EMRs2CSP : Mining Clinical Status Pathway from Electronic Medical Records

Yifei Chen, Ruihui Hou, Jingping Liu, Tong Ruan*
School of Information Science and Engineering,
East China University of Science and Technology, Shanghai, China
ruantong@ecust.edu.cn

Abstract

Many current studies focus on extracting tests or treatments when constructing clinical pathways, often neglecting the patient's symptoms and diagnosis, leading to incomplete diagnostic and therapeutic logic. Therefore, this paper aims to extract clinical pathways from electronic medical records that encompass complete diagnostic and therapeutic logic, including temporal information, patient symptoms, diagnosis, and tests or treatments. To achieve this objective, we propose a novel clinical pathway representation: the clinical status pathway. We also design a LLM-based pipeline framework for extracting clinical status pathway from electronic medical records, with the core concept being to improve extraction accuracy by modeling the diagnostic and treatment processes. In our experiments, we apply this framework to construct a comprehensive breast cancer-specific clinical status pathway and evaluate its performance on medical question-answering and decision-support tasks, demonstrating significant improvements over traditional clinical pathways. The code is publicly available at https://github.com/finnchen11/EMRs2CSP.

1 Introduction

The clinical pathways (CPs) is a set of standardized diagnostic, therapeutic, and care protocols established for specific diseases based on evidence-based medicine and diagnostic and treatment guidelines. It encompasses the entire diagnostic and therapeutic process, from patient admission to discharge, and outlines the specific medical activities at each stage, along with their chronological sequence (Alahmar and Alkhatib, 2022). Clinical pathways play a significant role in standardizing medical behavior and enhancing the quality and safety of medical care (Wang et al., 2018; Alahmar et al., 2018).

*Corresponding author.

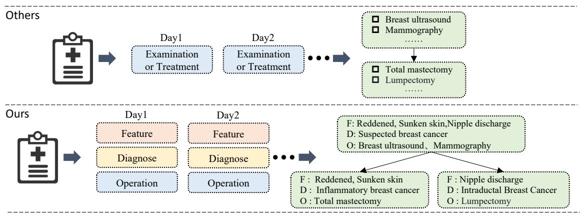

Figure 1: Differences between our approach and others with respect to extraction and representation.

However, current clinical pathway research primarily focuses on medical activities such as examinations or treatments (Yang and Hwang, 2006; dos Santos Garcia et al., 2019; Pika et al., 2020) and the analysis of temporal sequences (Kaymak et al., 2012; Xu et al., 2017), often neglecting the patient's symptoms and diagnostic results during the diagnostic and treatment process. In clinical diagnostic and therapeutic logic, determining the most appropriate treatment path requires identifying the examinations or treatments the patient needs based on their symptoms and diagnosis. Therefore, considering only medical activities and chronological order results in incomplete diagnostic and therapeutic logic in CPs, leading to high variation and incompleteness in their practical application (Huang et al., 2016; Tehrani, 2016). As illustrated in Figure 1, the physician will make an appropriate diagnosis based on the presence of reddened skin or nipple discharge, followed by determining whether a total or partial mastectomy is necessary. Therefore, an effective clinical pathway should include chronological information, the patient's symptoms, diagnosis, and the required examinations and treatments.

To address the aforementioned issues, We propose the Clinical Status Pathway (CSP), a new method for representing clinical pathways. Unlike traditional clinical pathways, which construct a series of tests and treatments based on the length of hospital stay, our representation employs three

Findings of the Association for Computational Linguistics: ACL 2025, pages 17235–17251
July 27 - August 1, 2025

states—feature, diagnosis, and operation—to represent the patient's treatment process, integrating decision conditions and diagnostic and therapeutic information. To derive the CSP, we design an LLM-based pipeline framework for extracting it from electronic medical records (EMRs). This framework first filters the data, then extracts patient state information from EMRs, constructs the clinical status pathway from individual records by integrating the length of hospital stay and the sequence of states, and finally consolidates results from multiple records to form the CSP for a specific disease.

To evaluate the effectiveness of our representations, we use various representations as knowledge bases in LLMs for medical quizzing and decision support tasks. We found that using CSP as a knowledge base outperforms other data types in both medical quizzing and decision support effectiveness. Furthermore, to assess the effectiveness of our extraction framework, we perform the extraction of clinical status pathway using multiple methods. The results demonstrate that the clinical status pathway obtained using our extraction method outperform alternative methods in terms of recall and accuracy. These findings validate the rationality of our representation and the efficacy of our extraction framework.

The main contributions of this paper are as follows:

- We propose a representation known as the clinical status pathway, which incorporates more decision-making conditions and diagnostic and therapeutic information than traditional clinical pathways.
- We propose a LLM-based framework for clinical status pathway extraction that automates the process of extracting clinical status pathway from electronic medical records.
- We conducted extensive experiments using both private and public datasets to evaluate the effectiveness of our framework. The experimental results indicate that our framework outperforms others in both extraction efficacy and practical application.

2 Related Work

Process Mining. Process mining refers to the extraction of event types and timing information from event logs to model business processes. It is a widely used business process analysis method (dos Santos Garcia et al., 2019; Diba et al., 2020), with applications in the healthcare domain (Munoz-Gama et al., 2022; Dallagassa et al., 2022). These studies analyzed healthcare logs to identify frequent patterns, highlighting its importance in understanding CPs. Yang et al. proposed a process mining algorithm to systematically detect healthcare fraud and CPs abuse (Yang and Hwang, 2006). Mans et al. applied process mining techniques to discover treatment patterns for stroke patients across hospitals (Mans et al., 2008). Additionally, Huang et al. introduced a process mining algorithm that generates concise summaries from clinical event logs (Huang et al., 2016). However, due to the complexity of diagnosis and treatment processes, traditional process mining techniques often produce models that are difficult to interpret (**?**Xu et al., 2017). These models lack clarity in representing causal relationships within medical activities, limiting their utility in clinical pathway analysis and applications. To address this challenge, we propose a LLM-based pipeline framework to extract clinical pathways with richer causal relationships from electronic medical records, enhancing their interpretability.

