

Original Paper

# An Entity-Based Visual Analytics System Enhancing Medical Expertise Acquisition: Development and Verification Study

Xiao Pang<sup>1</sup>, MS; Chang Liu<sup>1</sup>, MD; Yan Huang<sup>1</sup>, MS; MingYou Liu<sup>2\*</sup>, MD; Jiyuan Liu<sup>1\*</sup>, MD

<sup>1</sup>Department of Information Management, State Key Laboratory of Oral Diseases & National Center for Stomatology & National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu, China

<sup>2</sup>Department of Information and Network Security, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China

\*these authors contributed equally

**Corresponding Author:**

Jiyuan Liu, MD

Department of Information Management

State Key Laboratory of Oral Diseases & National Center for Stomatology & National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University

No. 14, 3rd Section of Ren Min Nan Rd.

Chengdu 610041

China

Phone: 86 13880737189

Email: [wchsljy@scu.edu.cn](mailto:wchsljy@scu.edu.cn)

## Abstract

**Background:** Acquiring medical expertise from the vast body of medical text is a critical component of medical education. However, the majority of medical knowledge resides in unstructured texts. Data heterogeneity across institutions and strict privacy regulations hinder the use of general-purpose analysis tools. This creates a substantial barrier to the efficient acquisition of expertise for learners.

**Objective:** This study aimed to design, develop, and evaluate MExplore, an interactive visual analytics system to facilitate the acquisition of medical expertise from unstructured clinical texts.

**Methods:** We propose a localized, cost-effective workflow for the automatic extraction of medical entities. Building on this workflow, MExplore provides a multilevel visual framework featuring coordinated views for progressive, entity-centered exploration. The system was evaluated through case studies conducted with domain experts, a user study, and semistructured interviews.

**Results:** The evaluation demonstrated that MExplore significantly enhances the medical expertise acquisition process. Our findings confirm that MExplore provides an effective and interactive approach for structuring complex knowledge, facilitating the construction of illness scripts, and strengthening knowledge retention.

**Conclusions:** MExplore provides an intuitive and powerful approach for acquiring medical expertise. The results suggest that it effectively supports medical learners in conducting in-depth data exploration and developing robust clinical reasoning skills.

*JMIR Med Inform* 2026;14:e83785; doi: [10.2196/83785](https://doi.org/10.2196/83785)

**Keywords:** medical text; medical expertise; illness script; visualization; visual analytics

## Introduction

The rapid expansion of medical education has intensified the demand for efficient expertise acquisition [1,2]. Medical expertise is characterized by a highly organized and differentiated knowledge base, technical skills, and perceptual capabilities acquired through extensive domain-related practice [3]. The acquisition of medical expertise is a complex cognitive process that involves constructing abstract

knowledge networks and iteratively refining them into narrative structures known as “illness scripts” [4]—cognitive frameworks that link clinical features to diagnoses and enable clinicians to perform complex reasoning tasks efficiently [5]. Traditional learning relies heavily on textbooks and often overlooks real-world medical documents (MDs) such as electronic medical records. Although these documents provide rich, detailed cases for refining illness scripts, their

unstructured nature and volume make it difficult for novices to extract and synthesize information.

While large language models (LLMs) have emerged as tools for information synthesis, their utility is limited by practical barriers, such as privacy regulations and computational costs [6]. Furthermore, their direct application in medical education is fraught with risks. Research indicates that LLMs are prone to “deceptive expertise,” generating plausible yet factually incorrect information while failing to recognize their own limitations [7,8]. Overreliance may lead to cognitive offloading, impairing learners’ ability to critically evaluate artificial intelligence-generated outputs [9-12]—a particularly concerning issue in medical education, where independent clinical reasoning is essential. Consequently, there is a pressing need for frameworks that support exploratory learning and active engagement with data, rather than systems that encourage the passive acceptance of output [9,13].

Visual analytics, which integrates human expertise with machine intelligence in a “human-in-the-loop” paradigm, offers a viable approach for such exploratory frameworks. By visualizing complex data, these systems can facilitate the construction of mental models. However, the application of visual analytics to medical expertise acquisition remains underdeveloped. Much of the existing work in this domain has focused on analyzing individual patient electronic medical records to support clinical decision-making for specific cases [14,15], which does not facilitate the higher-level organization of generalizable medical knowledge required for education. Other studies have visualized health care data but often neglect the underlying semantics of the content [16,17], limiting the user’s ability to understand deep clinical relationships. Therefore, a gap exists for a system that can aggregate unstructured text and present it in a semantically meaningful, multilevel structure suitable for learning.

Named entity recognition (NER) serves as a key technique for transforming unstructured text into structured representations. It involves identifying and classifying domain-specific concepts, such as symptoms, diseases, and treatments, within a text [18,19]. While recent advancements in deep

learning and transformer-based architectures such as BERT (Bidirectional Encoder Representations From Transformers) have significantly improved the accuracy of extracting these medical entities (MEs) [20-23], the potential of using extracted entities as “anchors” to construct visual knowledge networks remains largely unexplored. Integrating NER with visual analytics offers a novel pathway to transform raw clinical text into structured learning pathways, thereby accelerating the construction of illness scripts.

We present MExplore, a system that leverages NER to extract MEs from unstructured text and organizes them into a coherent visual structure based on a multilevel metaphor. This design allows learners to explore medical knowledge at varying granularities—including MDs, medical paragraphs (MPs), medical entity sets (MESs), and individual MEs. This approach actively engages users in reasoning, reducing the passive learning common in chat-based interfaces [24, 25], and reinforcing the cognitive links necessary for robust illness scripts. We validate the system through a multifaceted evaluation involving domain expert case studies, a controlled user study with medical students, and expert interviews to assess clinical validity and user acceptance [26].

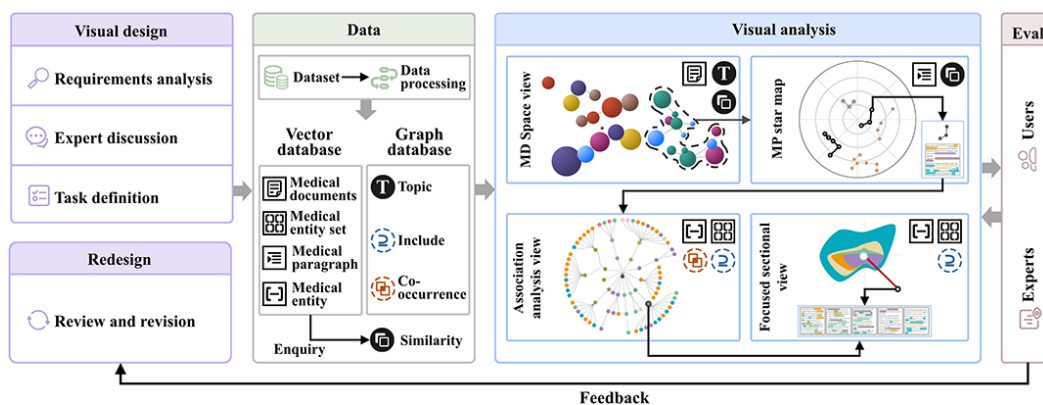
## Methods

### Study Design

#### Design Methodology and Workflow

We used a user-centered design process [27], as illustrated in Figure 1. Initially, we collaborated with domain experts to analyze requirements and define specific visual analytics tasks. Guided by these tasks, we established the hierarchy, structural organization, and computational metrics for the textual data. We subsequently designed and developed a multilevel visual analytics framework tailored to these data specifications. Throughout the development lifecycle, we engaged in close collaboration with experts and users, conducting multiple rounds of testing and validation. The feedback and recommendations from these sessions drove iterative redesigns, ensuring continuous system optimization.

**Figure 1.** The MExplore workflow encompasses visual design, data processing, storage, and query handling, culminating in multilevel visual analysis. The process incorporates iterative redesigns driven by user and expert feedback. MD: medical document; MP: medical paragraph.



## Requirements Analysis

### *Requirement Elicitation*

We collaborated closely with 10 domain experts (E1-E10). E1 and E2 are PhD candidates specializing in medical studies, with 4 and 5 years of focused research experience; E3 is a professor at a medical college with 8 years of teaching experience; E4 is a medical researcher with a decade of research expertise; E5 is a dentist with 10 years of professional practice; E6-E8 are experienced physicians with 5-10 years of clinical practice; and E9 and E10 are visualization researchers, with 5-10 years of experience in visual analytics.

Through iterative discussions, the experts (specifically E3-E8) emphasized that acquiring expertise involves refining knowledge into “illness scripts” [4]. These scripts enable clinicians to integrate disease-relevant information and enhance both recall and application [28,29]. To support this, we identified 4 key requirements.

### *R1: Extraction of Core Knowledge Units From Complex Texts*

Unstructured medical texts create a high cognitive load for novice learners due to their density and complexity [30]. Novices often face a “reading bottleneck” [31] where resources are consumed by decoding information rather than deep comprehension. Therefore, it is essential to extract core knowledge units and filter out redundant information to mitigate extraneous cognitive load [32,33].

### *R2: Organization of Text With Varying Knowledge Density to Support Gradual Exploration*

Presenting excessive details simultaneously causes cognitive overload [32,34]. Experts suggest that information should be presented at appropriate levels of abstraction [35]. Medical texts must be restructured with varying knowledge densities to support an incremental learning process, moving from high-level overviews to granular details [36].

### *R3: Revealing the Interconnections Between Knowledge Units*

Rote memorization is insufficient for long-term retention [37]. Learners must actively integrate new information with existing schemas [38]. Consequently, there is a need to visually reveal the logical interconnections and inclusion relationships between knowledge units to support structural understanding.

### *R4: Focused Analysis of the Key Knowledge Units*

Effective knowledge acquisition relies on identifying and analyzing key knowledge units that serve as cognitive “anchors” [39,40]. Focused analysis of these units not only deepens understanding but also facilitates the establishment of broader connections within the medical field [40], thereby reinforcing long-term retention.

## Visual Analytics Tasks

Based on the requirements (R1-R4), we derived 4 visual analytics tasks (T1-T4) to guide the system implementation.

### *T1: Extract MEs From Real-World Medical Texts*

The system must automatically extract MEs from raw medical texts and categorize them. This transforms unstructured data into discrete, manageable knowledge units (R1), which serve as the foundational elements for visualization.

### *T2: Support for Cascading Visual Analysis and Exploration of Texts With Varying Knowledge Densities*

The system should organize extracted data into a hierarchical structure. It must provide cascading views that allow users to transition seamlessly from a high-level thematic overview to coarse-grained paragraph relations, and finally to fine-grained entity details. This capability facilitates incremental exploration and understanding of complex medical knowledge (R2).

### *T3: Establishing a Clear Structure for Association Analysis*

The system needs to construct and visualize a clear mapping of relationships, specifically the inclusion of MEs within paragraphs and the co-occurrence of elements. This visualization should clarify how knowledge units are structurally connected (R3).

### *T4: Support for Entity-Centric Pattern Analysis*

The system must provide interaction mechanisms to select a specific ME or MES as a focal point. Upon selection, the visualization should be dynamically reconfigured to display the distribution, composition, and associated context centered around the target knowledge unit (R4).

## Dataset Construction

### Data Source

In this paper, we analyze and process 2 realistic datasets. The first dataset is the Chinese Biomedical Language Understanding Evaluation (CBLUE) provided by the Key Laboratory of Computational Linguistics (Peking University, China) [41]. CBLUE is a comprehensive benchmark compiled from authoritative medical textbooks and clinical practice records. The dataset contains approximately 96,000 MDs, covering a wide range of more than 500 common diseases. To ensure high quality, the MEs within the dataset were meticulously annotated by medical experts. These MEs are categorized into 9 classes on the basis of Chinese ME annotation standards [41]: disease (dis), clinical symptoms (sym), drugs (dru), medical equipment (equ), medical procedures (pro), body (bod), medical examination items (ite), microorganisms (mic), and department (dep).

The second dataset, the Medical Text of West China School (Hospital) of Stomatology (MWCSS), was collected between 2023 and 2025 and provided by the West

China School (Hospital) of Stomatology (Sichuan University, China). This dataset contains a large volume of unstructured text data from clinical practice. It includes approximately 100,000 MDs, covering various facets of medical processes, such as clinical diagnoses, treatment processes, specialist examinations, auxiliary examinations, treatment plans, interventions, and drug instructions.

