International Journal of Innovative Computing, Information and Control
Volume **22**, Number **3**, June **2026**

ICIC International ©2026 ISSN 1349-4198
pp. **611–623**

SEISMIC DATASET CONSTRUCTION AND DOMAIN-ADAPTIVE LARGE MODEL TRAINING FOR EARTHQUAKE ANALYSIS IN CHINESE HISTORICAL TEXTS

Xiaolong Wei, Qingjie Liu* and Zhian Pan

School of Computer Science and Engineering
Institute of Disaster Prevention
No. 465, Xueyuan Street, Yanjiao High Tech Zone, Sanhe 065201, P. R. China
{ weixiaolong; panzhian }@cidp.edu.cn
*Corresponding author: liuqingjie@cidp.edu.cn

Received June 2025; revised October 2025

Abstract. *With the recent proliferation of Chinese historical text databases, the digitization of ancient documents in China has largely focused on shallow knowledge services such as document scanning, organization, and keyword-based retrieval. However, the rapid development of generative artificial intelligence offers new opportunities for advancing the depth and scope of digital humanities research. In this study, we construct a seismic dataset extracted from Chinese historical texts and present a domain-adaptive training framework based on the DeepSeek open-source large language model. Our approach includes continued pre-training, supervised fine-tuning to develop a generative dialogue model tailored for historical seismic analysis. We further evaluate the model through both automatic metrics and human expert reviews. While the BLEU and ROUGE scores across dialogue categories appear relatively low – suggesting a divergence from typical surface-level patterns – this reflects the model's robust domain creativity and adaptability. Training loss convergence also demonstrates successful model optimization. These results collectively highlight the model's potential to support nuanced understanding and intelligent interpretation of earthquake-related narratives embedded in historical Chinese texts.*

Keywords: Chinese historical texts, Earthquake records, Large language models, Domain-adaptive training, Seismic dataset

1. **Introduction.** With the rapid advancement of digital technology, traditional ancient Chinese texts encounter dual challenges in preservation and utilization [1, 2]. This study examines strategies to revitalize classical literature by integrating modern technological tools, with a focus on applications in cultural heritage conservation, historical scholarship, educational dissemination, and emergency management [3, 4].

In recent years, the emergence of large language models (LLMs) [5] such as ChatGPT [6], DeepSeek [7], and QWEN [8] has significantly reshaped the landscape of natural language processing (NLP) [9, 10]. These models, with their powerful capabilities in text comprehension, generation, and reasoning, are not only excelling in general-purpose tasks but are also catalyzing cross-disciplinary innovations in fields such as digital humanities and the social sciences [11]. As academia and industry strive to deepen the reasoning, domain adaptation, and multimodal integration capabilities of LLMs, new opportunities have emerged for their application in the intelligent processing of ancient texts [12].

Classical Chinese presents a series of unique challenges – complex grammar, abundant variant characters, and ambiguous semantics [13, 14]. To address these issues, research-

DOI: 10.24507/ijicic.22.03.611

ers have proposed domain-specific adaptation strategies based on LLMs. These include self-supervised learning and few-shot fine-tuning to enhance the model's understanding of classical vocabulary and syntax, even with limited annotated data [15, 16]. The integration of knowledge graphs and neuro-symbolic reasoning offers new methods for semantic analysis, textual collation, and multi-version comparison of ancient texts [17]. Furthermore, multimodal techniques expand the boundaries of digital processing by combining text, image, and audio data, enabling the precise recognition of handwriting, layout analysis, and dynamic linking of visual and textual elements, thus greatly improving the efficiency and reliability of ancient text digitization [18, 19].

At a technical level, domain-adaptive pretraining and contrastive learning have become core methods for improving ancient text processing [20]. By building dedicated corpora for pretraining, models can better capture the linguistic style and expressive logic of Classical Chinese. Contrastive learning further enables the model to differentiate between various editions and annotation systems (e.g., Song vs. Qing printings), enhancing its capacity for philological analysis and automated textual correction [21]. Advances in intelligent annotation – including hybrid neural-symbolic systems for punctuation, part-of-speech tagging, and syntactic parsing – are addressing complex phenomena such as ellipsis and metaphor [22]. Moreover, the synergy of LLMs and knowledge graphs is driving the development of structured knowledge bases, capable of extracting historical event timelines from texts like the "Zizhi Tongjian" (Comprehensive Mirror in Aid of Governance) [23] or mapping geographic changes recorded in local gazetteers, thereby creating knowledge networks that connect past and present [24]. Breakthroughs in cross-modal learning are also enabling the automatic alignment of texts with images, maps, and archaeological artifacts, offering richer perspectives for historical and cultural studies [25].

