region as an Example. Ethnic Art Studies, 2025, 38 (02): 152-160.
- [9] National Administration of Traditional Chinese Medicine, National Medical Products Administration. Notice

on Issuing the Catalogue of Ancient Classic Prescriptions (the Second Batch). <http://www.natcm.gov.cn/kejisi/gongzuodongtai/2023-09-01/31743.html/>, 2023-09-01.

- [10] Li H T, 2025, Digital Technical Specifications for Photolithographic Publication of Traditional Chinese Medicine Ancient Books. Chinese Journal of Library and Information Science for Traditional Chinese Medicine, 49 (02): 1-5.

Publisher's note

Whioce Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.