Topic Models. Topic models, a technique for unsupervised text representation learning applied to unconnected documents, offer an effective probabilistic model for extracting hidden topic semantics from healthcare data, based on the assumption that a document consists of different word combinations across multiple topics (Jelodar et al., 2019). In clinical pathway design, automatic extraction of treatment events from large-scale clinical data has become an active research area (Aspland et al., 2021; Quintano Neira et al., 2019; Kempa-Liehr et al., 2020). As medical data includes various informational dimensions, researchers increasingly use topic models to analyze these characteristics. Chen et al. proposed using Latent Dirichlet Allocation to mine hospital charge item data, distinguishing similarities and differences among topics by comparing data from different hospitals (Chen et al., 2015). Huang et al. contributed to mining clinical treatment patterns with topic models (Huang et al., 2013, 2014, 2015). They considered each hospitalization as a document and each treatment activity as a word, mining potential treatment patterns through topic models (Huang et al., 2013). They also incorporated examination results into the topic model, enriching discovered patterns (Huang et al., 2015). Some studies have explored adding temporal information into topic models. Huang et al. integrated temporal data into a topic model to mine time-series treatment patterns (Huang et al., 2014). Xu et al. combined process mining with topic models (Xu et al., 2017), identifying daily topics in treatment processes and revealing their temporal relationships. Li et al. developed a temporal topic model capturing treatment topics and timestamps (Li et al., 2024). However, topic model-based approaches mainly focus on discovering treatment events. Due to the complexity of electronic medical records, these models often produce semantic discrepancies and fail to capture event correlations. Therefore, we modeled the admission process to represent medical information relationships, extracting clinical pathways containing decision logic through an LLM-based pipeline approach.

3 Method

The goal of our approach is to extract CSP from EMRs. In this section, we first introduce a new representation for clinical state pathways, then outline the process of filtering high-quality information from EMR, followed by an explanation of how to extract clinical state pathways from the filtered EMRs, and finally, we discuss how to integrate the extraction results from multiple EMRs into a comprehensive clinical state pathway.Figure 2 illustrates the overall framework of our approach.

3.1 CSP Representation

Our objective is to develop a data storage framework that encapsulates both clinical treatment processes and patient information. Therefore, our objective is to identify a general, abstract, and structured data model to serve as a repository for complex clinical information.

While more sophisticated methods exist to represent clinical information, we are constructing clinical status pathway based on modeling the actual admission process of a patient. Patients typically present with specific symptoms and seek treatment at the hospital. Doctors assess the patient's symptoms and formulate an initial diagnosis. To facilitate further diagnosis or treatment of the currently diagnosed condition, the doctor will prescribe relevant tests or treatments for the patient. After examination or treatment, the patient's symptoms change. When the doctor determines that the patient's condition meets the criteria for discharge, the patient may be discharged from the hospital.

Based on the aforementioned admission process, we categorize the patient's stay in the hospital into three states: Feature, Diagnose, and Operation, which are defined as follows: Feature refers to the patient's symptoms, Diagnose refers to the patient's diagnosis, and Operation refers to the tests or treatments the patient needs to undergo. The transitions between the states are as follows: following the doctor's evaluation, Feature is converted to Diagnose; after the doctor's prescription, Diagnose is converted to Operation; based on the results of examination or treatment, Operation is converted to Feature, and so on, until the doctor's evaluation of Diagnose reaches the criteria for discharge. In addition to the patient states described above, temporal states are represented by Days and Order, which denote the number of days of admission and the sequence of patient states occurring on that day, respectively. The format of the clinical status pathway is illustrated in Figure 3. The content of each node includes three components of patient information: Feature, Diagnose, and Operation, while the edges represent the temporal information, specifically Days and Order.

3.2 Extraction Algorithm

To extract CSP from EMRs, we designed a extraction algorithm based on LLM. The algorithm primarily consists of a state screening module, an extraction module, a time extraction module, and an assembly module. The input to the extraction algorithm is the screened EMR, followed by the time extraction and state extraction modules to obtain the quintuple (days, order, feature, diagnose, operation). Finally, the assembly module processes the data to generate the CSP for the EMR.Details of the extraction algorithm are provided in Appendix B.2, B.3 and B.4.

Screening for Valid Information. The first step in our extraction framework is to filter fields from the EMR that are relevant to the construction of clinical pathways, ensuring the accuracy of the information extraction process. Therefore, our objective is to filter out EMR fields containing diagnostic and treatment information and decision logic to serve as input data for the extraction framework. Given that the diagnostic and treatment process, as well as decision logic, are critical components of the clinical status pathway, we employ a strategy of eliminating irrelevant information while retain-