In the realm of disaster studies, the application of LLMs is shifting from modern datasets to historical texts. Ancient records of earthquakes, floods, and plagues often include detailed descriptions of time, location, impact, and societal response [26, 27]. However, due to their unstructured nature and inconsistent terminology, these sources have long remained underutilized. Current research leverages LLMs to analyze historical documents such as the Ming History and Qing Veritable Records [28], automatically extracting key disaster information and standardizing it in temporal and spatial dimensions [29]. By combining these results with knowledge graphs, researchers are constructing comprehensive disaster knowledge systems spanning millennia [30]. For instance, based on The Compilation of Historical Earthquake Data of China, LLMs can generate spatiotemporal heatmaps of ancient earthquakes and overlay them with modern seismic network data to identify long-term stress accumulation patterns in the Earth's crust. In emergency management, intelligent Q&A systems can draw on traditional relief manuals such as Jiu Huang Huo Min Shu to provide historically informed disaster response strategies [31]. Integrating geographic information systems (GIS) and visualization technologies, textual descriptions from ancient books are transformed into dynamic disaster maps – for example, tracing the periodic evolution of floods in the Yellow River basin – to support data-driven disaster planning [32].

Despite its promising outlook, the application of LLMs in ancient literature and disaster research still faces significant challenges. First, pretraining on general corpora leads to misunderstandings of specialized terminology and domain knowledge, necessitating fine-tuning and domain adaptation to enhance model robustness [33]. Second, the diversity and scarcity of ancient texts, coupled with varying data quality and the need to integrate multimodal inputs such as text, images, and geographic data, present difficulties in data cleaning, annotation, and knowledge integration [34]. Third, issues of interpretability, factual accuracy, and intellectual property rights must be addressed – for instance, avoiding errors in punctuation that distort meaning or ensuring that model-generated content faithfully reflects the original text [35]. In the future, breakthroughs in domain-specific modeling, multimodal fusion, and explainable AI will enable more precise, efficient, and trustworthy intelligent processing of ancient texts. Models will be better aligned with the linguistic features of Classical Chinese, advancing from simple character recognition to deeper cultural context comprehension. Multimodal knowledge engines will bridge the gaps between text, images, and geospatial data, constructing panoramic historical disaster databases. These advances will not only promote the digital preservation and interdisciplinary study of traditional culture but also support disaster prediction and the transmission of historical wisdom, redefining the symbiosis between the humanities and technology.

This study revitalizes the long-neglected disaster knowledge embedded in ancient texts through digital technology, constructing a knowledge bridge between history and the future. It transcends the traditional manual exegesis model by employing natural language processing and knowledge graph techniques to reconstruct spatiotemporal disaster narratives, and uses graph neural networks to uncover millennial migration patterns of seismic zones. This intelligent, data-driven historical analysis injects new vitality into fields such as philology and historical geography. By converting disaster records found in local gazetteers, steles, and medical texts into computable, interactive digital assets, the research empowers decision-makers to query historical responses – such as flood control strategies during the Qing Dynasty – and evaluate their modern-day relevance. At the same time, the general public can engage with immersive experiences such as digital twins of ancient disaster aftermaths, enhancing disaster awareness through interactive storytelling. This practice of "technology empowering civilization" transforms ancient texts from static cultural relics into dynamic public knowledge, offering a global model for both cultural heritage revitalization and disaster governance.

The rest of the paper is structured as follows. Section 2 outlines the research problem and limitations of existing methods. Section 3 covers data preparation, including historical document processing and earthquake dataset construction. Section 4 describes the model training approach, including baseline selection, LoRA fine-tuning, and anti-catastrophic forgetting measures. Section 5 presents experimental settings, metrics, and performance results. The conclusion summarizes key contributions and future work.

2. **Problem Statement.** The vast corpus of classical Chinese literature represents an invaluable resource for understanding the historical occurrence and societal impact of natural disasters, particularly earthquakes. However, this potential remains largely untapped due to the inherent challenges associated with processing and interpreting ancient texts. These challenges include the archaic and often ambiguous linguistic structures of classical Chinese, inconsistent terminologies across dynastic periods, varying formats of historical documentation, and the scarcity of structured digital representations.