## ME Extraction Comparative Analysis

The accuracy of ME extraction is critical to the effectiveness of data processing and visual analytics. Therefore, we systematically assessed the performance and computational resource demands of several models for ME extraction.

Due to strict data confidentiality protocols, which preclude the transmission of data to external servers, models must be deployable locally. In light of the practical limitations of resource-constrained environments, and based on the latest benchmark list from FlagEval [42], we selected high-performing, open-source LLMs for evaluation:

DeepSeek-70B and Qwen-32B. Additionally, we included MacBERT, an enhanced BERT model with a novel masked language modeling correction pretraining task [43], which has been identified as the top-performing BERT-based model according to the CBLUE benchmark, for comparison.

Fine-tuning and inference were performed via two NVIDIA A10 GPUs or a single NVIDIA RTX 4060 Ti GPU for MacBERT. The fine-tuning parameters were as follows: learning rate of  $3e-5$ , 2 epochs, batch size of 16, and warm-up ratio of 0.1. For DeepSeek-70B and Qwen-32B, both models required two A10 GPUs and were instruction-tuned via 4-bit quantized models with QLoRA (Quantized Low-Rank Adaptation) [44] (LoRA [Low-Rank Adaptation] settings:  $r=16$ ,  $\alpha=64$ , dropout=0.05). All the fine-tuning parameters and LLM prompt words were selected based on benchmark tests [41] and relevant comparative experiments [45] and have been validated. The computational resource consumption for each model is summarized in Table 1.

**Table 1.** Computational resource consumption comparison.

Model	GPU configuration (NVIDIA)	Fine-tuning time (hours)	Inference time (hours)
DeepSeek-70B	2 A10	167.23	134.27
Qwen-32B	2 A10	101.46	83.58
MacBERT	2 A10	0.77	0.18
MacBERT	1 RTX 4060 Ti	5.16	0.35

The results highlight a significant disparity in resource demand. While LLMs (DeepSeek-70B and Qwen-32B) require substantial fine-tuning and inference time on dual A10 GPUs, MacBERT demonstrates remarkable efficiency, completing the same tasks in a fraction of the time. Notably, MacBERT's inference speed on consumer-grade hardware (a single RTX 4060 Ti) suggests that it is highly suitable for local deployment in practical medical environments where high-end computational clusters may not be available.

Table 2 summarizes the performance metrics for each model. Fine-tuning consistently improved performance across all the models. Notably, the fine-tuned MacBERT model achieves the highest  $F_1$ -score (0.605), surpassing both LLMs, and its precision approaches human annotation levels [41].

**Table 2.** Performance comparison of models for medical entity extraction.

Model	Type	Precision	Recall	$F_1$ -score
DeepSeek-70B	Base	0.548	0.487	0.516
DeepSeek-70B	Fine-tuned	0.593	0.521	0.535
Qwen-32B	Base	0.554	0.491	0.525
Qwen-32B	Fine-tuned	0.608	0.542	0.561
MacBERT	Base	0.595	0.549	0.571
MacBERT	Fine-tuned	0.636	0.579	0.605

This analysis demonstrates that MacBERT not only delivers the best performance after fine-tuning but also requires significantly fewer computational resources than the LLMs. MacBERT, therefore, emerges as the optimal choice for ME extraction tasks. This conclusion aligns with the literature [46], which emphasizes that while LLMs excel in generative tasks, BERT-based models retain distinct advantages in NER and other specialized domains [47] and sentiment analysis [48].

Beyond common medical conditions, the system's robustness concerning rare diseases and less common

terminology warrants further discussion. Although our evaluation used a dataset of common diseases—where MacBERT demonstrated strong performance—recent research [49] suggests that the performance landscape may shift when encountering “long-tail” medical data. Specifically, while fine-tuned BERT-based models generally maintain overall superiority, LLMs can exhibit superior capabilities in identifying rare disease entities in few-shot settings. Therefore, for specialized tasks focusing on rare diseases where relevant samples and sufficient computational resources are available, LLMs may be used to achieve higher

precision. Nevertheless, within our current visual analytics framework centered on general-purpose unstructured medical text, MacBERT provides the most reliable performance-to-cost ratio.

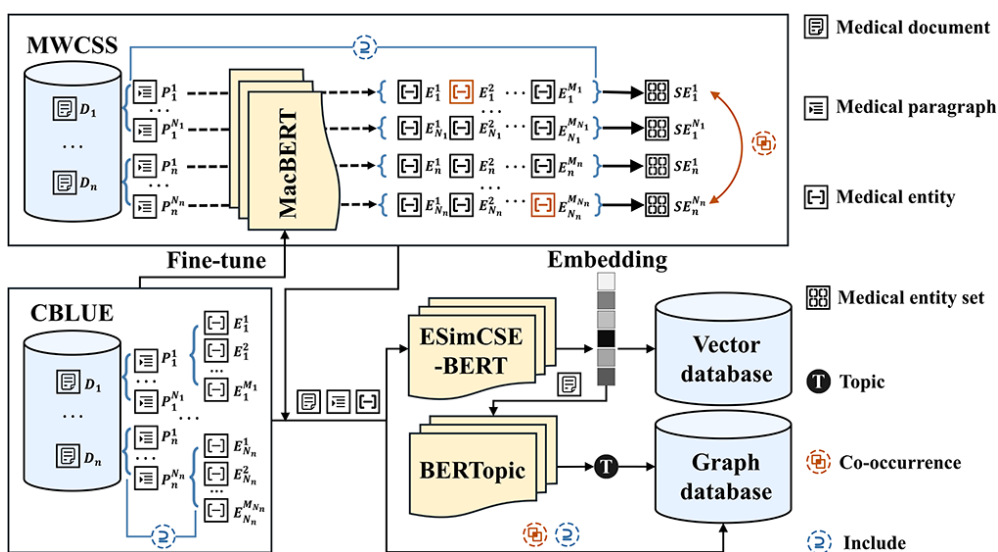
For NER requiring deep semantic insight, language-specific pretrained models often outperform general multi-lingual models [50]. In this study, we use MacBERT—the top-performing model on the CBLUE dataset—to maximize extraction accuracy for Chinese corpora. Notably,

the underlying pipeline remains highly adaptable; the base encoder can be seamlessly replaced (eg, substituting BioBERT [51] for MacBERT) to process English or other language datasets.

### Data Processing

Figure 2 shows the pipeline of data processing. The process can be described as follows.

**Figure 2.** Pipeline of data processing. First, MacBERT was fine-tuned using labeled data from CBLUE. Individual MPs within the MDs are processed via the fine-tuned MacBERT model to extract MEs, which are subsequently organized into MESs. These texts are input into ESimCSE-BERT to generate embeddings, which are then stored in a vector database. The MD embeddings serve as the basis for topic clustering in BERTopic. The resulting topics of MDs, along with all medical texts and their inclusion and co-occurrence relationships, are then structured into a graph and stored in a graph database. CBLUE: Chinese Biomedical Language Understanding Evaluation; *D*: medical document; *E*: medical entity; MD: medical document; ME: medical entity; MES: medical entity set; MP: medical paragraph; MWCSS: Medical Text of West China School (Hospital) of Stomatology; *P*: medical paragraph; *SE*: medical entity set.



### Step 1: ME Extraction and Structuring

According to the results of the ME Extraction comparative analysis, we used the MacBERT model fine-tuned with the CBLUE dataset to extract MEs from MWCSS, using MP of MD *D<sub>i</sub>* as the extraction unit, thereby constructing MESs.

### Step 2: Vectorization

All the text from CBLUE and MWCSS, including MDs, MPs, MESs, and MEs, is processed via ESimCSE-BERT [52]—an efficient and improved unsupervised sentence embedding method that generates high-quality sentence vectors.

### Step 3: Topic Clustering

The high-quality MD vectors from step 2 were input into the BERTopic model. By leveraging these superior embeddings,

the clustering process produces topic assignments for each MD. The results yielded a topic diversity score [53] of 0.975, indicating a high degree of distinctness among the generated categories. To further validate the clinical utility, a 5-point Likert scale assessment was conducted by medical experts (E3, E4, E6, E7, and E8) to evaluate the effectiveness of the generated topics and keywords (Table 3). Following the evaluation framework proposed in [54], the results suggest that the interpretability, distinctiveness, and relevance of the topic clusters are at a high level. These findings further corroborate the superiority of transformer-based topic models as discussed in [54].

**Table 3.** Evaluation of topic clustering (1-5 Likert scale).

Evaluation	Explanation	Score, mean (SD)
Interpretability	How easily a human can assign a coherent meaning to the topic.	4.3 (0.57)
Distinctiveness	How well the topic is differentiated from other topics.	4.4 (0.41)
Relevance	How well the topic captures meaningful domain content.	4.6 (0.22)

### Step 4: Relationship Mapping and Storage

To support complex analytical queries, we implemented a dual-storage strategy addressing the inherent relational complexity and interconnectedness of medical data.

#### Semantic Storage in the Vector Database

All textual elements (MDs, MPs, MESs, and MEs) are embedded into high-dimensional vectors and stored in a vector database. This infrastructure leverages various optimization algorithms to enable the efficient execution of similarity-based queries and fast calculations across datasets in the tens of millions [55].

#### Structural Storage in the Graph Database

Each text unit is modeled as a vertex, with MD vertices carrying the topic attributes derived from clustering. The inclusion relationships between texts and the co-occurrence of MEs within MESs are represented as edges. This graph structure is stored in a graph database to facilitate high-performance queries across complex entity relationships [56]. The scalability and query efficiency of this graph-based approach have become increasingly evident with larger datasets [57].

### Visual Analytics Systems Development

#### System Overview and Workflow

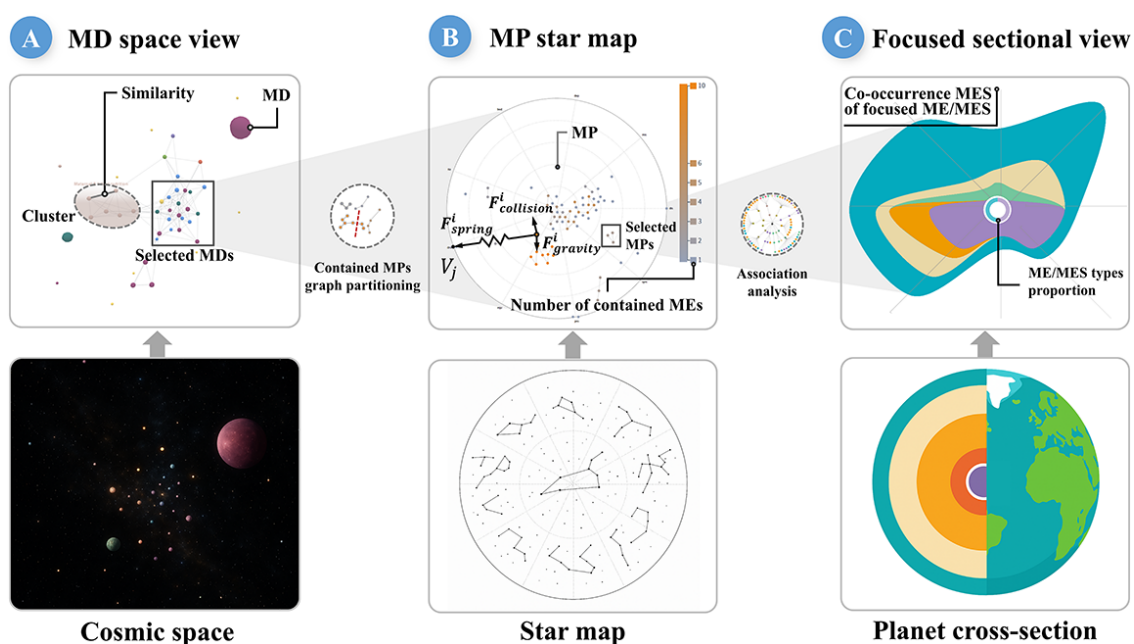
Building upon the constructed dataset, we developed MExplore, a multilevel visual analytics system designed to facilitate progressive exploration and cascading visual

analysis (T2). The system adopts a 3-tiered metaphor—cosmic space, star map, and planet cross-section—which inspires the design of the MD space view (Figure 3A), the MP star map (Figure 3B), and the focused sectional view (Figure 3C). We use a hierarchical layout that systematically guides users from macrolevel document retrieval to microlevel detail exploration.