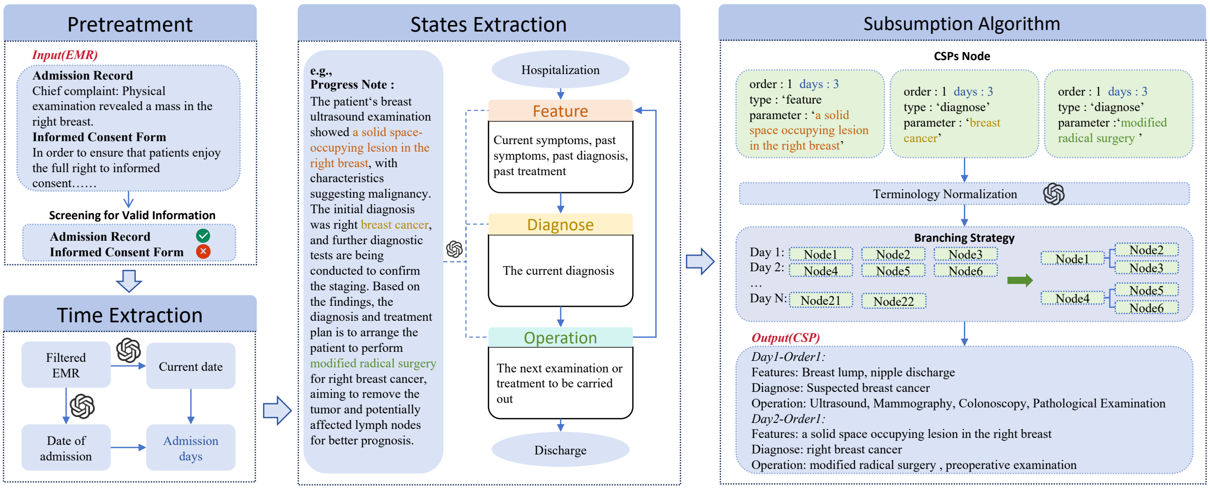

Figure 2: The overall framework of our extraction method

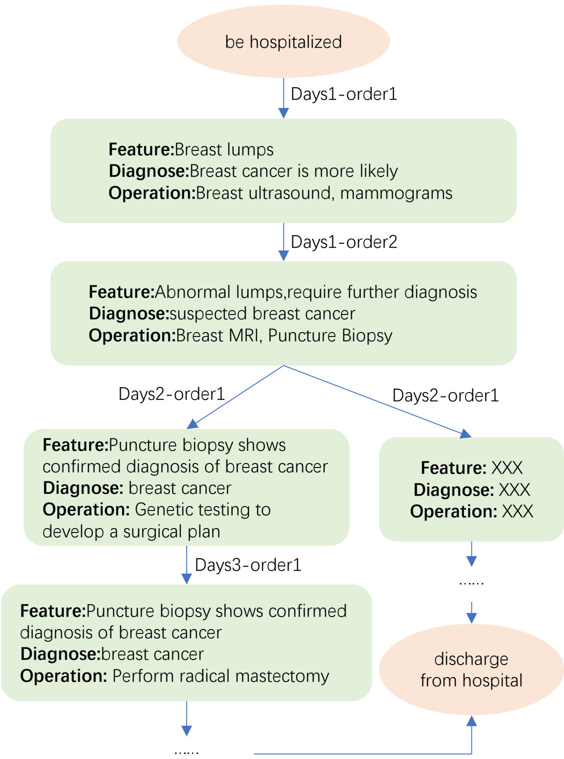

Figure 3: Structure of the Clinical Status Pathway

ing only those fields directly related to these two key elements. Specifically, we engaged experts to review the fields of each medical record to assess whether they contained treatment information and decision logic. Fields that meet the criteria, such as admission records, medical histories, and preoperative summaries, are retained, while fields that do not meet the criteria, such as informed consent for surgery, are discarded. The processed input data are not only of high quality but also information-dense, which minimizes the potential interference of irrelevant data on the extraction results.

Time Extraction Module. To extract the days and order, we designed a time extraction module to determine the specific admission days associated with the current state and the order of events during the treatment process on those days. First, we use a LLM to extract time information from the new admission evaluation form of the current medical record to determine the admission days. Next, we extract the specific time of each subsequent field to determine the order, and based on the admission days, we calculate the current admission day, thereby obtaining the chronological information for the entire process.

State Extraction Module. The State Extraction Module is responsible for extracting the feature, diagnose, and operation, and is the core component of the entire framework. We explicitly define the information contained in the three states: Feature represents the patient's current symptoms, as well as past medical history, treatment history, etc.; Diagnose refers to the diagnosis based on the patient's current feature; and Operation refers to the examination or treatment based on the patient's current diagnosis. First, we traverse each field from the preprocessed EMR and extract a set of Feature, Diagnose, and Operation from each field. We design a set of prompt templates for the LLM to extract information, specifically: first, we have the model determine the tense of each sentence in the current field, indicating whether it occurred in the past or present; then, we have the model perform patient entity extraction for symptoms, diagnosis, test modality, and treatment modality. Finally, the extraction results are summarized according to various timestamps and entity labels. Symptoms, diagnosis, and treatments in the past tense are rewritten as medical history, surgical history, and current symptoms, then added to the list of features; diagnosis in the present tense are added to the list of diagnosis; and tests or treatments in the present tense are added to the list of operations.

Assembly Module. After the previous steps, a quintuple is obtained for each field in the filtered EMR. To facilitate the subsequent merging algorithm, we need to assemble all quintuples from the EMR into a clinical status pathway. Specifically, the module achieves this by defining multiple keys corresponding to the Days. Each key contains an ordered list of treatment records, where each record is represented by a dictionary with the keys Order, Feature, Diagnose, and Operation, and the values correspond to the specific contents of the quintuple.

3.3 Subsumption Algorithm

By following the aforementioned steps, we obtain the clinical pathway for a single electronic medical record. However, data from a single medical record is insufficient to represent the entire specialty pathway. Therefore, we designed the subsumption algorithm to summarize and integrate the clinical pathways from multiple medical records, yielding a more comprehensive clinical pathway. Initially, we divided each pathway in the set of clinical status pathway based on the number of days and arranged the resulting segments in chronological order. Next, for each group, we perform the merging operation as follows:

1. We first identify parts with different features but the same diagnosis and order. During merging, the distinct features are combined into a new feature, while the diagnosis remains unchanged. This process is repeated recursively until no further merging is possible.

2. Next, we identify node parts that share the same feature and diagnosis but have different operations under the same order. During merging, the feature and diagnosis are retained, and the distinct operations are merged into a new operation. This process is repeated recursively until no further merging is possible.

3. Finally, we merge the nodes in chronological order to derive the final clinical status pathway.

It is important to note that the merging process is conducted by an agent. We employ the LLM to merge each part based on the aforementioned logic. After merging, the agent performs a think-aloud judgment to verify whether any intellectual errors exist in the merged result. If the agent identifies an error, it re-executes the merge operation and repeats the verification process three times. If an error persists after three verifications, the agent labels the subpath with the correct knowledge for manual validation. If no error is found during verification, the next round of merging operations is performed.See Appendix B.5 for details.

4 Experiments

4.1 Setup

Dataset. In this study on clinical status pathway extraction, we utilized a private dataset consisting of 400 expert-reviewed and error-free electronic medical records from a breast cancer specialty in a tertiary care hospital. We utilized 70% of this data for clinical status pathway extraction, with the remaining 30% used for the patient information portion of the decision aid task. This dataset ensured the accuracy of the medical record data, effectively preventing biased extraction results caused by errors in the input source. To evaluate the effectiveness of the clinical status pathway using the medical QA task, we selected 1,127 data points from Huatuo-26M (Li et al., 2023) related to breast cancer as the benchmark dataset for evaluation. Additionally, we obtained traditional breast cancer clinical pathways from the Internet and used them to compare with our clinical status pathway. Details of the Huatuo-26M-based screening method and the clinical pathway acquisition method are provided in Appendix B.6.

Baselines. We considered three scenarios: electronic medical records, traditional clinical pathways, and clinical status pathway as knowledge bases. Based on these three scenarios, we performed medical quizzing and assisted decision-making tasks using four models: Chatglm3-6B (GLM et al., 2024), qwen2.5-7B (Yang et al., 2024), Chatgpt-4o (Achiam et al., 2023), and Llama3-Chinese-8B (Shenzhi Wang, 2024), to evaluate which knowledge base performs best on the two tasks.

Evaluation. In assessing the effectiveness of the Q&A task, we utilized metrics such as BLEU-1, BLEU-2, BLEU-3, BLEU-4 (Papineni et al., 2002), ROUGE-1, ROUGE-2 (Lin, 2004), and BertScore (Zhang et al., 2019) to quantify the performance of different databases. To evaluate the effectiveness of assisted decision-making, we introduced GPT-4o, which integrates four aspects of the diagnosis and treatment process: accuracy, completeness, rationality, and security. Finally, we employed accuracy and recall to further assess the effectiveness of our extraction algorithm.