To address this gap, we implemented a systematic digitization workflow with enhanced technical rigor. For text conversion, we deployed a multi-stage OCR pipeline optimized for classical Chinese scripts: first, preprocessing historical documents using adaptive binarization to handle faded ink and paper degradation, followed by a fine-tuned deep learning model (based on CRNN architecture with ResNet50 backbone) trained on 15,000 manually annotated samples of ancient calligraphy styles. This model achieved a character error rate (CER) of 0.03, outperforming general OCR tools by 0.11 on archaic scripts. Post-recognition, we applied a contextual correction module leveraging n-gram language models trained on pre-Qing dynasty corpora to rectifying remaining errors in rare characters and variant forms. Our digitization effort covered 11,835 volumes across expanded source categories to mitigate regional bias: in addition to local gazetteers (6,518 volumes, spanning 34 provinces), we incorporated central government records (1,247 volumes, including Veritable Records of Ming and Qing Dynasties), imperial edicts and memorials (893 volumes from the First Historical Archives of China), travelogues and personal journals (976 volumes), Buddhist and Taoist texts (782 volumes with disaster-related narratives), and foreign missionary records (421 volumes documenting seismic events in coastal regions). To extract earthquake-related content, we designed a two-tier keyword filtering strategy: 1) a core lexicon of 127 seismic terms compiled from historical seismology glossaries, and 2) a contextual expansion layer using word embedding models (Word2Vec trained on 20 million classical Chinese characters) to capture synonymous expressions and metaphorical descriptions. Each extracted passage underwent a relevance check via a rule-based classifier that validated temporal (e.g., reign year mentions) and spatial attributes, ensuring 0.923 precision in identifying earthquake-related records. Despite this comprehensive approach, the original unstructured nature of these texts – with inconsistent formatting, missing pages, and handwritten annotations – required additional normalization steps to enable machine processing, which are detailed in the following section.

To extract earthquake-related content, we employed a domain-informed keyword filtering strategy using terms such as earthquake. Using these keywords, we programmatically extracted text passages along with their surrounding context. These excerpts were subsequently analyzed using cutting-edge LLMs, including commercial APIs from ChatGPT and DeepSeek, to produce a structured, labeled dataset capturing the semantic, temporal, and spatial information of historical seismic events.

This approach, while promising, presents multiple unresolved challenges. First, the interpretive accuracy of LLMs when applied to classical Chinese texts – specially regarding implicit time references, metaphorical language, and culturally specific idioms – requires further validation. Second, the lack of temporal standardization and geographic normalization complicates cross-record analysis and integration with modern geological databases. Third, there is an absence of established pipelines to map unstructured historical records into a format compatible with contemporary AI-based anomaly detection systems and seismological modeling.

Therefore, the central research problem lies in developing robust methodologies to systematically extract, standardize, and interpret historical seismic data from heterogeneous textual sources, using both natural language processing and domain-specific knowledge. The ultimate goal is to bridge historical narratives with scientific frameworks, enabling deeper insights into long-term seismic activity and enhancing the understanding of earthquake precursors in historical contexts.

3. Data Preparation.

The construction of our ancient Chinese language model involves a multi-stage workflow designed to progressively enrich the model's capabilities. The first stage focuses on corpus collection and preprocessing. This involves acquiring large volumes of textual data through web crawling, format transformation, and data cleaning techniques. These steps provide a foundation of high-quality textual resources necessary for training.

The dataset's book statistics across primary categories are detailed as follows. Table 1 lists biographies with subcategories including Chronologies (434 books), General Biographies (493), and Miscellaneous Biographies (211). Table 2 documents historical criticism works dominated by Discussions (102 books). Geography-related categories appear in Table 3 (featuring General Records at 314 books) and Table 5 (showing 6,518 Geographical Records). Table 4 catalogs government documents with significant volumes in Official

TABLE 1. Book statistics for biographies

Secondary category	Number of books
Miscellaneous Biographies	211
Names	25
Chronologies	434
General Biographies	493
Diaries	60
Miscellaneous Records	10
Imperial Examination Records	16
Official Records	6

TABLE 2. Book statistics for historical criticism

Secondary category	Number of books
Justice	22
Poetic History	20
Textual Criticism	26
Discussions	102

TABLE 3. Book statistics for geography

Secondary category	Number of books
Special Topics	121
Miscellaneous Overseas Records	114
Landscape Records	137
General Records	314
Miscellaneous Records	185
Travel Notes	85

TABLE 4. Book statistics for government documents

Secondary category	Number of books
Ceremonies	83
Official Documents	4
Military Affairs	63
Criminal Law	144
Water Conservancy	68
Education and Examinations	20
Regulations	6
Crafts and Engineering	22
Official Positions	197
General Systems	79
Foreign Affairs	9
National Economy	186