The analytical workflow follows a top-down trajectory (Figure 3), starting from the MD space view, where users filter and define a subset of MDs as the basis for all subsequent analysis based on cluster and spatial distributions. Once the analytical scope is established, the system transitions to the MP star map, enabling users to identify structural patterns and semantic clusters within MP subgraphs. From there, users can incorporate specific subgraphs into the association analysis view to reveal intricate relationships between MESs and MEs. For the most granular inquiry, users can select individual nodes within this view to generate a corresponding focused sectional view, facilitating a deep dive into the target knowledge unit’s context.

By structuring the analysis as an interactive, stepwise process, MExplore allows users to progressively increase analytical granularity without losing global contextual coherence [58]. Furthermore, the design leverages grounded metaphors to map complex information onto pre-existing cognitive schemas [59], effectively mitigating the visual complexity encountered by the user and enhancing memory encoding for knowledge retention [60]. Ultimately, this multitiered approach facilitates a continuous analytical loop.

**Figure 3.** The MExplore framework: (A) the MD space view, inspired by cosmic space, where users select MDs and construct and partition graphs of the contained MPs; (B) the MP Star map, inspired by the star map, where users choose corresponding MP subgraphs and perform relational analysis; (C) the focused sectional view, inspired by the planet cross-section, which enables detailed, focused analysis of the selected data.  $F_{collision}$ : the collision force;  $F_{gravity}$ : the gravity force;  $F_{spring}$ : the spring force; MD: medical document; ME: medical entity; MES: medical entity set; MP: medical paragraph.



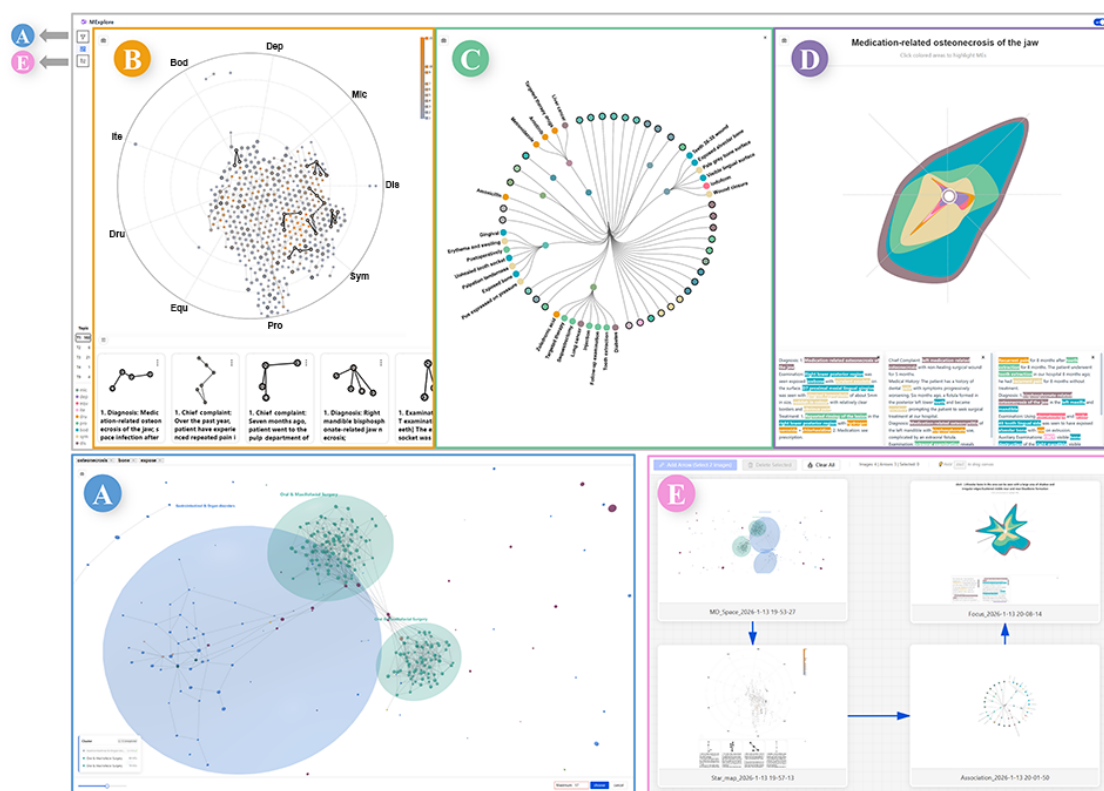
### MD Space View

As shown in Figure 4A, upon entering keywords, the system executes a full-text vector search to retrieve relevant MDs. The retrieved MDs are mapped into a 3D force-directed graph where each MD is represented as a planet—its size encodes text length, and its position is determined by semantic similarity. Interdocument links visualize this similarity. Users can adjust a threshold slider to filter connections by similarity, aggregating MDs into topic-based clusters represented as nebulae labeled with topic titles. This layout emphasizes core structural relationships and allows users to follow a progressive approach starting from these nebulae, clicking a specific nebula to enter Focus Mode to explore individual MDs in detail. This interaction mechanism leverages spatial proximity to facilitate the efficient identification of areas of interest [61].

The colors of the nebulae and their constituent planets represent their respective topics. This color encoding is discarded upon transitioning to detailed analysis views, preventing semantic interference between the MD topic and ME attributes.

To facilitate comprehensive discovery, a guide for unexplored nebulae is displayed in the lower-left corner, allowing for thorough exploration and rapid focusing by clicking. Simultaneously, the panel below ranks MDs by vector similarity scores to support instant focused viewing and prevent users from overlooking important MDs. Throughout this process, users iteratively select MDs to build an analytical collection, which serves as the foundation for all subsequent analysis views.

**Figure 4.** The MExplore system: (A) MD space view; (B) MP star map; (C) association analysis view; (D) focused sectional view; (E) provenance view. MD: medical document; MP: medical paragraph.

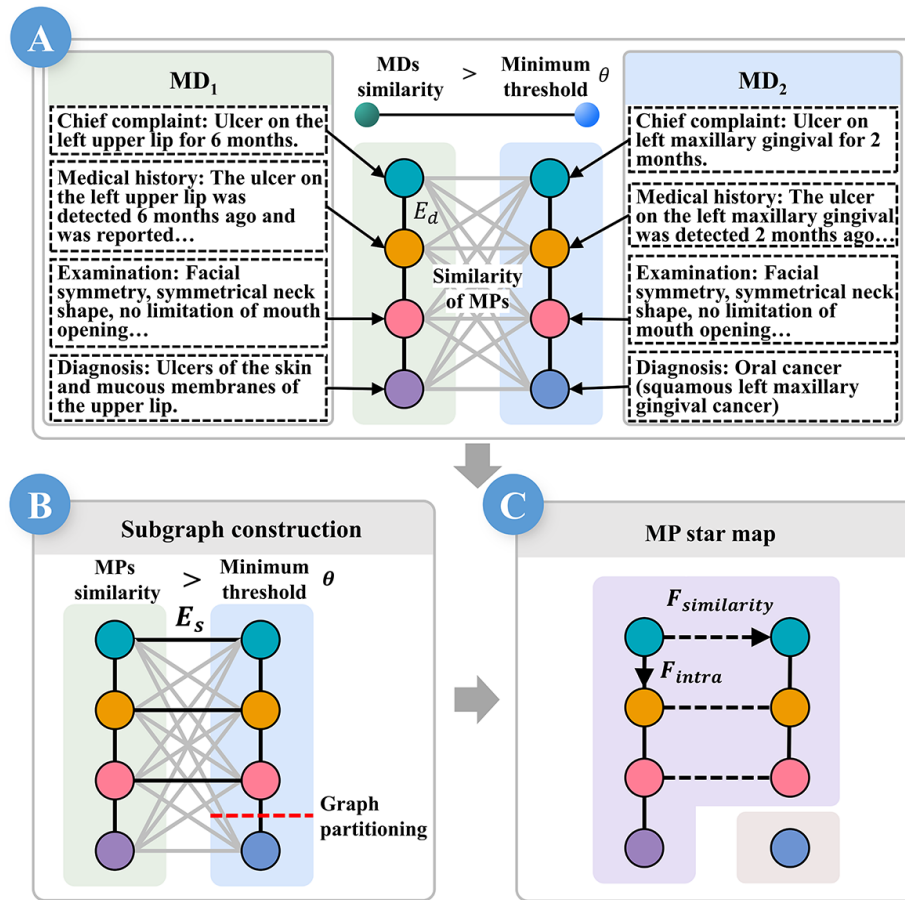


### MP Star Map

Relying solely on full-text embedding similarity can introduce bias, as document representations may disproportionately reflect longer sections while overshadowing concise but critical information. For instance, detailed medical histories in outpatient records can overshadow concise diagnosis sections during similarity computation.

To mitigate this, MDs are decomposed into MPs, each modeled as a graph vertex. Intradocument edges  $E_d$  connect the MPs within the same MD. Cross-document MP pairs (Figure 5A) exceeding the similarity threshold set in Figure 4A are linked by similarity edges  $E_s$ . The KaFFPa algorithm [62] then partitions the graph, segregating semantically distinct MPs into subgraphs (Figure 5B).

**Figure 5.** For detailed analysis, MDs are decomposed into MPs. (A) MPs are represented as vertices, with  $E_d$  connecting MPs within the same MD, and the similarity of MPs across connected MDs is computed. (B) MPs whose similarity is greater  $\theta$  than are connected by  $E_s$ , and the graph is partitioned by KaFFPa. (C) The partitioned subgraphs are visualized via the constellation metaphor, where  $E_d$  are shown as lines and the MPs are shown as stars. Each star is subject to  $F_{intra}$  from  $E_d$  and  $F_{similarity}$  from  $E_s$ . MD: medical document; MP: medical paragraph.



Each resulting subgraph is visualized as a constellation in the MP star map (Figure 4B), the MPs are depicted as stars, and  $E_d$  is rendered as a constellation line, whereas  $E_s$  is omitted to reduce clutter but is retained as latent factors influencing the layout, the similarity-based gravitational force  $F_{similarity}$ , and the structural gravitational force  $F_{intra}$  (Figure 4C). The combined gravitational force of the star  $i$  is calculated as follows:

$$F_{gravity}^i = F_{intra}^i + F_{similarity}^i \quad (1)$$

$$F_{similarity}^i = \sum_{(i,j)}^{similarity_i^j > \theta} (\|F\| \cdot similarity_i^j) \cdot \hat{d}_i^j \quad (2)$$

where  $F_{intra}$  is the link force connecting the MPs within the same MD, with a magnitude of unit force  $\|F\|$ ,  $\theta$  is the threshold set by the user, and  $similarity_i^j$  is the similarity between MP  $i$  and MP  $j$ .  $\hat{d}_i^j$  is the unit vector of the force direction from MP  $i$  to MP  $j$ .

In addition, each star  $i$  is subjected to spring forces  $F_{spring}^i$  from 9 ME-type poles uniformly distributed on the circular boundary of the star map:

$$F_{spring}^i = \sum_{j=1}^9 \frac{num_i^j}{num_i} \cdot \|F\| \cdot \hat{d}_i^j \quad (3)$$

where  $num_i^j$  is the number of MEs of type  $j$  in MP  $i$ , and  $num_i$  is the total number of MEs in MP  $i$ .  $\hat{d}_i^j$  is the unit vector of the force direction from MP  $i$  to  $V_j$ .

To prevent occlusion, each star  $i$  is subject to the collision force  $F_{collision}^i$  of stars that may overlap it. The final combined force  $F_{collision}^i$  is calculated as follows:

$$F_{combined}^i = F_{gravity}^i + F_{spring}^i + F_{collision}^i \quad (4)$$

This multiforce design achieved semantically coherent clustering ( $F_{spring}$ ), representation of cross-document similarity and structural relationships ( $F_{gravity}$ ), and clear visualization ( $F_{collision}$ ) (Figure 4B). Stars within the same constellation are assigned a uniform color, allowing users to quickly identify them even on a densely populated map. This color is modulated by a sequential luminance gradient [63] to encode the star count: brighter constellations denote a greater number of constituent stars, enabling users to efficiently pinpoint pattern-dense regions. The luminance of star borders

encodes the ME count, providing information density cues (T2). To preserve macrocontext awareness during the analysis of MPs and subsequent tasks, a topic navigation list at the bottom of the sidebar (Figure 4) visualizes the topic information of selected MDs. In the MP star map and subsequent views, users can click on MP/MES nodes or document cards to highlight the topic of the corresponding document. This mechanism enables on-demand backtracking, ensuring users remain oriented within the macrolevel topic context. This design facilitates progressive schema construction [33,34] and supports iterative exploratory learning.