Implementation Details. Initially, we extracted the clinical status pathway from the electronic medical records using GPT-4. We then utilized the BGE (Chen et al., 2024) model to vectorize the electronic medical records, traditional clinical pathways, and clinical status pathway, storing them in the Faiss (Douze et al., 2024) Loud database. From this, we obtained three distinct knowledge bases: the electronic medical records knowledge base, the clinical pathways knowledge base, and the clinical status pathway knowledge base. When utilizing the knowledge bases, we access the vectorized repository based on the inputs of each experiment to retrieve the relevant content, which is then used as external knowledge inputs to the LLM. In evaluating the effectiveness of medical quizzing, we use the questions from Huatuo-26M as inputs and compute the BLEU and ROUGE scores. To evaluate the effectiveness of assisted decision-making, we used the patient's chief complaint from the electronic medical record as input to predict and provide answers for the patient's subsequent treatment process, utilizing different knowledge bases. We compared and analyzed the different responses using GPT-4 in terms of accuracy, completeness, rationality, and safety of the treatment process. In evaluating the extraction algorithm, we used a private dataset of breast cancer EMRs and asked experts to manually annotate the status of each field based on the EMRs to create a manually annotated status set. We then evaluated the algorithm's performance by comparing this set with the status sets extracted by different algorithms and calculating the accuracy and recall. Details of the construction of the knowledge base, as well as the implementation and evaluation of the medical quiz and assisted decision-making, are provided in Appendix B.7, B.8 and B.9.

4.2 Main Results

In this study, we employed BLEU,ROUGE and BertScore as indicators of medical questioning performance, and conducted a comprehensive evaluation of the effectiveness of the assisted decision-making system across four dimensions of the diagnosis and treatment process: accuracy, completeness, reasonableness, and safety, to validate the efficacy and practicality of our proposed method. Table 1 presents the results of a comprehensive comparison between this method and other established methods in terms of medical Q&A capability and assisted decision-making efficacy.

4.2.1 Medical Q&A

The BLEU metric evaluates the quality of the output by measuring the overlap of n-grams between the generated output and the reference text. This metric reflects the similarity and consistency between the generated and reference texts. ROUGE calculates the frequency of n-gram occurrences in the generated output relative to the reference text, reflecting the coverage of the generated content. In general, higher BLEU and ROUGE scores indicate closer alignment between the generated content and the reference text in terms of semantics, structure, and expression. In this study, we employed a specific method to extract and characterize the clinical pathway information and tested the Q&A effect on several LLMs, including ChatGLM-3.6B, Qwen-2.5-7B, ChatGPT-4o, and Llama3-Chinese-8B. The experimental results show that, compared to other methods, the clinical status pathway demonstrates superior performance when used as a knowledge base for medical Q&A. Specifically, the metrics for BLEU-1, BLEU-2, BLEU-3, and BLEU-4 on multiple models are, on average, 0.0140, 0.0105, 0.0085, and 0.0091 higher than those for the electronic medical record knowledge base, and 0.0073, 0.0063, 0.0074, and 0.0064 higher than those for the clinical pathway knowledge base. The performance of ROUGE-1 and ROUGE-2 is on average 0.0243 and 0.0048 higher than the EMR knowledge base, and 0.0181 and 0.0038 higher than the clinical pathway knowledge base. Our approach achieves the best results on BertScore, which effectively proves the effectiveness of our method.

4.2.2 Assisted Decision-making

Accuracy refers to whether the tests or treatments implemented for a patient's condition and diagnosis are precise, providing reliable data support for subsequent medical decisions. Completeness emphasizes whether the diagnostic and therapeutic steps are effectively connected throughout the entire medical process, from admission to discharge, ensuring the coherence and systematic nature of the treatment. Rationality focuses on whether the treatment design considers economy and efficiency, and whether it can effectively control medical costs

Model	Knowledge Base	Medical Q&A							Supported Decision Making			
		BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-1	ROUGE-2	BertScore	Accuracy	Completeness	Rationality	Safety
Chatglm3-6B	EMRs	0.1753	0.0889	0.0502	0.0292	0.2501	0.0665	0.6582	0	0	0	0
	CPs	0.1803	0.0921	0.0528	0.0314	0.2527	0.0685	0.6606	+1.9	+2.9	+2.8	+1.0
	CSP	**0.1836**	**0.0932**	**0.0620**	**0.0397**	**0.2894**	**0.0696**	**0.6638**	**+5.1**	**+6.5**	**+5.7**	**+2.9**
Qwen2.5-7B	EMRs	0.1537	0.0739	0.0375	0.0193	0.2339	0.0534	0.6399	0	0	0	0
	CPs	0.1578	0.0746	0.0375	0.0192	0.2347	0.0527	0.6414	+2.2	+2.4	+1.8	+1.7
	CSP	**0.1698**	**0.0817**	**0.0422**	**0.0226**	**0.2458**	**0.0576**	**0.6508**	**+4.2**	**+6.1**	**+5.0**	**+2.7**
Chatgpt-4o	EMRs	0.2353	0.1266	0.0761	0.0475	0.2915	0.0897	0.7006	0	0	0	0
	CPs	0.2412	0.1302	0.0798	0.0512	0.2976	0.0932	0.7178	+1.2	+1.7	+1.9	+0.8
	CSP	**0.2457**	**0.1366**	**0.0888**	**0.0576**	**0.3128**	**0.0989**	**0.7243**	**+2.8**	**+4.7**	**+4.1**	**+2.1**
Llama3-Chinese-8B	EMRs	0.1645	0.0743	0.0377	0.0203	0.2378	0.0533	0.6531	0	0	0	0
	CPs	0.1690	0.0786	0.0389	0.0244	0.2385	0.0574	0.6561	+2.0	+2.5	+2.3	+0.9
	CSP	**0.1728**	**0.0812**	**0.0421**	**0.0278**	**0.2396**	**0.0597**	**0.6580**	**+4.5**	**+6.2**	**+5.5**	**+2.7**
DeepSeek R1	EMRs	0.2248	0.1200	0.0831	0.0475	0.2598	0.0987	0.6896	0	0	0	0
	CPs	0.2420	0.1315	0.0805	0.0520	0.2980	0.0940	0.7150	+1.5	+2.0	+2.2	+1.0
	CSP	**0.2490**	**0.1380**	**0.0905**	**0.0585**	**0.3130**	**0.1000**	**0.7480**	**+3.2**	**+5.0**	**+4.5**	**+2.5**
Claude3.5 Sonnet	EMRs	0.2153	0.1289	0.0702	0.0392	0.2501	0.0665	0.6582	0	0	0	0
	CPs	0.2210	0.1330	0.0730	0.0420	0.2530	0.0690	0.6420	+1.8	+2.7	+2.5	+1.2
	CSP	**0.2294**	**0.1470**	**0.0835**	**0.0505**	**0.2900**	**0.0710**	**0.6700**	**+4.8**	**+6.3**	**+5.8**	**+3.0**
Gemini 2.0	EMRs	0.2137	0.1139	0.0775	0.0293	0.2339	0.0634	0.6399	0	0	0	0
	CPs	0.2185	0.1155	0.0780	0.0302	0.2357	0.0640	0.6420	+2.1	+2.3	+2.0	+1.5
	CSP	**0.2305**	**0.1220**	**0.0830**	**0.0385**	**0.2465**	**0.0685**	**0.6511**	**+4.0**	**+5.8**	**+5.2**	**+2.8**