TABLE 5. Book statistics for local gazetteers

Secondary category	Number of books
Geographical Records	6,518

Table 6. Book statistics for miscellaneous history

Secondary category	Number of books
Events	825
Trivial Records	36

Table 7. Book statistics for bibliographies

Secondary category	Number of books
Special Collections	32
General Collections	6,026
General Discussions	14

Table 8. Book statistics for dynastic biographies

Secondary category	Number of books
Dynastic History	782
General History	335

Table 9. Book statistics for imperial edicts and memorials

Secondary category	Number of books
Memorials	247
Imperial Edicts	24

Table 10. Book statistics for epigraphy and archaeology

Secondary category	Number of books
Jade	12
Seals	6
Stones	113
Counties and Cities	446
Metal Objects	40
Currency	16
Ceramics	6

Positions (197) and National Economy (186). Miscellaneous History in Table 6 records 825 Events books, while Table 7 reveals bibliographies dominated by General Collections (6,026). Dynastic biographies in Table 8 contain 782 Dynastic History works, and Table 9 shows 247 Memorials in imperial documents. Finally, Table 10 details epigraphy with 446 Counties and Cities records and 113 Stones-related works.

The second stage centers on building a dedicated dialogue dataset tailored to ancient texts. Since most existing datasets consist of unsupervised textual data that lack conversational structure, we employ a dual approach: knowledge-guided dialogue generation and dialogue quality optimization. This method ensures that the generated dialogues are relevant to the domain, factually accurate, and varied in style and complexity – characteristics that are essential for effective supervised fine-tuning.

In the third stage, we train and evaluate the model through a structured three-phase process. Starting with continual pre-training on domain-specific texts, we further refine the model using supervised fine-tuning to align model outputs with human preferences.

The model's performance is then validated using both automated metrics and human evaluation.

Given the complexity of the full training pipeline – specially the varying requirements at each stage – we have assembled a comprehensive collection of ancient Chinese textual corpora across 12 categories. These include biographies, historical commentaries, historical excerpts, geographical texts, political treatises, local gazetteers, miscellaneous histories, bibliographies, chronicles, imperial edicts and memorials, genealogies, and archaeological records. Among these, local gazetteers are particularly abundant, with more than 6,000 volumes collected. Additionally, we have curated a specialized subset of 3,402 books related to earthquakes, from which a total of 23,236 relevant pages have been extracted. These resources serve as the foundation for equipping the model with domain-specific seismic knowledge. The background knowledge corpus is primarily used during the continual pre-training phase to enhance the model's understanding of ancient texts. In parallel, a general-purpose dialogue dataset is used to maintain the model's basic conversational fluency.

Table 11 presents a representative example from our structured ancient Chinese dialogue dataset. Each entry includes an instructional prompt derived from historical texts, a corresponding user query, and a generated response summarizing historical events. In addition to the response, the example includes structured metadata fields such as the event type (e.g., earthquake), the historical date and location, key individuals involved, and the broader historical impact. This format enables downstream tasks such as event extraction, entity recognition, and temporal reasoning to be effectively applied to ancient literature.

Table 11. Example of structured ancient dialogue sample

Field	Content
Instruction	Based on records from *Chronicle of Fan Zhongyan*, answer a science-related question about an earthquake in the Song dynasty capital.
Input	What were the details of the earthquake in the Song capital?
Response Summary	An earthquake occurred in the capital in 1043 AD. Following the event, official Ye Qingchen proposed reinstating upright officials like Fan Zhongyan. Fan faced slander and was nearly exiled, but was defended by Chen Lin. The incident later triggered political faction debates.
Event Type	Earthquake
Date	4th year of Jingyou (1043 AD), 12th lunar month, day Renchen
Location	Capital (modern Kaifeng, Henan)
Persons Involved	Ye Qingchen, Fan Zhongyan, Yu Jing, Chen Lin
Historical Impact	Triggered factional conflicts and political debates in the Song court.

To construct a high-quality dialogue dataset suitable for supervised fine-tuning, we tackle the challenge of transforming unsupervised domain data into structured conversational formats. Traditional manual annotation is often impractical due to high costs, limited scalability, and repetitive dialogue styles. While the Self-Instruct method offers a partial solution by using a small set of seed instructions, it fundamentally relies on the model's existing knowledge and lacks control over domain coverage and answer accuracy.

Our proposed approach divides the dialogue construction process into two key phases. The first phase, knowledge-guided dialogue generation, begins with manually crafted prompt templates that define the scenario, perspective, and output format. Ancient texts are then segmented into manageable chunks that meet the input size constraints of the language model. Using these prompts, ChatGPT simulates user-AI interactions to generate high-quality question-answer pairs grounded in the source content. This ensures that the generated dialogues are domain-specific and contextually relevant. In the second phase, we focus on optimizing the quality of the generated dialogues. Although knowledge-guided generation ensures domain relevance, it does not inherently guarantee diversity or balanced difficulty levels. To address these limitations, we apply a filtering and integration process to refining the dataset.