## Association Analysis View

Leveraging the concepts within MPs to create visual representations of hierarchical structures can significantly

increase the efficiency of knowledge navigation [64,65]. Therefore, we extract MEs from each MP (T1), compose the MES, and visualize the resulting data via a radial dendrogram. Compared with alternative text visualization methods, such as word clouds or Sankey diagrams, this approach effectively captures both inclusion and relational connections among elements while accurately displaying hierarchical structures. To further enhance visualization, we propose an algorithm that groups MESs containing shared MEs into the same tree branch, thereby emphasizing co-occurrence relationships between the MESs (T3). The algorithm is outlined in Textbox 1.

### Textbox 1. Algorithm for tree construction.

#### Input:

1. Root node of the tree:  $r_{tree}$
2. Set of nodes to be added:  $N_{add}$

#### Output: Root node of the tree: $r_{tree}$

1. **for** each node  $n_i$  in  $N_{add}$  **do**
2.  $C = GetMEs(n_i)$
3.  $N, M_{checked} = TraverseTree(r_{tree}, C)$
4. **end for**
5.  $N_{children} \leftarrow N \cup M_{checked}$
6. **for** each node  $n_i$  in  $N$  **do**
7.  $RemoveFromTree(r_{tree}, n_i)$
8. **end for**
9. **if**  $L_N > 1$  **then**
10.  $n_f \leftarrow CommonFatherNode(r_{tree}, N)$
11. **if**  $n_f$  is not (*null or  $r_{tree}$* ) **then**
12.  $n_f \cdot Children \leftarrow N_{children}$
13. **else**
14.  $r_{tree} \cdot children \leftarrow r_{tree} \cdot Children \cup N_{children}$
15. **end if**
16. **end if**
17. **return**  $r_{tree}$

For each MP, the extracted MEs form an MES, which acts as a node. The newly added node set  $N_{add}$ , along with the root node of the tree  $r_{tree}$ , is provided as input. For each node  $n_i$  in  $N_{add}$ , the algorithm retrieves its associated MEs via  $GetMEs(n_i)$  function. It then traverses the tree through  $TraverseTree$  to identify the set  $M_{checked}$  of co-occurring MEs, along with the node set  $N$  containing these MEs. They are all grouped into the set  $N_{children}$ , which are subsequently added as children to the appropriate parent nodes.

If multiple nodes intersect, their common ancestor is identified by  $CommonFatherNode$ , and the new nodes are added as children of this ancestor. The nodes are added directly to the root if no common ancestor is found. The process concludes by returning the updated tree structure.

Based on the algorithm described above, the association analysis view is constructed (Figure 4C). To manage the

cognitive load and ensure visual consistency, the system uses a synchronized color encoding strategy and a progressive disclosure mechanism. Categorical colors are assigned to ME nodes based on their entity types, while for MES nodes, the color is derived from a weighted mixture of its constituent ME types. To maintain cross-view coherence, this encoding scheme is applied uniformly across both the association analysis and focused sectional views, supported by a persistent legend in the sidebar that ensures color-to-type mappings remain readily accessible throughout the analytical process.

The risk of information overload is further mitigated by initially displaying only the aggregated MES nodes, allowing users to perceive the high-level relational structure without being overwhelmed by individual entities. Users can then interactively expand specific MES nodes to reveal their

constituent MEs as independent child nodes. This progressive disclosure strategy enables users to manage information density dynamically according to their analytical needs.

These design choices facilitate the perception of complex hierarchical and relational patterns, thereby supporting schema construction and clinical reasoning by externalizing the learner's cognitive processes [66,67]. Furthermore, the system architecture supports the dynamic addition of new MPs, enabling real-time reconstruction of the association tree. This capability allows seamless integration of new clinical information into the existing cognitive framework of the user, promoting iterative, exploratory learning and long-term memory retrieval [68].

### Focused Sectional View

To enable granular analysis of a specific ME or MES, the focused sectional view (Figure 4D) uses a "planetary layer" metaphor to organize complex co-occurrence data. The central ring functions as a donut chart that encodes the proportional distribution of ME classifications. If the analytical focus is a single ME, this layer identifies its specific classification. Surrounding this core, the mantle layer uses a polar-axis-aligned area chart to visualize co-occurring MESs. In this arrangement, each radial axis represents a distinct MES, with the axis height being  $N_i^k * h_t$ , which represents the number of MEs belonging to class  $k$  in the  $i$ th MES. To emphasize semantic continuity across the dataset, areas representing the same ME classification are connected across adjacent axes. Compared with traditional area or stream charts, this radial layout effectively portrays the compositional richness and semantic associations of the target entity within a compact structural signature.

To bridge the gap between visualization and raw text, the system allows users to click a radial axis to query the underlying database and retrieve the corresponding MD, which is then displayed as a document card in the panel below. This design enables users to revert to the original evidence for detailed, word-by-word analysis. To further prevent highlighting overload in the text view, the system implements a selective filtering strategy. While the card highlights various extracted MEs, users can click specific semantic regions in the polar area chart to isolate and highlight only the ME category of interest. This interaction triggers the document card to automatically scroll to the relevant text segment, enabling an efficient and focused review of the original evidence without cognitive distraction.

The resulting structural signatures illustrate the contextual association patterns unique to each focused ME/MES (T4). Furthermore, the mantle's radial patterns facilitate the interactive comparison of MESs with similar ME configurations. This comparative analysis fosters deeper cognitive engagement and more enduring conceptual impressions, supporting the learner's ability to internalize complex clinical relationships [67].

### Provenance View

To facilitate knowledge retention and structure the analytical process, we developed the provenance view (Figure 4E). This view enables users to capture high-fidelity snapshots from any analytical view on demand. These snapshots function as interactive nodes on a scalable, zoomable canvas, allowing users to freely organize layouts and establish logical links between visual findings. By externalizing exploration history into an explicit visual learning path, users can revisit and inspect detailed analytical states at any time, effectively transforming ephemeral interactions into organized knowledge assets.

### Evaluation Design

#### Evaluation Overview

To comprehensively evaluate the usability, effectiveness, and expertise acquisition capabilities of MExplore, we used a mixed methods evaluation strategy comprising three components: (1) case studies with domain experts to assess the practicality, (2) a comparative user study with medical students to quantify learning outcomes, and (3) semistructured interviews with experts to evaluate usability and future potential of the system.

#### Implementation and Apparatus

As a web-based visual analysis system, MExplore was developed using the d3.js and Django frameworks. A Microsoft Windows platform with a 2.71 GHz Intel Core i5-7300 CPU and 8 GB of memory was used as the front-end page server. The evaluation experiments were performed via a Google Chrome web browser.

#### Case Study

The case study can demonstrate feasibility and usability in performing real-world tasks [69]. Therefore, we conducted case studies with domain experts (E3 and E5) to evaluate MExplore within authentic analytical scenarios. Through extensive discussion, the experts identified three critical stages of medical expertise acquisition as core tasks: (1) the discovery of areas of interest, (2) the association analysis, and (3) the construction of illness scripts. The session commenced with a 20-minute tutorial on MExplore's features, followed by 1 hour of autonomous exploration focused on the predefined tasks. To capture nuanced qualitative feedback, a think-aloud protocol was used, which encouraged participants to externalize their thought process, ask questions, or provide verbal comments throughout the session.

#### User Study

##### Study Overview

The primary objective of the user study was to quantitatively assess the effectiveness of MExplore in facilitating the acquisition of medical expertise. We adopted illness script construction as the evaluative metric to operationalize the measurement of expertise [4]. The experimental design and comparative study protocols were informed by established

methodologies in visual analytics and medical education research [15,70-72].

### Participants

We recruited 20 second-year undergraduate medical students (10 male and 10 female). None of the participants had systematic prior knowledge of the specific diseases selected for the task, thereby minimizing potential confounding effects from background knowledge.

### Tasks and Materials

Three diseases—oral candidiasis (D1), meningitis (D2), and herpes zoster (D3)—were selected from a list of diseases for which illness scripts could be activated [73]. The participants had not studied these diseases systematically beforehand, thereby eliminating the potential confounding effect of prior knowledge. Following prior studies [70], the task began by providing the participants with a brief description of typical cases for each disease, including relevant medical history and examination results, excluding the disease name. Based on this information, participants were asked to identify the disease and complete an illness script template, the information of which was divided into 3 categories: enabling conditions (EC), fault (FT), and consequences (CQ) [28,73].

### Experimental Procedure

Participants were randomly assigned to two balanced groups ( $n=10$  each, 5 males and 5 females). The experimental group (MEX) used MExplore to complete the tasks. The control group (OTH) was permitted to use external resources (eg, textbooks, case retrieval systems, and LLMs) but was restricted from using MExplore. Before the main task, the MEX group received a detailed introduction and a 20-minute

training session to familiarize themselves with the visual views.

### Metrics

Performance was evaluated based on the accuracy and time taken to form illness scripts. Accuracy was scored by comparing responses to a standardized answer key established by the expert panel. Each correct information unit was awarded one point. The total score was normalized to a percentage [74]. Furthermore, to assess the long-term impact of MExplore on expertise retention, a follow-up test was conducted 2 weeks after the initial experiment. The participants were asked to fill out the answer sheets again based solely on recall, without access to the system or external aids [75].

### Statistical Analysis

All analyses were conducted via Python (version 3.10). The performance was evaluated via 3 metrics: accuracy, retention accuracy, and completion time. Both accuracy and retention accuracy are expressed as percentages. Independent sample  $t$  tests were used to compare the mean differences between the MEX and OTH groups. Two-sided  $P<.05$  was considered statistically significant.

### Expert Interview

We conducted semistructured interviews with domain experts (E1-E10); such qualitative feedback is essential to verify whether users benefit from the system support in their specific domain problems [27]. Each expert subsequently completed a 10-item standardized questionnaire (Table 4) via a 5-point Likert scale to assess their attitudes toward MExplore.

**Table 4.** Expert interview questionnaire<sup>a</sup>.

Question	Question content
Q1	MExplore is very easy (difficult) to learn.
Q2	MExplore is very easy (difficult) to use.
Q3	The visual design of MExplore is easy (difficult) to understand.
Q4	The visual interactions of MExplore are easy (difficult) to use.
Q5	I am very willing (unwilling) to use MExplore in exploring and acquiring medical expertise.
Q6	Using MExplore, I can (cannot) efficiently identify core knowledge units within complex medical texts. (R1)
Q7	Using MExplore, I can (cannot) explore medical texts in a gradual, structured manner. (R2)
Q8	Using MExplore, I can (cannot) identify and analyze the interconnections between knowledge units. (R3)
Q9	Using MExplore, I can (cannot) focus on and analyze key knowledge units in detail. (R4)
Q10	MExplore can (cannot) help me construct a comprehensive and reliable illness script.

<sup>a</sup>Q1-Q5 focus on assessing the system performance of MExplore, Q6-Q9 evaluate whether the key analysis requirements (R1-R4) are satisfied, and Q10 is the overall objective.

### Ethical Considerations

The study received ethics approval from the Institutional Review Board of West China Hospital of Stomatology, Sichuan University (ethical approval number: WCHSIRB-D-2024-335; approval date: August 9, 2024). Due to its retrospective nature, this study required no informed consent. All collected data were fully anonymized, and no personally

identifiable information was included. No financial or material incentives were provided to the participants.