Table 1: Evaluation of medical Q&A and supported decision making using various knowledge bases

while ensuring that treatment outcomes meet expectations. Safety requires that during treatment, the patient's underlying disease, special physical condition, and other factors be fully considered to ensure that the potential risk of the chosen treatment to the patient's condition remains manageable. Compared with other methods, the clinical status pathway has shown significant improvement in its effectiveness in assisting decision-making. On four specific indicators, the average improvement over the electronic medical record knowledge base was 3.94, 5.80, 5.11, and 2.67, while the improvement over the clinical pathway knowledge base was 2.27, 3.47, 2.90, and 1.51. This fully demonstrates the effectiveness of our method.

4.3 Comparison of Extraction Methods

Method	Recall	Acc	F1
Inductive Miner	76%	70%	73%
LDA	81%	69%	75%
ICL	79%	88%	83%
Ours	**89%**	**95%**	**92%**

Table 2: Evaluation of different methods for extracting CSP

To further validate the effectiveness of our extraction framework, using GPT-4o as the LLM and using EMRs as a dataset to extract CSP, we compared it with other extraction methods in terms of the reliability of the results. We used accuracy and recall of patient status to assess the reliability of the extraction results. As shown in the Table 2, our method outperforms Inductive Miner (van Detten et al., 2023) and LDA (Jelodar et al., 2019) in extracting CSP from EMRs.

4.4 Detailed Analysis

We conduct a detailed analysis of the ablation of extraction frames and investigate the reasons why CSP yield better results in medical questioning and assisted decision-making.

Ablation Study In the task of extracting clinical pathways from electronic medical records, we propose a pipeline extraction method based on a LLM. To evaluate the impact of different modules in the method, we conduct ablation experiments on our extraction framework. Specifically, we eliminate the time extraction module and the state extraction module, evaluating the effectiveness of the remaining component combinations.

From Table 3, it is clear that removing both the time and state extraction modules significantly degrades information extraction performance. Specifically, removing the time extraction module causes a noticeable regression in all medical quiz metrics. The BLEU-1 score drops by an average of 0.008, and ROUGE-1 decreases by 0.012. This leads to a decline in the effectiveness of the extraction in supporting decision-making, particularly in diagnostic and treatment processes. Completeness sees a notable drop, averaging 2.85. Likewise, removing the state extraction module causes a significant

Model	Method	BLEU-1	ROUGE-1	Accuracy	Completeness	Rationality	Safety
Chatglm3-6B	w/o TEM	0.1788	0.2675	+3.7	+3.1	+4.3	+2.5
	w/o SEM	0.1765	0.2621	+3.0	+5.5	+3.5	+1.0
	Ours	**0.1836**	**0.2894**	**+5.1**	**+6.5**	**+5.7**	**+2.9**
Qwen2.5-7B	w/o TEM	0.1578	0.2388	+3.5	+3.1	+4.0	+1.9
	w/o SEM	0.1601	0.2442	+3.0	+5.9	+3.8	+0.9
	Ours	**0.1698**	**0.2458**	**+4.2**	**+6.1**	**+5.0**	**+2.7**
Chatgpt-4o	w/o TEM	0.2365	0.2927	+1.0	+2.7	+3.1	+1.5
	w/o SEM	0.2372	0.2931	+0.8	+3.7	+2.9	+1.0
	Ours	**0.2457**	**0.3128**	**+2.8**	**+4.7**	**+4.1**	**+2.1**
Llama3-Chinese-8B	w/o TEM	0.1658	0.2391	+3.7	+3.2	+4.3	+2.0
	w/o SEM	0.1662	0.2385	+2.5	+5.9	+3.9	+0.9
	Ours	**0.1728**	**0.2396**	**+4.5**	**+6.2**	**+5.5**	**+2.7**

Table 3: The ablation study on our extraction framework

performance decline across various models, with BLEU-1 and ROUGE-1 scores dropping by 0.007 and 0.012, respectively. The diagnostic and treatment processes also suffer in accuracy, completeness, rationality, and safety, with safety declining by an average of 1.3. In conclusion, both the time and state extraction modules are essential for improving the accuracy and rationality of information extraction in clinical pathways.

Cause Analysis Traditional clinical pathways outline general processes but overlook patient-specific details and decision-making. By extracting patient states from electronic medical records, clinical status pathway can be created, providing clear state-transition logic. This approach maps a patient's condition to feature and diagnostic states, helping identify appropriate procedures. In addition, extracting rare cases from EMR can improve the coverage of paths. In Q&A tasks, the model selects the most relevant pathway based on the question, while for decision-making support, it analyzes the patient's state to identify the best pathway. In contrast, traditional pathways lack state-transition logic, limiting decision-making to the model's inherent knowledge, without fully reflecting the patient's actual condition.

4.5 Case Study

To optimize our method and improve its effectiveness, we classified the error samples from the experiments and conducted a detailed analysis for each error type. As shown in Figure 4, we identified three main types of errors: state extraction (56%), time extraction (25%), and subsumption errors (19%).In-depth analysis revealed that state extraction errors are primarily caused by FEATURE extraction errors. This issue arises from

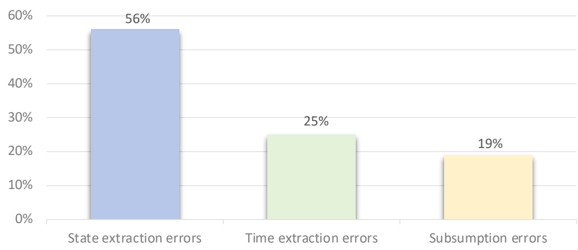

Figure 4: Distribution of error causes

the complexity of medical record information, including the diversity of patients' conditions and the inconsistency in electronic medical record standards. The complexity of medical records can hinder the model's ability to accurately follow the preset extraction logic. Time extraction errors typically occur when the timestamp of the current instrument is not explicitly labeled in the medical records, preventing the model from accurately parsing the admission time. Subsumption errors are more complex. They can result from inaccurate input data due to errors in state and time extraction, or from the Agent's knowledge not fully covering all clinical scenarios, which can introduce bias when calibrating subsumption results.