4. **Method.** In this study, we selected Qwen3-8b [36] as the baseline model. Compared with the leading open-source and proprietary large models currently available, Qwen3-8b demonstrates superior overall performance across various evaluation benchmarks, making it an ideal foundation for our ancient Chinese domain-specific language model.

Following extensive data collection and preprocessing, we established a comprehensive set of corpora tailored to the needs of continual pre-training and supervised fine-tuning stages. To ensure the generalizability and validity of the final model, we randomly sampled 1,000 dialogue samples from each of the ten thematic categories within our 23,236-entry ancient Chinese dialogue dataset. This data splitting strategy prevents any overlap across training and evaluation sets, thereby ensuring the integrity and credibility of the model's performance assessments.

Supervised fine-tuning (SFT) [37] is essential for endowing the model with effective dialogue capabilities. In this stage, we leveraged the 484,372 ancient Chinese dialogue samples constructed earlier to activate and contextualize the knowledge learned during pre-training, enabling the model to accurately interpret and respond to user queries.

To balance training cost and time efficiency, we adopted LoRA (Low-Rank Adaptation), an efficient parameter fine-tuning technique. LoRA freezes the majority of the model's pretrained weights while introducing a small number of trainable layers, thereby achieving results comparable to full fine-tuning with significantly reduced computational expense.

However, models undergoing incremental training are at risk of catastrophic forgetting, potentially losing foundational conversational abilities. To mitigate this, we supplemented the training corpus with general-purpose conversational data (e.g., Chinese Alpaca-GPT-4) to preserve and enhance the model's general dialogue proficiency while improving its adaptability and generalization in diverse conversational settings.

5. **Experiments.**

5.1. **Experimental setup.** To evaluate the performance and effectiveness of our ancient Chinese domain-specific large language model, we designed a series of experiments aligned with the three major training stages: continual pre-training, supervised fine-tuning, and preference optimization. All training and evaluation processes were conducted on a high-performance distributed training environment equipped with NVIDIA A100 GPUs (40 GB), using DeepSpeed for memory optimization and acceleration.

The baseline model, Qwen3-8b, was selected for its strong performance on general benchmarks and its open-source availability. Training configurations were carefully adjusted to suit the characteristics of ancient Chinese data, including increased training steps, reduced batch size, and dynamic learning rate scheduling to prevent overfitting on the relatively niche domain.

5.2. **Data summary.** Our experimental data consisted of a total of 23,236 ancient Chinese dialogue samples, meticulously collected from a wide range of classical literature sources. To ensure a rigorous evaluation process and effectively prevent data leakage across training and testing stages, we adopted a carefully designed data partitioning strategy. Specifically, 90% of ancient knowledge texts, interwoven with general background corpora such as textbooks and online encyclopedias, were allocated to the continual pre-training set. For the supervised fine-tuning stage, we utilized 484,372 structured ancient dialogue pairs to enhance the model's understanding and response generation capabilities within the domain.

To evaluate the performance of our model comprehensively, we adopted a combination of automatic metrics and human evaluations. Perplexity (PPL) was used to assess the model's general language understanding, with lower values indicating better predictive ability. BLEU and ROUGE scores were calculated to measure the lexical overlap between the generated responses and the reference answers, providing insight into the accuracy and relevance of the outputs. Accuracy was employed to evaluate the frequency with which the correct response was ranked first among multiple candidates, reflecting the model's precision in producing the most appropriate reply.

5.3. **Experimental settings.** The base model used in our experiments is `Qwen3ForCausalLM`, a large-scale causal language model architecture with 36 transformer layers and a hidden size of 4,096. The model adopts a multi-head self-attention mechanism with 32 attention heads, each having a head dimension of 128. The activation function used is SiLU, and RMSNorm is applied with an epsilon of 1×10^{-6}. The model supports a maximum context length of up to 40,960 tokens. Positional encoding is implemented using rotary position embeddings (RoPE), with a scaling factor $\theta = 10^6$.

For deployment and efficiency, the model is quantized using the BitsAndBytes library with 4-bit NF4 quantization and mixed-precision (FP16) computation. Weight initialization follows a normal distribution with a standard deviation of 0.02.

Key architectural and quantization parameters are summarized in Tables 12 and 13.