## Results

### Case Study

#### Case 1: Identification and Exploration of Areas of Interest

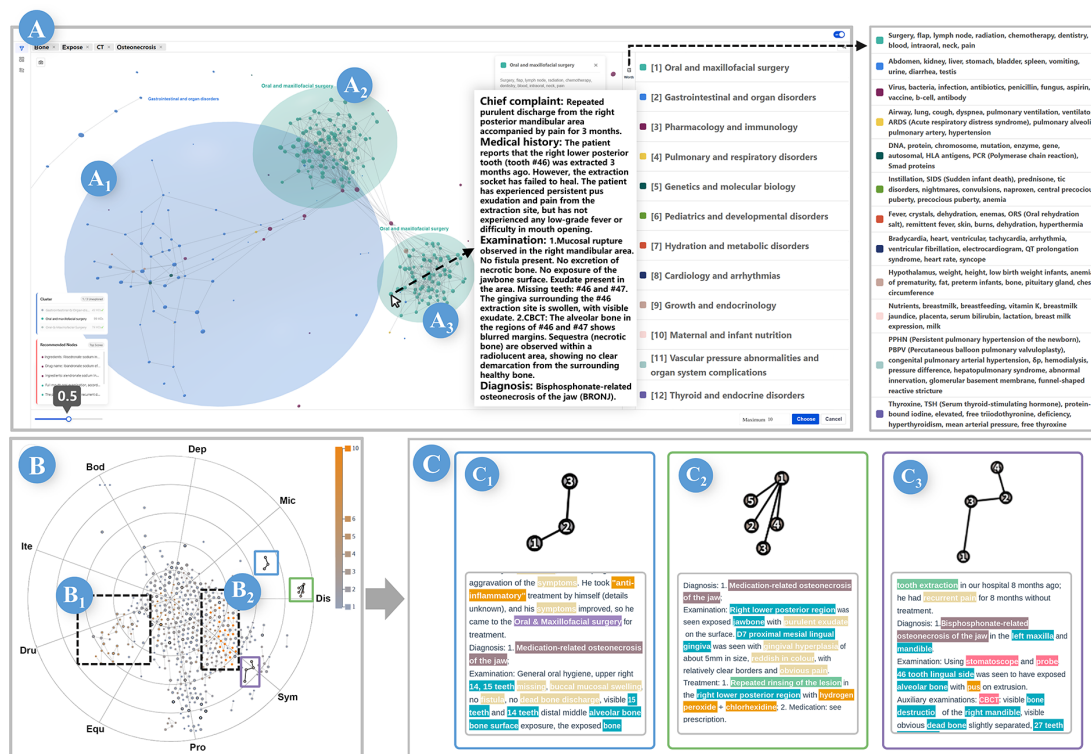
The case study begins with a real scenario with exposed bone in the facial area. A CT scan confirmed the presence of osteonecrosis. The experts entered the keywords *bone*, *expose*, *CT*, and *osteonecrosis* in MExplore. This query generated an MD space where, upon setting the similarity threshold to 0.5, a distinct structure comprising 3 nebulae (A1, A2, and A3) emerged (Figure 6A).

The color of nebula A1 corresponded to topic 2 (*Gastrointestinal and Organ Disorders*), whereas A2 and A3 aligned with topic 1 (*Oral and Maxillofacial Surgery*). Consistent with the patient’s clinical presentation in the maxillofacial region, the experts disregarded A1 to concentrate their analysis on A2 and A3. A detailed exploration of A3 revealed that it contained MDs describing symptoms highly congruent

with the patient’s condition. Notably, these MDs demonstrated strong connectivity to nodes in topic 3 (*Pharmacology and Immunology*), which also appeared in the recommendation list, confirming their high relevance to the search query. Consequently, E3 selected these MDs, along with those in A3, and transitioned them to the MP star map. The maximum size of the subgraph was set to 10, which was determined during the experts’ free exploration and proved effective for segmenting common MDs.

In the MP star map (Figure 6B), E5 first examined subgraphs close to *dis* and *sym* to identify the specific conditions. Three subgraphs were selected to review the text. After these MDs were compared with the original MDs (Figure 6C), the experts concluded that the target disease was medication-related osteonecrosis of the jaw (MRONJ). E5 commented, “Compared with standard search engines, MExplore facilitates the discovery of trustworthy MDs. The structured visual representations of knowledge units enable users to identify the target disease through interactive reasoning rather than randomly searching through uncertain sources.”

**Figure 6.** (A) Topics and distribution of keyword-related MDs, with a connection threshold set to 0.5. (B) Distribution of MP subgraphs within the MDs in A2. (C) The original MD texts of the selected MP subgraphs, with border color annotations indicating their corresponding constellation in the star map. MD: medical document; MP: medical paraphraph.



#### Case 2: Association Analysis and Identification of Key Factors

Upon identifying the target disease, the experts conducted an association analysis to determine the key factors related to the illness script. As defined by Feltovich and Barrows [29], illness scripts consist of 3 components: EC, FT, and CQ. In this case, the disease’s name suggested a connection to the medication, leading the experts to select subgraphs

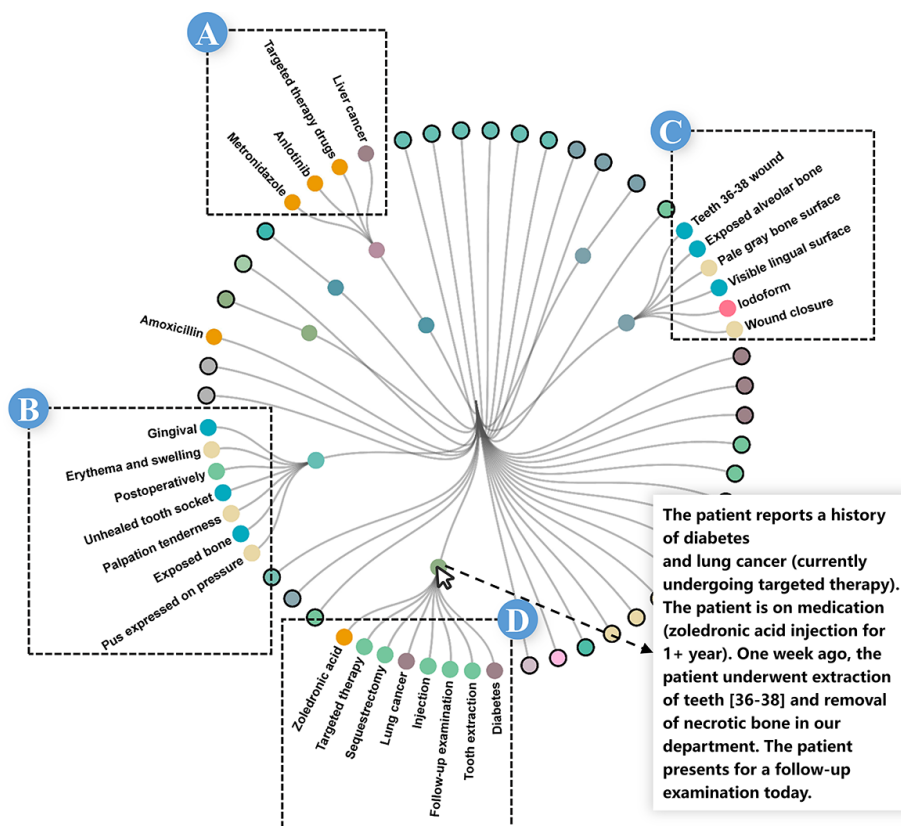
that distribute bias toward *dru* (Figure 6B1). Additionally, the subgraphs adjacent to the target disease with a high total number of MEs were selected (Figure 6B2).

The selected subgraphs were organized into the association analysis view (Figure 7). The experts first analyzed the MES (Figure 7A), which is distinguished by a high proportion of *dru* MEs based on its color encoding. Upon clicking to expand this MES, it was found to contain the *targeted therapy drugs*, *Anlotinib*, which is closely associated with

the *liver cancer* ME, indicating the patient had a history of malignancy and related drug treatment. The text in [Figure 7D](#) further corroborated this by containing *lung cancer* and *targeted therapy*, noting specifically that the patient was using *zoledronic acid*. Therefore, E3 concluded that cancer status and related medications were factors associated with

EC and FT. Regarding CQ, E3 analyzed the MES characterized by high *bod* and *sym* proportions ([Figure 7B and C](#)), identifying bone exposure and exudate as primary CQ factors. E3 concluded, “Performing association analyses on real-case factors can be integrated into clinical reasoning training, enhancing learners’ clinical analysis skills.”

**Figure 7.** The association analysis view of the MES and MEs within the selected subgraph facilitates the exploratory analysis of relationships. ME: medical entity; MES: medical entity set.



### Case 3: In-Depth Analysis and Construction of Illness Scripts

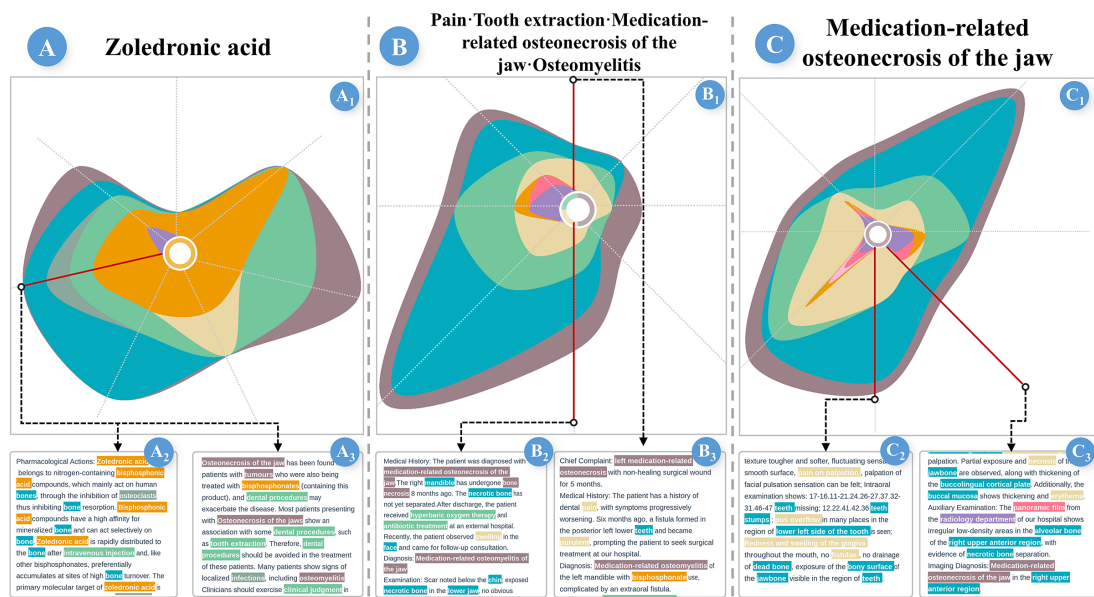
To construct a comprehensive illness script, the experts began by analyzing the factors presented in case 2. E5 was initially interested in the role of *zoledronic acid* as an EC for the development of MRONJ, so he focused on it and generated a focused sectional view ([Figure 8A1](#)). The relevant MES includes a *mic* ME potentially related to the FT. A review of the corresponding MD ([Figure 8A2](#)) revealed that this ME pertains to *osteoclasts*, detailing the drug’s pharmacological action of inhibiting osteoclast growth, which may impair bone repair. The MD further indicates that when bisphosphonates are used in cancer patients, concurrent dental procedures such as tooth extraction may trigger MRONJ ([Figure 8A3](#)).

Subsequently, E3 examined a medical history MES involving *tooth extraction*, such MESs typically feature a

relatively high density of *pro* and *dis* MEs ([Figure 8B1](#)). By examining the MDs of associated chief complaints and other medical history MESs ([Figure 8B2 and B3](#)), multiple co-occurring *pro* MEs were found to mention *osteomyelitis*, which was also referenced in [Figure 8A3](#), suggesting that osteomyelitis may have developed as a CQ of MRONJ, likely secondary to an underlying infection.

To obtain a more definitive CQ, E5 performed a focus analysis of MRONJ ([Figure 8C1](#)), reviewing the MDs corresponding to the MESs that have *sym* and *equ* MEs ([Figure 8C2 and C3](#)). The analysis revealed that bone exposure and pus discharge were prominent clinical symptoms, whereas necrotic bone and radiolucent areas observed on panoramic radiographs emerged as distinctive imaging features. These findings can be used as reliable criteria for clinical diagnosis.

**Figure 8.** (A) Focused analysis of zoledronic acid. (B) Focused analysis of the medical history of MES, including tooth extraction. (C) Focused analysis of MRONJ. MES: medical entity set; MRONJ: medication-related osteonecrosis of the jaw.



This analysis enabled the experts to construct a comprehensive illness script for MRONJ as follows:

- **EC:** Received treatment with bisphosphonates or related medications; a history of cancer or related systemic diseases; recent dental procedures.
- **FT:** Medications such as bisphosphonates inhibit osteoclast function, while dental procedures, particularly tooth extractions, can cause trauma, both of which contribute to the onset of MRONJ.
- **CQ:** Exposed bone, pus drainage, and potential for subsequent infections, such as osteomyelitis.

As E3 noted, “MExplore enables learners to construct illness scripts in a short period, facilitating rapid clinical reasoning and enhancing long-term retention of script knowledge.”