5 Conclusion

In this paper, we propose representation method called the Clinical Status Pathway which describes the diagnosis and treatment process through the patient's state, enriching the pathway with more comprehensive diagnosis and treatment logic, thereby improving its effectiveness in application. Meanwhile, to mine clinical pathways from EMRs, we design a LLM-based pipeline method that automates the construction of clinical pathways based on the abstract admission process we define, thus reducing the time required for manual construction by doctors. Experimental results demonstrate that the CSP extracted using our method yield superior performance and are more effective on downstream tasks, thereby validating the effectiveness of our approach. We believe this work will offer new insights into clinical pathway construction and promote the development of LLM-based clinical pathway execution. Our goal is to establish an intelligent healthcare system based on LLMs and CSP to address the challenges and limitations of current clinical pathway systems.

Limitations

There are several limitations in extracting clinical status pathway from electronic medical records. Since the input data are textual, incorporating more multimodal data, such as patient imaging results, electrocardiograms, etc., could enhance the extracted clinical status pathway by providing more comprehensive patient information, thereby improving their effectiveness. Although we have the capability to use multimodal data as input, additional methods are required to mitigate the generation of LLM illusions and efficiency issues when processing multimodal data. Additionally, incorrect EMRs and the instability of LLMs within the extraction framework can result in inaccurate extraction outcomes, impacting downstream applications and causing fluctuations in experimental results. Consequently, manual review of the EMRs and careful selection of cue words are necessary to complete this task.

Ethical Considerations

Our EMR dataset was primarily sourced from a tertiary care hospital, with permission granted for data usage. All EMRs and codes were de-identified twice by the ethics committee and experts, in strict adherence to the guidelines. Ethical approval was also obtained from the partner hospital prior to submission.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (No. 62306112), Shanghai Sailing Program (No. 23YF1409400), and Shanghai Pilot Program for Basic Research (No. 22TQ1400100-20).

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Ayman Alahmar and Ola Alkhatib. 2022. Computerization of clinical pathways: A literature review and directions for future research. *arXiv preprint arXiv:2203.00815*.

Ayman Alahmar, Emad Mohammed, and Rachid Benlamri. 2018. Application of data mining techniques to predict the length of stay of hospitalized patients with diabetes. In *2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)*, pages 38–43. IEEE.

Emma Aspland, Daniel Gartner, and Paul Harper. 2021. Clinical pathway modelling: a literature review. *Health Systems*, 10(1):1–23.

Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2402.03216*.

You Chen, Joydeep Ghosh, Cosmin Adrian Bejan, Carl A Gunter, Siddharth Gupta, Abel Kho, David Liebovitz, Jimeng Sun, Joshua Denny, and Bradley Malin. 2015. Building bridges across electronic health record systems through inferred phenotypic topics. *Journal of biomedical informatics*, 55:82–93.

Marcelo Rosano Dallagassa, Cleiton dos Santos Garcia, Edson Emilio Scalabrin, Sergio Ossamu Ioshii, and Deborah Ribeiro Carvalho. 2022. Opportunities and challenges for applying process mining in healthcare: a systematic mapping study. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–18.

Kiarash Diba, Kimon Batoulis, Matthias Weidlich, and Mathias Weske. 2020. Extraction, correlation, and abstraction of event data for process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1346.

Cleiton dos Santos Garcia, Alex Meincheim, Elio Ribeiro Faria Junior, Marcelo Rosano Dallagassa, Denise Maria Vecino Sato, Deborah Ribeiro Carvalho, Eduardo Alves Portela Santos, and Edson Emilio Scalabrin. 2019. Process mining techniques and applications–a systematic mapping study. *Expert Systems with Applications*, 133:260–295.

Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2024. The faiss library. *arXiv preprint arXiv:2401.08281*.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu

Feng, Hanlin Zhao, et al. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.

Zhengxing Huang, Wei Dong, Peter Bath, Lei Ji, and Huilong Duan. 2015. On mining latent treatment patterns from electronic medical records. *Data mining and knowledge discovery*, 29:914–949.

Zhengxing Huang, Wei Dong, Lei Ji, Chenxi Gan, Xudong Lu, and Huilong Duan. 2014. Discovery of clinical pathway patterns from event logs using probabilistic topic models. *Journal of biomedical informatics*, 47:39–57.

Zhengxing Huang, Wei Dong, Lei Ji, Chunhua He, and Huilong Duan. 2016. Incorporating comorbidities into latent treatment pattern mining for clinical pathways. *Journal of Biomedical Informatics*, 59:227–239.

Zhengxing Huang, Xudong Lu, and Huilong Duan. 2013. Latent treatment pattern discovery for clinical processes. *Journal of medical systems*, 37:1–10.

Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, 78:15169–15211.

Uzay Kaymak, Ronny Mans, Tim Van de Steeg, and Meghan Dierks. 2012. On process mining in health care. In *2012 IEEE international conference on Systems, Man, and Cybernetics (SMC)*, pages 1859–1864. IEEE.

Andreas W Kempa-Liehr, Christina Yin-Chieh Lin, Randall Britten, Delwyn Armstrong, Jonathan Wallace, Dylan Mordaunt, and Michael O'Sullivan. 2020. Healthcare pathway discovery and probabilistic machine learning. *International journal of medical informatics*, 137:104087.

Jianquan Li, Xidong Wang, Xiangbo Wu, Zhiyi Zhang, Xiaolong Xu, Jie Fu, Prayag Tiwari, Xiang Wan, and Benyou Wang. 2023. Huatuo-26m, a large-scale chinese medical qa dataset. *arXiv preprint arXiv:2305.01526*.

Wei Li, Xin Min, Panpan Ye, Weidong Xie, and Dazhe Zhao. 2024. Temporal topic model for clinical pathway mining from electronic medical records. *BMC Medical Informatics and Decision Making*, 24(1):20.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Ronny Mans, Helen Schonenberg, Giorgio Leonardi, Silvia Panzarasa, Anna Cavallini, Silvana Quaglini, Wil Van Der Aalst, et al. 2008. Process mining techniques: an application to stroke care. In *MIE*, volume 136, pages 573–578.