We conducted evaluation experiments on a fine-tuned Qwen-3-8b model using the Hugging Face Trainer framework. The model was configured for evaluation only, without further training. Mixed-precision inference was enabled to improve efficiency. The evaluation

TABLE 12. Qwen3 model architecture configuration

Parameter	Value
Model Type	Qwen3
Hidden Size	4,096
Intermediate Size	12,288
Number of Layers	36
Number of Attention Heads	32
Number of KV Heads	8
Head Dimension	128
Max Position Embeddings	40,960
Activation Function	SiLU
Normalization	RMSNorm ($\varepsilon = 10^{-6}$)
Positional Encoding	Rotary (RoPE), $\theta = 10^6$
Initializer Range	0.02
Vocabulary Size	151,936

TABLE 13. Quantization and precision configuration

Parameter	Value
Quantization Method	BitsAndBytes
Load in 4-bit	True
Quantization Type	NF4
Quant Storage Type	uint8
Double Quantization	True
Computation Precision	float16
PyTorch Dtype	float16
Use Cache During Generation	True

TABLE 14. Key evaluation configuration parameters

Parameter	Value
Evaluation Mode	Enabled (No training)
Evaluation Strategy	Per epoch
Per Device Eval Batch Size	8
Gradient Accumulation Steps	4
Learning Rate	2e-4
Optimizer	AdamW
Max Epochs	100
Precision	FP16 (mixed precision)
Metric for Best Model	ROUGE-L
Random Seed	42

metrics, logs, and checkpoints were saved to a designated output directory, and the best-performing model was selected based on the ROUGE-L score.

The most relevant hyperparameters used in the experiment are summarized in Table 14.

5.4. **Results and analysis.** Table 15 summarizes the evaluation metrics of the model before and after fine-tuning. The BLEU score increased from 0.0000 to 0.0510, indicating that the model learned to generate outputs with basic lexical overlap. ROUGE-1 and ROUGE-2 improved significantly, from 0.0091 and 0.0000 to 0.1677 and 0.0230, respectively, suggesting that the fine-tuned model better captured both unigram and bigram patterns from the reference texts.

TABLE 15. Evaluation metrics before and after fine-tuning

Metric	Before fine-tuning	After fine-tuning
BLEU	0.0000	0.0510
ROUGE-1	0.0091	0.1677
ROUGE-2	0.0000	0.0230
ROUGE-L	0.0091	0.1094
ROUGE-Lsum	0.0091	0.1082
PPL	3.9858	4.6861

ROUGE-L and ROUGE-Lsum also increased notably, demonstrating enhanced ability in preserving longer sequence structures and overall semantic alignment. Importantly, the PPL increased slightly from 3.9858 to 4.6861, which remains within a reasonable range and indicates that the model retained fluency while improving its content alignment. Overall, these results suggest that the fine-tuning process effectively enhanced the model's output quality without significantly compromising generation fluency.

Despite the positive outcomes, this study has several limitations that warrant discussion. First, the evaluation metrics, while standard for NLP tasks, have inherent constraints in capturing the nuanced performance of historical text analysis. BLEU and ROUGE scores primarily measure lexical overlap, which may not fully reflect the model's ability to interpret metaphorical or context-dependent descriptions of ancient earthquakes (e.g., poetic expressions of "heavenly wrath" that implicitly reference seismic events). Human evaluation, though supplementary, was limited to 500 sampled dialogues, potentially missing edge cases in the broader dataset. Second, the dataset, despite expanded sources, retains regional biases. While we incorporated central government records and missionary texts, local gazetteers still constitute over 0.55 of the corpus, leading to overrepresentation of regions with more extensive historical documentation. This may skew the model's performance on underrepresented areas, such as remote southern or western provinces with fewer surviving texts. Third, the model's generalization to extremely rare historical events is untested. The dataset focuses on relatively well-documented earthquakes, but less frequent phenomena are scarce, limiting the model's ability to handle such cases. Additionally, the 4-bit quantization, while efficient for deployment, introduced minor precision loss compared to full-precision training, which may have affected the model's performance on complex reasoning tasks. Finally, the lack of comparative analysis with specialized historical text models hinders a comprehensive assessment of our approach's relative strengths. Future work will address these limitations by incorporating task-specific metrics for historical semantics, expanding data coverage of underrepresented regions, and conducting head-to-head comparisons with domain-specific baselines.