### User Study

Tables 5 and 6 present the mean accuracy and retention scores, whereas Table 7 provides the detailed completion time for each disease across various information classification methods. A more granular comparison of accuracy and retention performance is illustrated in Figure 9.

**Table 5.** User study accuracy score results.

	Max score	Accuracy score		
		MEX <sup>a</sup> , mean (SD)	OTH <sup>b</sup> , mean (SD)	P value
<b>Oral candidiasis</b>				
Enabling conditions	7	6.3 (0.48)	5.9 (0.57)	.11
Fault	7	6.1 (0.32)	5.7 (0.48)	.04
Consequences	8	7.3 (0.48)	7 (0.47)	.18
<b>Meningitis</b>				
Enabling conditions	7	6.1 (0.56)	5.8 (0.92)	.39
Fault	10	9.1 (0.32)	8.4 (0.52)	.002
Consequences	10	9.3 (0.67)	8 (0.47)	<.001
<b>Herpes zoster</b>				
Enabling conditions	9	7.9 (0.88)	7.1 (0.57)	.03
Fault	10	8.9 (0.57)	8.2 (0.42)	.006
Consequences	8	7.4 (0.52)	6.9 (0.57)	.05

<sup>a</sup>Acquiring expertise via MExplore.

<sup>b</sup>Acquiring expertise without MExplore but using other resources.

**Table 6.** User study retention accuracy score results.

	Max score	Retention accuracy score		
		MEX <sup>a</sup> , mean (SD)	OTH <sup>b</sup> , mean (SD)	P value
<b>Oral candidiasis</b>				
Enabling conditions	7	5.7 (0.67)	5 (1.05)	.09
Fault	7	5.6 (0.70)	4.9 (0.74)	.04
Consequences	8	6.7 (0.67)	5.7 (1.34)	.05
<b>Meningitis</b>				
Enabling conditions	7	5.5 (0.53)	5.1 (0.99)	.28
Fault	10	8.4 (0.70)	7.1 (0.88)	.002
Consequences	10	8.6 (0.84)	7.4 (0.97)	.008
<b>Herpes zoster</b>				
Enabling conditions	9	7.2 (1.03)	6.7 (1.06)	.30
Fault	10	8.1 (0.74)	6.9 (0.99)	.007
Consequences	8	6.8 (1.03)	6 (1.15)	.11

<sup>a</sup>Acquiring expertise via MExplore.

<sup>b</sup>Acquiring expertise without MExplore but using other resources.

**Table 7.** User study completion time results.

	Completion time (minutes)		P value
	MEX <sup>a</sup> , mean (SD)	OTH <sup>b</sup> , mean (SD)	
Oral candidiasis	10.45 (1.41)	12.32 (1.88)	.02
Meningitis	13.40 (1.73)	13.85 (2.41)	.64
Herpes zoster	11.82 (1.76)	15.05 (2.54)	.004

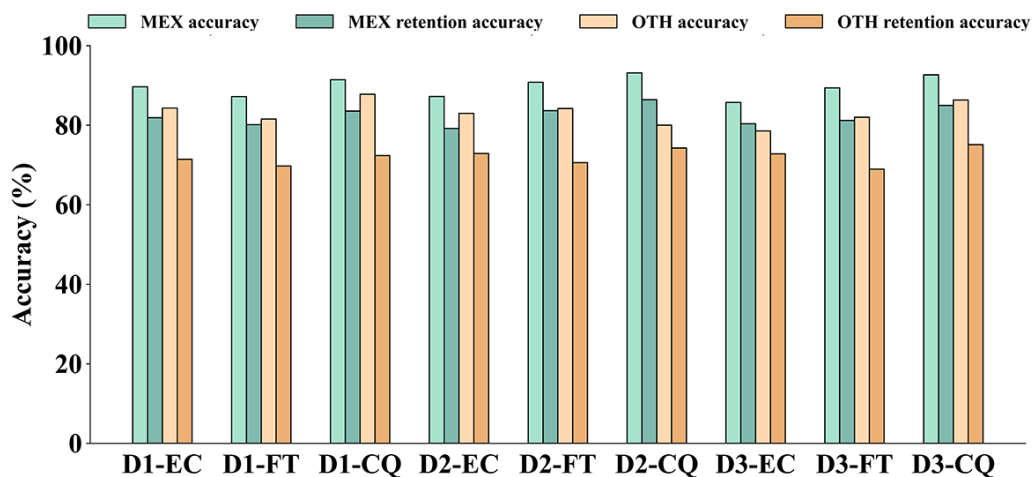
<sup>a</sup>Acquiring expertise via MExplore.

<sup>b</sup>Acquiring expertise without MExplore but using other resources.

The results showed that the MEX group demonstrated superior performance. Independent *t* tests confirmed significant improvements in overall mean accuracy (68.3 out of 76 (90%) vs 63.1 out of 76 (83%),  $P < .001$ ). This advantage persisted in long-term retention: the MEX group achieved an overall mean retention accuracy of 62.5 out of 76 (82%), significantly surpassing the 54.7 out of 76 (72%) observed in the OTH group ( $P = .003$ ). These findings suggest that the multilevel metaphors and interactive exploration in

MEX may facilitate knowledge consolidation into long-term memory. Regarding efficiency, the MEX group had significantly shorter completion times for D1 ( $P = .02$ ) and D3 ( $P = .004$ ). Beyond speed, the MEX group exhibited consistently lower SDs across all tasks (Table 7), suggesting that the structured exploration path may reduce user disorientation and cognitive load when navigating complex medical knowledge.

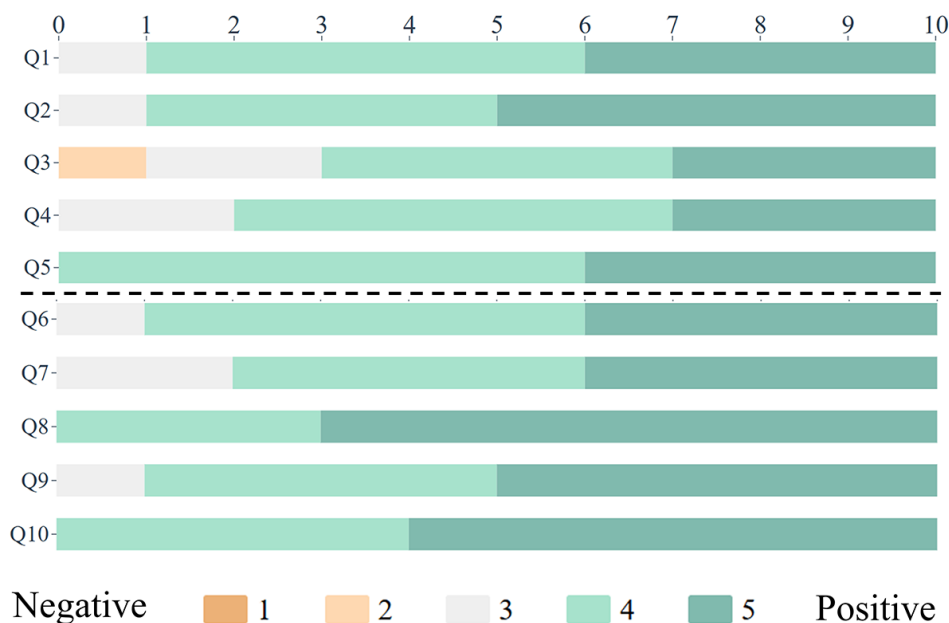
**Figure 9.** User study results: comparison of accuracy and retention accuracy. CQ: consequences; D1: oral candidiasis; D2: meningitis; D3: herpes zoster; EC: enabling conditions; FT: fault; MEX: experimental group; OTH: control group.



## Expert Interview

The results of the questionnaire are presented in Figure 10.

**Figure 10.** Results of the questionnaire. The number at the top represents the number of expert responses. The color of the bar represents the degree of negative (1 and 2), positive (4 and 5), or neutral (3) response.



## System Performance

The experts all recommended MExplore for its user-friendly interface and interaction capabilities. The system’s hierarchical visualization, organized by data granularity and incremental exploration ability, significantly enhances user understanding and analysis. Although some of the experts (E5-E8) had no prior experience with visual analytics systems, they were able to easily comprehend and effectively use the system after receiving minimal training. E4 concurred that the ME-based exploration learning process in MExplore allows learners to focus on the core concepts of the target domain while creating an independent exploration path. E2 suggested incorporating features such as exploration playback and snapshot functionality to further increase the system’s usability. In summary, the experts highly praised the system’s visual design and interaction approach, asserting that it holds significant potential to improve the efficiency of knowledge acquisition for medical professionals.

## Analysis Requirements

As shown in Figure 10, the experts believe that most of the key analytical requirements have been well met. E9 noted that the design of the association analysis view and the focused sectional view enabled users to correlate and concentrate their analysis on key knowledge points, thereby identifying relevant and valuable information—an essential aspect of the knowledge acquisition process. E8 noted that compared with search-based learning methods or chatbots, which have become popular in recent years, this approach can exercise the learner’s thinking ability and improve knowledge retention and accumulation. E5 further highlighted, “In addition to assisting novices in acquiring expertise, MExplore

can also support experts and researchers in disease studies by aiding in research, summarization, and the discovery of new features.”

## Discussion

### Lessons Learned

The key insight from this study is the importance of comprehensibility. Learners must be able to focus on acquiring knowledge throughout the learning process, and the visualizations and interactions should be designed in a way that does not overwhelm or distract users. Through an iterative design process involving both experts and users, we find a clear preference for simple and intuitive visual forms. Based on this feedback, we developed a multilevel metaphor visual analytics framework to reduce cognitive strain and facilitate a shift in paradigm from traditional learning methods to more interactive forms of knowledge acquisition.

### Generalizability

#### Language Adaptability

Although MExplore uses a model pretrained on Chinese corpora, the framework is not strictly bound by linguistic characteristics. Transformer-based architectures have demonstrated robust concept understanding across diverse languages when fine-tuned on domain-specific data [76]. Consequently, the pipeline can support cross-lingual applications by incorporating domain-specific clinical corpora.

## User Adaptability

As experts indicated in the Analysis Requirements section, MExplore's extraction and knowledge association capabilities benefit a broad spectrum of users, not limited to novices. It can further assist senior clinicians and researchers in identifying novel disease features and advancing clinical studies.

## Data Adaptability

To ensure broad generalizability, our retrieval and filtering mechanisms primarily leverage the semantic embedding similarity of unstructured text. However, we recognize that practical medical education often requires precise constraints. Our framework allows users to incorporate a deterministic metadata filtering layer (eg, filtering by visit date, department, and document type) into the search module tailored to their specific datasets. This enables the framework to meet curriculum-specific requirements and adapt to diverse educational goals beyond unstructured content analysis.

## Domain Adaptability

The methodology can be extended beyond medicine to other domains requiring structured knowledge exploration, such as jurisprudence. By fine-tuning predictive models on annotated domain-specific texts, the MExplore framework can be repurposed to visualize entity relationships and facilitate knowledge acquisition in nonclinical fields.

## Limitations

A notable limitation is the linguistic specificity of the current dataset. Clinical documents in different languages exhibit unique syntactic complexities and sublanguages. While our pipeline is architecture-agnostic, performance in non-Chinese contexts depends on the availability of high-quality

domain-specific corpora. Future iterations could incorporate neural machine translation and annotation projection to automatically generate training resources for low-resource languages, thereby enhancing multilingual applicability. Additionally, we plan to expand the dataset through multi-institutional collaboration to cover a broader range of diseases. While the system currently supports large-scale queries, further optimizations—such as approximate nearest neighbor search and WebGL rendering—will be explored to accommodate expanding data volumes. Lastly, the user study was conducted with a cohort of medical students (n=20). While suitable for testing learnability, this demographic may not reflect expert clinical decision-making. Future work will expand evaluations to practicing clinicians to better validate the system's professional utility and generalizability.

## Conclusions

In this paper, we introduced MExplore, a visual analysis tool designed to support medical learners in exploring and acquiring medical knowledge. MExplore extracts MEs from large-scale medical texts and constructs a knowledge structure, offering a multilevel metaphor visual analytics framework. This framework includes 4 coordinated views: the MD space view, the MP star map, the association analysis view, and the focused sectional view. Together, these views enable users to tailor their learning paths, fostering deeper comprehension and improved retention. By facilitating the construction of complex, interconnected schemas typical of medicine, MExplore significantly supports learners in acquiring medical expertise.