Jorge Munoz-Gama, Niels Martin, Carlos Fernandez-Llatas, Owen A Johnson, Marcos Sepúlveda, Emmanuel Helm, Victor Galvez-Yanjari, Eric Rojas, Antonio Martinez-Millana, Davide Aloini, et al. 2022. Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, 127:103994.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Anastasiia Pika, Moe T Wynn, Stephanus Budiono, Arthur HM Ter Hofstede, Wil MP van der Aalst, and Hajo A Reijers. 2020. Privacy-preserving process mining in healthcare. *International journal of environmental research and public health*, 17(5):1612.

Ricardo Alfredo Quintano Neira, Bart Franciscus Antonius Hompes, J Gert-Jan de Vries, Bruno F Mazza, Samantha L Simões de Almeida, Erin Stretton, Joos CAM Buijs, and Silvio Hamacher. 2019. Analysis and optimization of a sepsis clinical pathway using process mining. In *Business Process Management Workshops: BPM 2019 International Workshops, Vienna, Austria, September 1–6, 2019, Revised Selected Papers 17*, pages 459–470. Springer.

Shenzhi Wang. 2024. Llama3-8b-chinese-chat (revision 6622a23).

Jasmine Tehrani. 2016. Computerisation of clinical pathways: Based on a semiotically inspired methodology. In *E-Health and Telemedicine: Concepts, Methodologies, Tools, and Applications*, pages 25–48. IGI Global.

Jan Niklas van Detten, Pol Schumacher, and Sander JJ Leemans. 2023. An approximate inductive miner. In *2023 5th International Conference on Process Mining (ICPM)*, pages 129–136. IEEE.

Xinyu Wang, Shaofei Su, Hao Jiang, Jiaying Wang, Xi Li, and Meina Liu. 2018. Short-and long-term effects of clinical pathway on the quality of surgical non-small cell lung cancer care in china: an interrupted time series study. *International Journal for Quality in Health Care*, 30(4):276–282.

Xiao Xu, Tao Jin, Zhijie Wei, and Jianmin Wang. 2017. Incorporating topic assignment constraint and topic correlation limitation into clinical goal discovering for clinical pathway mining. *Journal of healthcare engineering*, 2017(1):5208072.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, et al. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.

Wan-Shiou Yang and San-Yih Hwang. 2006. A process-mining framework for the detection of healthcare fraud and abuse. *Expert systems with Applications*, 31(1):56–68.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

A Electronic Medical Record Analysis

To understand the characteristics of our EMR, we analyzed our EMR dataset in terms of days of admission, instrument fields. Figure 5 illustrates the types and counts of clerical fields across all EMRs, while

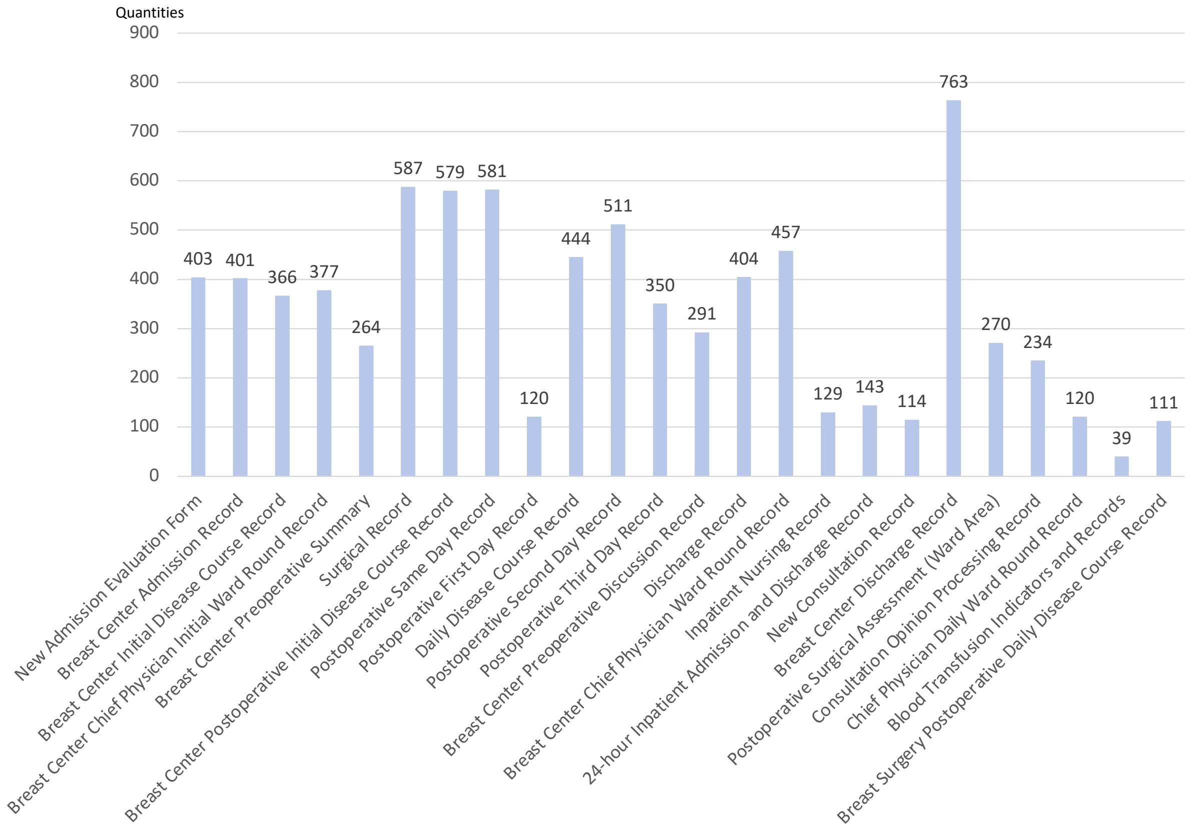

Figure 5: Analysis of Instrument Fields

Figure 6 depicts the corresponding maximum admission days and their frequencies. The filtered EMRs exhibit a wide range of instrument fields and admission days. This indicates that our dataset encompasses more comprehensive medical information, along with more intricate diagnosis and treatment logic, thereby enhancing the credibility of our extraction results.

B Implementation Details

B.1 CSP Representation

We identify several key factors as follows:

1. Chronology of the treatment process: In the course of treatment, each step is performed in a specific sequence and time frame to ensure the continuity and effectiveness of care. Thus, we establish the CSP based on the process sequence to ensure its effectiveness in standardizing medical practice and maintaining medical quality.