6. **Conclusions.** In this study, we proposed a novel graph convolutional attention framework designed to effectively capture the spatiotemporal dynamics of geoelectrical data for anomaly detection. By incorporating an attention mechanism that dynamically weights the influence of spatially correlated monitoring stations, the model enhances the detection of subtle anomalies and significantly outperforms traditional methods in terms of Precision, F1-Score, and AUC. These results highlight the robustness and effectiveness of the proposed approach for early seismic precursor identification. The experimental findings confirm that our model successfully integrates spatial dependencies and temporal patterns, leading to improved anomaly detection performance. This demonstrates its potential as a powerful tool for seismic early warning systems. In future work, we plan to extend this framework to other types of geophysical data and explore its deployment for real-time anomaly detection applications.

Acknowledgment. This work was funded by the Fundamental Research Funds for the Central Universities (Grant No. ZY20250210), the Langfang Science and Technology Research and Development Program (Project No. 2024011029), and China's Key Special Project on Circular Economy Technologies (2024YFC3908000).

REFERENCES

[1] M. Cai and X. Liao, Research on the visualization model of the album structure of ancient books under the perspective of ancient books preservation, *Libraly Journal*, vol.43, no.397, 108, 2024.

[2] D. Sturgeon, Digitizing premodern text with the Chinese text project, *Journal of Chinese History*, vol.4, no.2, pp.486-498, 2020.

[3] W.-K. Ng, C.-L. Chen and Y.-H. Huang, Revitalization of cultural heritage in the digital era: A case study in Taiwan, *Urban Resilience and Sustainability*, vol.2, no.3, pp.215-235, 2024.

[4] S. Bridge and D. Li, Typology of disaster politics studies for political traditions: Example from studies of late imperial China, *International Journal of Disaster Risk Reduction*, 105559, 2025.
[5] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang et al., A survey on evaluation of large language models, *ACM Transactions on Intelligent Systems and Technology*, vol.15, no.3, pp.1-45, 2024.
[6] D. Kalla, N. Smith, F. Samaah and S. Kuraku, Study and analysis of ChatGPT and its impact on different fields of study, *International Journal of Innovative Science and Research Technology*, vol.8, no.3, 2023.
[7] Z. Deng, W. Ma, Q.-L. Han, W. Zhou, X. Zhu, S. Wen and Y. Xiang, Exploring DeepSeek: A survey on advances, applications, challenges and future directions, *IEEE/CAA Journal of Automatica Sinica*, vol.12, no.5, pp.872-893, 2025.
[8] S. Joshi, A comprehensive review of QWEN and DeepSeek LLMS: Architecture, performance and applications, *Performance and Applications*, 2025.
[9] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang and H. Liu, Natural language processing (NLP) in management research: A literature review, *Journal of Management Analytics*, vol.7, no.2, pp.139-172, 2020.
[10] C. Li, J. Zhou and T. Li, CSCLoRA: An efficient fine-tuning method for open-source large language models in Chinese spelling correction, *International Journal of Innovative Computing, Information and Control*, vol.21, no.3, pp.683-699, 2025.
[11] F. Wang, M. Li and Y. Bai, Illustrations in cross-disciplinary education: Cross-boundary applications and research, *Education*, vol.5, no.21, pp.64-70, 2023.
[12] L. Xu, M. Mani et al., AI-enhanced literary education: Unveiling the potential of generative AI in literary education, *International Journal of Academic Research in Business and Social Sciences*, vol.15, no.1, pp.1780-1792, 2025.
[13] Q. Yu, An organic syntactic complexity measure for the Chinese language: The TC-unit, *Applied Linguistics*, vol.42, no.1, pp.60-92, 2021.
[14] L. I. Lied, Digitization and manuscripts as visual objects: Reflections from a media studies perspective, *Digital Biblical Studies*, 15, 2019.
[15] T. Gao, A. Fisch and D. Chen, Making pre-trained language models better few-shot learners, *arXiv Preprint*, arXiv: 2012.15723, 2020.
[16] X. Zhang, R. Mao, K. He and E. Cambria, Neuro-symbolic sentiment analysis with dynamic word sense disambiguation, *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp.8772-8783, 2023.
[17] F. Su and S. Zhao, Analysis of Chinese classical literature texts and prediction of Internet language trends based on natural language processing, *Journal of Computational Methods in Sciences and Engineering*, vol.25, no.2, pp.1246-1260, 2025.
[18] M. Terras, Artefacts and errors: Acknowledging issues of representation in the digital imaging of ancient texts, *Kodikologie und Paläographie im digitalen Zeitalter 2/Codicology and Palaeography in the Digital Age 2*, pp.43-61, 2011.
[19] J. Wan, H. Zhang, J. Zou, A. Zou, Y. Chen, Q. Zeng, X. Li and Q. Wang, WUMKG: A Chinese painting and calligraphy multimodal knowledge graph, *Heritage Science*, vol.12, no.1, 159, 2024.
[20] S. Chen, Y. Pei, Z. Ke and W. Silamu, Low-resource named entity recognition via the pre-training model, *Symmetry*, vol.13, no.5, 786, 2021.
[21] Y. Zhang, B. He, Y. Chen, H. Li, H. Yue, S. Zhang, H. Dou, J. Yan, Z. Liu, Y. Zhang et al., PhiloGPT: A philology-oriented large language model for ancient Chinese manuscripts with dunhuang as case study, *Proc. of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp.2784-2801, 2024.
[22] X. Xia, Exclamation points and elliptical hope: Reading wild grass against modern Chinese punctuation reform, *Modern Chinese Literature and Culture*, vol.32, no.1, pp.194-232, 2020.
[23] X. Guan and D. Qi, Research on the translation of official titles of Zizhi Tongjian with computer aid, *Journal of Physics: Conference Series*, vol.1992, 022004, 2021.
[24] X. Zheng, M. Li, Z. Wan and Y. Zhang, Knowledge mining and graph visualization of ancient Chinese scientific and technological documents bibliographic summaries based on digital humanities, *Library Hi Tech*, vol.42, no.6, pp.1693-1721, 2024.
[25] S. Chen, P.-L. Guhur, C. Schmid and I. Laptev, History aware multimodal transformer for vision-and-language navigation, *Advances in Neural Information Processing Systems*, vol.34, pp.5834-5847, 2021.