We conducted 3 case studies, a user study with real-world datasets, and expert interviews. The results highlight MExplore's effectiveness in facilitating the exploration and analysis of medical texts, significantly enhancing learning efficiency and supporting the acquisition of medical expertise.

## Funding

This work was supported in part by the Industry-University-Research Innovation Fund for Chinese Universities (2024XL004), the Sichuan Science and Technology Program (2025ZNSFSC0459), and the Comprehensive Reform Project on Innovation-Oriented Practical Education Empowered by Artificial Intelligence at Sichuan University.

## Data Availability

The Chinese Biomedical Language Understanding Evaluation (CBLUE) dataset analyzed in this study is publicly available through the Aliyun platform [77]. The Medical Text of West China School (Hospital) of Stomatology (MWCSS) dataset, which contains sensitive clinical data, is subject to strict confidentiality agreements and institutional privacy policies and is not publicly available. However, deidentified data may be shared by the authors upon reasonable request, subject to institutional approval and a formal data-use agreement. Both the source code and the fine-tuned MacBERT model weights are publicly accessible in our GitHub repository [78].

## Authors' Contributions

XP and CL conceptualized the study. JL and YH contributed to data curation. XP, CL, and YH developed and validated the software. XP drafted the original manuscript. JL and ML reviewed and edited the manuscript. All the authors have read and approved the final manuscript. JL and ML contributed equally as the corresponding authors of this manuscript.

## Conflicts of Interest

None declared.

## References

1. Medical school enrollment reaches a new high. Association of American Medical Colleges. Jan 9, 2025. URL: <https://www.aamc.org/news/medical-school-enrollment-reaches-new-high> [Accessed 2026-4-10]

2. China Health Statistics Yearbook 2022. National Health Commission of the People's Republic of China. Jan 24, 2022. URL: <https://www.nhc.gov.cn/mohwsbwstjxxzx/tjnj/202501/8193a8edda0f49df80eb5a8ef5e2547c.shtml> [Accessed 2026-4-10]
3. Patel VL, Kaufman DR. Medical expertise, cognitive psychology of. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. Pergamon; 2001:9515-9518. ISBN: 978-0-08-043076-8
4. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev.* Sep 1993;5(3):205-221. [doi: [10.1007/BF01323044](https://doi.org/10.1007/BF01323044)]
5. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med.* Feb 2000;75(2):182-190. [doi: [10.1097/00001888-200002000-00020](https://doi.org/10.1097/00001888-200002000-00020)] [Medline: [10693854](https://pubmed.ncbi.nlm.nih.gov/10693854/)]
6. Xiao H, Zhou F, Liu X, et al. A comprehensive survey of large language models and multimodal large language models in medicine. *Inf Fusion.* May 2025;117:102888. [doi: [10.1016/j.inffus.2024.102888](https://doi.org/10.1016/j.inffus.2024.102888)]
7. Griot M, Hemptinne C, Vanderdonck J, Yuksel D. Large language models lack essential metacognition for reliable medical reasoning. *Nat Commun.* Jan 14, 2025;16(1):642. [doi: [10.1038/s41467-024-55628-6](https://doi.org/10.1038/s41467-024-55628-6)] [Medline: [39809759](https://pubmed.ncbi.nlm.nih.gov/39809759/)]
8. Kim Y, Jeong H, Chen S, et al. Medical hallucination in foundation models and their impact on healthcare. arXiv. Preprint posted online on Feb 26, 2025. [doi: [10.48550/arXiv.2503.05777](https://doi.org/10.48550/arXiv.2503.05777)]
9. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The role of large language models in medical education: applications and implications. *JMIR Med Educ.* 2023;9:e50945. [doi: [10.2196/50945](https://doi.org/10.2196/50945)]
10. Dergaa I, Ben Saad H, Glenn JM, et al. From tools to threats: a reflection on the impact of artificial-intelligence chatbots on cognitive health. *Front Psychol.* 2024;15:1259845. [doi: [10.3389/fpsyg.2024.1259845](https://doi.org/10.3389/fpsyg.2024.1259845)] [Medline: [38629037](https://pubmed.ncbi.nlm.nih.gov/38629037/)]
11. Stadler M, Bannert M, Sailer M. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Comput Human Behav.* Nov 2024;160:108386. [doi: [10.1016/j.chb.2024.108386](https://doi.org/10.1016/j.chb.2024.108386)]
12. Lee HP, Sarkar A, Tankelevitch L, et al. The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In: CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2025. [doi: [10.1145/3706598.3713778](https://doi.org/10.1145/3706598.3713778)]
13. Roustan D, Bastardot F. The clinicians' guide to large language models: a general perspective with a focus on hallucinations. *Interact J Med Res.* Jan 28, 2025;14(1):e59823. [doi: [10.2196/59823](https://doi.org/10.2196/59823)] [Medline: [39874574](https://pubmed.ncbi.nlm.nih.gov/39874574/)]
14. Jin Z, Cui S, Guo S, Gotz D, Sun J, Cao N. CarePre: an intelligent clinical decision assistance system. *ACM Trans Comput Healthc.* 2020;1(1):1-20. [doi: [10.1145/3344258](https://doi.org/10.1145/3344258)]
15. Xu T, Ma Y, Pan T, et al. Visual analytics of multidimensional oral health surveys: data mining study. *JMIR Med Inform.* Aug 1, 2023;11(1):e46275. [doi: [10.2196/46275](https://doi.org/10.2196/46275)] [Medline: [37526971](https://pubmed.ncbi.nlm.nih.gov/37526971/)]
16. Permana B, Harris PNA, Roberts LW, et al. HAIviz: an interactive dashboard for visualising and integrating healthcare-associated genomic epidemiological data. *Microb Genom.* Feb 2024;10(2):001200. [doi: [10.1099/mgen.0.001200](https://doi.org/10.1099/mgen.0.001200)] [Medline: [38358326](https://pubmed.ncbi.nlm.nih.gov/38358326/)]
17. Siirtola H, Gracia-Tabuenca J, Raisamo R, Niemi M, Reeve MP, Laitinen T. Glyph-based visualization of health trajectories. In: 2022 26th International Conference on Information Visualisation. IEEE; 2022. [doi: [10.1109/IV56949.2022.00075](https://doi.org/10.1109/IV56949.2022.00075)]
18. Grishman R, Sundheim B. Message understanding conference-6: a brief history. In: Proceedings of the 16th Conference on Computational Linguistics. Association for Computational Linguistics; 1996. [doi: [10.3115/992628.992709](https://doi.org/10.3115/992628.992709)]
19. Mehmood T, Serina I, Lavelli A, Putelli L, Gerevini A. On the use of knowledge transfer techniques for biomedical named entity recognition. *Future Internet.* 2023;15(2):79. [doi: [10.3390/fi15020079](https://doi.org/10.3390/fi15020079)]
20. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv. Preprint posted online on Aug 9, 2015. [doi: [10.48550/arXiv.1508.01991](https://doi.org/10.48550/arXiv.1508.01991)]
21. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural architectures for named entity recognition. arXiv. Preprint posted online on Mar 4, 2016. [doi: [10.48550/arXiv.1603.01360](https://doi.org/10.48550/arXiv.1603.01360)]
22. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics; 2019. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
23. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc; 2017. [doi: [10.5555/3295222.3295349](https://doi.org/10.5555/3295222.3295349)]
24. Fall LH, English R, Fulton TB, et al. Thinking slow more quickly: development of integrated illness scripts to support cognitively integrated learning and improve clinical decision-making. *Med Sci Educ.* Jun 2021;31(3):1005-1007. [doi: [10.1007/s40670-021-01293-z](https://doi.org/10.1007/s40670-021-01293-z)] [Medline: [34457943](https://pubmed.ncbi.nlm.nih.gov/34457943/)]

25. Cloude EB, Ballelos NAM, Azevedo R, et al. Designing intelligent systems to support medical diagnostic reasoning using process data. In: *Artificial Intelligence in Education*. Springer; 2021. [doi: [10.1007/978-3-030-78270-2\\_19](https://doi.org/10.1007/978-3-030-78270-2_19)]
26. Likert R. A technique for the measurement of attitudes. *Arch Psychol*. 1932;22(140):1-55. URL: <https://psycnet.apa.org/record/1933-01885-001> [Accessed 2026-04-10]
27. Munzner T. A nested model for visualization design and validation. *IEEE Trans Vis Comput Graph*. 2009;15(6):921-928. [doi: [10.1109/TVCG.2009.111](https://doi.org/10.1109/TVCG.2009.111)] [Medline: [19834155](https://pubmed.ncbi.nlm.nih.gov/19834155/)]
28. Custers EJFM. Thirty years of illness scripts: theoretical origins and practical applications. *Med Teach*. May 2015;37(5):457-462. [doi: [10.3109/0142159X.2014.956052](https://doi.org/10.3109/0142159X.2014.956052)] [Medline: [25180878](https://pubmed.ncbi.nlm.nih.gov/25180878/)]
29. Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, Volder ML, editors. *Tutorials in Problem-Based Learning: New Directions in Training for the Health Professions*. Van Gorcum; 1984:128-142. ISBN: 9023220641
30. Bagheri A, Giachanou A, Mosteiro P, Verberne S. Natural language processing and text mining (turning unstructured data into structured). In: Denaxas S, Oberski D, Moore J, editors. *Clinical Applications of Artificial Intelligence in Real-World Data*. Springer; 2023:69-93. ISBN: 978-3-031-36678-9\_5
31. Perfetti CA. Cognitive research can inform reading education. *J Res Read*. Sep 1995;18(2):106-115. [doi: [10.1111/j.1467-9817.1995.tb00076.x](https://doi.org/10.1111/j.1467-9817.1995.tb00076.x)]
32. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci*. Apr 1988;12(2):257-285. [doi: [10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)]
33. Cognitive load theory: a guide to applying cognitive load theory to your teaching. Office of Educational Improvement, Medical College of Wisconsin. May 2022. URL: <https://www.mcw.edu/-/media/MCW/Education/Academic-Affairs/OEI/Faculty-Quick-Guides/Cognitive-Load-Theory.pdf> [Accessed 2026-4-10]
34. Natesan S, Bailitz J, King A, et al. Clinical teaching: an evidence-based guide to best practices from the council of emergency medicine residency directors. *West J Emerg Med*. Jul 3, 2020;21(4):985-998. [doi: [10.5811/westjem.2020.4.46060](https://doi.org/10.5811/westjem.2020.4.46060)] [Medline: [32726274](https://pubmed.ncbi.nlm.nih.gov/32726274/)]
35. Ganascia JG. Abstraction of levels of abstraction. *J Exp Theor Artif Intell*. Jan 2, 2015;27(1):23-35. [doi: [10.1080/0952813X.2014.940685](https://doi.org/10.1080/0952813X.2014.940685)]
36. Qiao YQ, Shen J, Liang X, et al. Using cognitive theory to facilitate medical education. *BMC Med Educ*. Apr 14, 2014;14(1):79. [doi: [10.1186/1472-6920-14-79](https://doi.org/10.1186/1472-6920-14-79)] [Medline: [24731433](https://pubmed.ncbi.nlm.nih.gov/24731433/)]
37. Lujan HL, DiCarlo SE. The paradox of knowledge: why medical students know more but understand less. *Med Sci Educ*. Jun 2025;35(3):1761-1766. [doi: [10.1007/s40670-025-02379-8](https://doi.org/10.1007/s40670-025-02379-8)] [Medline: [40626000](https://pubmed.ncbi.nlm.nih.gov/40626000/)]
38. Mayer A, Hege I, Kononowicz AA, Müller A, Sudacka M. Collaborative development of feedback concept maps for virtual patient-based clinical reasoning education: mixed methods study. *JMIR Med Educ*. Jan 30, 2025;11:e57331. [doi: [10.2196/57331](https://doi.org/10.2196/57331)] [Medline: [39883032](https://pubmed.ncbi.nlm.nih.gov/39883032/)]
39. Etukakpan AU, Waldhuber MG, Janke KK, Netere AK, Angelo T, White PJ. Core concept identification in STEM and related domain education: a scoping review of rationales, methods, and outputs. *Front Educ*. 2025;10:1547994. [doi: [10.3389/educ.2025.1547994](https://doi.org/10.3389/educ.2025.1547994)]
40. Medwell J, Wray D. CONCEPT-based teaching and learning: a review of the research literature. In: *13th Annual International Conference of Education, Research and Innovation*. IATED; 2020. [doi: [10.21125/iceri.2020.0144](https://doi.org/10.21125/iceri.2020.0144)]
41. Zhang N, Chen M, Bi Z, et al. CBLUE: a Chinese biomedical language understanding evaluation benchmark. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics; 2022. [doi: [10.18653/v1/2022.acl-long.544](https://doi.org/10.18653/v1/2022.acl-long.544)]
42. FlagEval. Beijing Academy of Artificial Intelligence. URL: <https://flageval.baai.ac.cn/> [Accessed 2026-4-10]
43. Cui Y, Che W, Liu T, Qin B, Wang S, Hu G. Revisiting pre-trained models for Chinese natural language processing. *arXiv*. Preprint posted online on Apr 29, 2020. [doi: [10.48550/arXiv.2004.13922](https://doi.org/10.48550/arXiv.2004.13922)]
44. Dettmers T, Holtzman A, Pagnoni A, Zettlemoyer L. QLORA: efficient finetuning of quantized LLMs. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. Curran Associates Inc; 2023. [doi: [10.5555/3666122.3666563](https://doi.org/10.5555/3666122.3666563)]
45. Hu Y, Zuo X, Zhou Y, et al. Information extraction from clinical notes: are we ready to switch to large language models?. *arXiv*. Preprint posted online on Nov 15, 2024. [doi: [10.48550/arXiv.2411.10020](https://doi.org/10.48550/arXiv.2411.10020)]
46. Keraghel I, Morbieu S, Nadif M. A survey on recent advances in named entity recognition. *arXiv*. Preprint posted online on Dec 20, 2024. [doi: [10.13140/RG.2.2.35314.00965](https://doi.org/10.13140/RG.2.2.35314.00965)]
47. Mohan GB, Kumar RP, Elakkiya R, Pendem M. Fine-tuned BERT based multilingual model for named entity recognition in native Indian languages. In: *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*. IEEE; 2023. [doi: [10.1109/EASCT59475.2023.10392937](https://doi.org/10.1109/EASCT59475.2023.10392937)]