[26] T. Li, M. F. Cai and M. Cai, A review of mining-induced seismicity in China, *International Journal of Rock Mechanics and Mining Sciences*, vol.44, no.8, pp.1149-1171, 2007.
[27] A. Jafari, G. Fox, J. B. Rundle, A. Donnellan and L. G. Ludwig, Time series foundation models and deep learning architectures for earthquake temporal and spatial nowcasting, *GeoHazards*, vol.5, no.4, pp.1247-1274, 2024.
[28] Q. Zhang, *An Introduction to Chinese History and Culture*, Springer, 2015.
[29] X. Tang, Z. Deng, Q. Su, H. Yang and J. Wang, CHisIEC: An information extraction corpus for ancient Chinese history, *arXiv Preprint*, arXiv: 2403.15088, 2024.
[30] H. Mu, P. Wu and W. Su, Construction of knowledge graph for emergency resources, *International Journal of Intelligent Systems*, vol.2024, no.1, 6668559, 2024.
[31] W. Dong, *Relief Treatise for Famine and the People: With Supplement*, Reprint Edtion, Commercial Press, 1936.
[32] I. Liritzis, A. Westra and C. Miao, Disaster geoarchaeology and natural cataclysms in world cultural evolution: An overview, *Journal of Coastal Research*, vol.35, no.6, pp.1307-1330, 2019.
[33] A. Mohammed and R. Kora, A comprehensive overview and analysis of large language models: Trends and challenges, *IEEE Access*, 2025.
[34] D. Mekhaldi, Multimodal document alignment: Towards a fully-indexed multimedia archive, *Multimedia Information Retrieval*, 52, 2007.
[35] Y. Guo, J. H. Sohn, G. Leroy and T. Cohen, Are LLM-generated plain language summaries truly understandable? A large-scale crowdsourced evaluation, *arXiv Preprint*, arXiv: 2505.10409, 2025.
[36] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv et al., Qwen3 technical report, *arXiv Preprint*, arXiv: 2505.09388, 2025.
[37] G. Dong, H. Yuan, K. Lu, C. Li, M. Xue, D. Liu, W. Wang, Z. Yuan, C. Zhou and J. Zhou, How abilities in large language models are affected by supervised fine-tuning data composition, *arXiv Preprint*, arXiv: 2310.05492, 2023.

Author Biography

Xiaolong Wei obtained his Ph.D. degree in Information and Computing Technology from the Communication University of China, China, in 2024. He is presently a Lecturer at the Institute of Disaster Prevention, China. His primary research areas include disaster information monitoring and the processing of large-scale data.

Qingjie Liu obtained his Master's degree in Software Engineering from Beijing University of Technology, China, in 2021 and Bachelor's degree in Computer and Applications from Zhengzhou University, China, in 2001. He is presently a Professor at the Institute of Disaster Prevention, China. His primary research areas include disaster prevention information monitoring and big data processing.

Zhian Pan obtained his Master's degree in Computer Software and Theory from Taiyuan University of Technology, China, in 2010. He is presently a Professor, Master's Supervisor and Director of the Network Engineering Teaching and Research Section at the School of Computer Science and Engineering, Institute of Disaster Prevention, China. His primary research areas include big data processing, deep learning and network engineering.