48. Anas M, Saiyeda A, Sohail SS, Cambria E, Hussain A. Can generative AI models extract deeper sentiments as compared to traditional deep learning algorithms? *IEEE Intell Syst.* 2024;39(2):5-10. [doi: [10.1109/MIS.2024.3374582](https://doi.org/10.1109/MIS.2024.3374582)]
49. Shyr C, Hu Y, Bastarache L, et al. Identifying and extracting rare diseases and their phenotypes with large language models. *J Healthc Inform Res.* Jun 2024;8(2):438-461. [doi: [10.1007/s41666-023-00155-0](https://doi.org/10.1007/s41666-023-00155-0)] [Medline: [38681753](https://pubmed.ncbi.nlm.nih.gov/38681753/)]
50. Kim K, Park S, Min J, et al. Multifaceted natural language processing task-based evaluation of bidirectional encoder representations from transformers models for bilingual (Korean and English) clinical notes: algorithm development and validation. *JMIR Med Inform.* Oct 30, 2024;12:e52897. [doi: [10.2196/52897](https://doi.org/10.2196/52897)] [Medline: [39475725](https://pubmed.ncbi.nlm.nih.gov/39475725/)]
51. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics.* Feb 15, 2020;36(4):1234-1240. [doi: [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682)] [Medline: [31501885](https://pubmed.ncbi.nlm.nih.gov/31501885/)]
52. Wu X, Gao C, Zang L, Han J, Wang Z, Hu S. ESIMCSE: enhanced sample building method for contrastive learning of unsupervised sentence embedding. *arXiv.* Preprint posted online on Sep 9, 2021. [doi: [10.48550/arXiv.2109.04380](https://doi.org/10.48550/arXiv.2109.04380)]
53. Dieng AB, Ruiz FJR, Blei DM. Topic modeling in embedding spaces. *Trans Assoc Comput Linguist.* Dec 2020;8:439-453. [doi: [10.1162/tacl\\_a\\_00325](https://doi.org/10.1162/tacl_a_00325)]
54. Ajinaja MO, Fakoya JT, Ogunwale YE, et al. A comparative evaluation of probabilistic and transformer-based topic models across diverse and multilingual text corpora. *Neural Process Lett.* 2026;58(1):9. [doi: [10.1007/s11063-025-11820-3](https://doi.org/10.1007/s11063-025-11820-3)]
55. Wang J, Yi X, Guo R, et al. Milvus: a purpose-built vector data management system. In: *Proceedings of the 2021 International Conference on Management of Data.* Association for Computing Machinery; 2021. [doi: [10.1145/3448016.3457550](https://doi.org/10.1145/3448016.3457550)]
56. Hodler AE, Needham M. Graph data science using Neo4j. In: Bader DA, editor. *Massive Graph Analytics.* Chapman and Hall/CRC; 2022:433-457. ISBN: 9781003033707-20
57. Robinson I, Webber J, Eifrem E. *Graph Databases: New Opportunities for Connected Data.* O'Reilly Media; 2015. ISBN: 1491930896
58. Roberts JC. State of the art: coordinated & multiple views in exploratory visualization. In: *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization.* IEEE Computer Society; 2007. [doi: [10.1109/CMV.2007.20](https://doi.org/10.1109/CMV.2007.20)]
59. Carroll JM, Mack RL, Kellogg WA. Chapter 3: Interface metaphors and user interface design. In: Helander M, editor. *Handbook of Human-Computer Interaction.* North-Holland; 1988:67-85. ISBN: 978-0-444-70536-5
60. Shneiderman B, Plaisant C, Cohen M, Jacobs S, Elmqvist N, Diakopoulos N. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Pearson; 2016. ISBN: 013438038X
61. Ware C. *Information Visualization: Perception for Design.* Morgan Kaufmann; 2020. ISBN: 9780128128756
62. Sanders P, Schulz C. Engineering multilevel graph partitioning algorithms. In: *Proceedings of the 19th European Conference on Algorithms.* Springer; 2011. [doi: [10.5555/2040572.2040624](https://doi.org/10.5555/2040572.2040624)]
63. Wang L, Giesen J, McDonnell KT, Zolliker P, Mueller K. Color design for illustrative visualization. *IEEE Trans Vis Comput Graph.* 2008;14(6):1739-1746. [doi: [10.1109/TVCG.2008.118](https://doi.org/10.1109/TVCG.2008.118)] [Medline: [18989033](https://pubmed.ncbi.nlm.nih.gov/18989033/)]
64. Huang L, Xu M, Chen Z, Liu F. Syllabus design for teacher education MOOCs (massive open online courses): a mixed methods approach. In: *Technology in Education: Pedagogical Innovations.* Springer; 2019. [doi: [10.1007/978-981-13-9895-7\\_14](https://doi.org/10.1007/978-981-13-9895-7_14)]
65. Zhou Z, Cai L, Guo J, et al. ExeVis: concept-based visualization of exercises in online learning. *J Vis. Apr* 2024;27(2):235-254. [doi: [10.1007/s12650-024-00956-4](https://doi.org/10.1007/s12650-024-00956-4)]
66. Sweller J. Cognitive load theory, learning difficulty, and instructional design. *Learn Instr.* Jan 1994;4(4):295-312. [doi: [10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)]
67. Fonseca M, Broeiro-Gonçalves P, Barosa M, et al. Concept mapping to promote clinical reasoning in multimorbidity: a mixed methods study in undergraduate family medicine. *BMC Med Educ.* Dec 18, 2024;24(1):1478. [doi: [10.1186/s12909-024-06484-x](https://doi.org/10.1186/s12909-024-06484-x)] [Medline: [39695556](https://pubmed.ncbi.nlm.nih.gov/39695556/)]
68. Dong H, Lio J, Sherer R, Jiang I. Some learning theories for medical educators. *Med Sci Educ.* Jun 2021;31(3):1157-1172. [doi: [10.1007/s40670-021-01270-6](https://doi.org/10.1007/s40670-021-01270-6)] [Medline: [34457959](https://pubmed.ncbi.nlm.nih.gov/34457959/)]
69. Plaisant C. The challenge of information visualization evaluation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces.* Association for Computing Machinery; 2004. [doi: [10.1145/989863.989880](https://doi.org/10.1145/989863.989880)]
70. Keemink Y, Custers E, van Dijk S, Ten Cate O. Illness script development in pre-clinical education through case-based clinical reasoning training. *Int J Med Educ.* Feb 9, 2018;9:35-41. [doi: [10.5116/ijme.5a5b.24a9](https://doi.org/10.5116/ijme.5a5b.24a9)] [Medline: [29428911](https://pubmed.ncbi.nlm.nih.gov/29428911/)]
71. Moghadami M, Amini M, Moghadami M, Dalal B, Charlin B. Teaching clinical reasoning to undergraduate medical students by illness script method: a randomized controlled trial. *BMC Med Educ.* Feb 2, 2021;21(1):87. [doi: [10.1186/s12909-021-02522-0](https://doi.org/10.1186/s12909-021-02522-0)] [Medline: [33531017](https://pubmed.ncbi.nlm.nih.gov/33531017/)]

72. Young JQ, van Dijk SM, O'Sullivan PS, Custers EJ, Irby DM, ten Cate O. Influence of learner knowledge and case complexity on handover accuracy and cognitive load: results from a simulation study. *Med Educ*. Sep 2016;50(9):969-978. [doi: [10.1111/medu.13107](https://doi.org/10.1111/medu.13107)]
73. Custers E, Boshuizen HPA, Schmidt HG. The role of illness scripts in the development of medical diagnostic expertise: results from an interview study. *Cogn Instr*. Dec 1998;16(4):367-398. [doi: [10.1207/s1532690xci1604\\_1](https://doi.org/10.1207/s1532690xci1604_1)]
74. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med*. Apr 2005;80(4):395-399. [doi: [10.1097/00001888-200504000-00019](https://doi.org/10.1097/00001888-200504000-00019)] [Medline: [15793026](https://pubmed.ncbi.nlm.nih.gov/15793026/)]
75. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol Bull*. May 2006;132(3):354-380. [doi: [10.1037/0033-2909.132.3.354](https://doi.org/10.1037/0033-2909.132.3.354)] [Medline: [16719566](https://pubmed.ncbi.nlm.nih.gov/16719566/)]
76. Akcali Z, Cubuk HS, Oguz A, et al. Automated extraction of key entities from non-English mammography reports using named entity recognition with prompt engineering. *Bioengineering (Basel)*. Feb 10, 2025;12(2):168. [doi: [10.3390/bioengineering12020168](https://doi.org/10.3390/bioengineering12020168)] [Medline: [40001686](https://pubmed.ncbi.nlm.nih.gov/40001686/)]
77. Chinese medical information processing challenge list\_CBLUE. Alibaba Cloud. Feb 21, 2025. URL: <https://tianchi.aliyun.com/cblue> [Accessed 2026-04-11]
78. Pang X. MExplore. GitHub. URL: <https://github.com/Px956678784/MExplore> [Accessed 2026-4-11]

## Abbreviations

**BERT:** Bidirectional Encoder Representations From Transformers  
**CBLUE:** Chinese Biomedical Language Understanding Evaluation  
**CQ:** consequences  
**EC:** enabling conditions  
**FT:** fault  
**LLM:** large language model  
**LoRA:** Low-Rank Adaptation  
**MD:** medical document  
**ME:** medical entity  
**MES:** medical entity set  
**MEX:** experimental group  
**MP:** medical paragraph  
**MRONJ:** medication-related osteonecrosis of the jaw  
**MWCSS:** Medical Text of West China School (Hospital) of Stomatology  
**NER:** named entity recognition  
**OTH:** control group  
**QLoRA:** Quantized Low-Rank Adaptation

*Edited by Arriel Benis; peer-reviewed by Di Weng, Dimitrios Megaritis, Xiaolin Wen; submitted 08.Sep.2025; final revised version received 06.Mar.2026; accepted 06.Mar.2026; published 15.May.2026*

*Please cite as:*

*Pang X, Liu C, Huang Y, Liu M, Liu J*

*An Entity-Based Visual Analytics System Enhancing Medical Expertise Acquisition: Development and Verification Study*

*JMIR Med Inform 2026;14:e83785*

URL: <https://medinform.jmir.org/2026/1/e83785>

doi: [10.2196/83785](https://doi.org/10.2196/83785)

© Xiao Pang, Chang Liu, Yan Huang, MingYou Liu, Jiyuan Liu. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 15.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.