2. Clinical decision-making complexity: Doctors develop and implement optimal treatment strategies based on the patient's condition and medical history. This process is often nonlinear, involving complex decision-making logic. We enhance the expressiveness of CSP and improve their effectiveness in supporting treatment by incorporating state-to-state transfer logic.

B.2 Screening for Valid Information

The criteria used for screening valid information are outlined in Section 3.1, and following discussions with experts, we excluded the following fields:

1. Informed consent for hepatitis B program testing

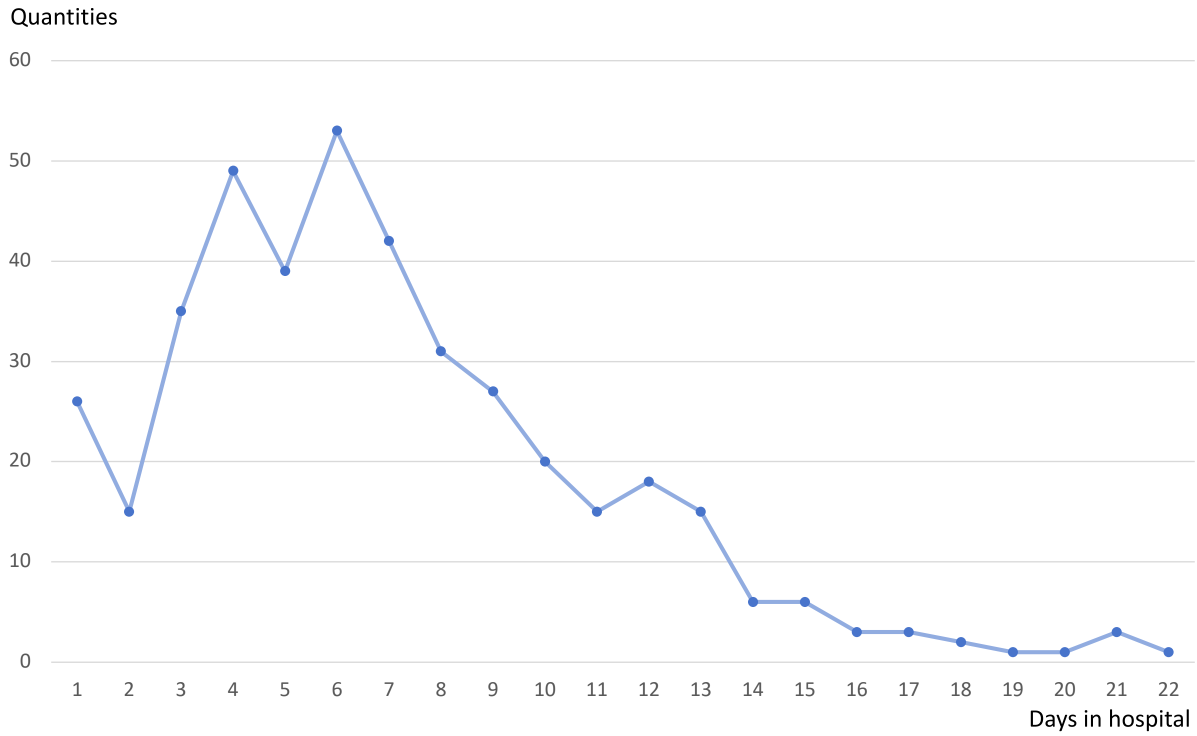

Figure 6: Analysis of Days of Admission

2. Informed consent for self-financed medications and medical consumables in the mammography department
3. Informed consent for frozen-section pathology examination
4. Admission notice
5. Patient authorization letter
6. Preoperative informed consent

B.3 Time Extraction Module

Figure 7 shows the system prompt we use for Time Extraction Module. The resulting temporal information was used as the basis for ordering the clinical status pathway.

```
Extract the "current time" information from the following electronic health record text.

Specific Requirements:
    Prioritize extracting the latest time in the medical record or the start time of the patient's visit.
    If there are multiple relevant times (such as examination, medication times), list all times in chronological order.
    Unify the time format to YYYY-MM-DD HH:MM in 24-hour format.

Example Text:
"On 2023-10-05 at 14:30, the patient complained of a headache with a body temperature of 37.8℃. A head CT scan was
performed at 15:00, and the results were normal."

Example Output:
["2023-10-05 14:30", "2023-10-05 15:00"]

Please process the following text:
{Input Text}
```

Figure 7: System prompt for Time Extraction

B.4 State Extraction Module

Figure 8 shows the system prompt we use for State Extraction Module. Specifically, we provide explicit definitions of the reasoning process for the LLM, the delineation of each state, and ultimately present an example alongside defining the output format.

To accurately extract key information from the provided electronic health record text and categorize it into Feature ,Diagnose, and Operation please follow these steps for analysis:
Step 1: Text Understanding
First, carefully read the text of the provided EMR to ensure a full understanding of the patient's condition, medical history notes, current symptoms, doctor's diagnosis, and recommended tests or treatments. As well as the time status of each piece of information, whether it occurred in the past or present.
Step 2: Extract Feature
Identify and list all descriptions of the patient's current state (such as symptoms, signs), past medical history, surgical history, or medication history. This information will be categorized under Feature.
Example: If the text mentions "headache", "body temperature 37.8℃", and "history of hypertension", these are all features of the patient.
Step 3: Extract Diagnose
Based on the Feature identified in Step 2, find any diagnostic conclusions explicitly stated in the text. These may include specific disease names or other medical terms.
Example: If the text states "preliminary diagnosis of tension headache", this is the diagnostic result related to the above features.
Step 4: Extract Operation
Based on the known Feature and Diagnose, identify all suggested examinations or treatment measures in the text. These recommendations aim to help manage or further evaluate the patient's condition.
Example: If the text suggests "take ibuprofen and monitor the condition", both pieces of advice fall under the operation category.
Step 5: Organize Output
Organize the information extracted in the previous three steps into the following dictionary format, where each key corresponds to a list to accommodate multiple related entries.
{ "Feature": ["headache", "body temperature 37.8℃", "history of hypertension"], "Diagnose": ["tension headache"], "Operation": ["take ibuprofen", "monitor the condition"] }

Figure 8: System prompt for State Extraction

B.5 Subsumption Algorithm

The pseudo-code for the Subsumption Algorithm is detailed in Listing 1.

B.6 Access to Clinical Pathways

We obtained clinical pathways for breast cancer specialties from the following websites. http://www.nhc.gov.cn/yzygj/s7659/202001/b3c9e097b0c1471a969d7a63be471759.shtml

B.7 Building the Knowledge Base

Since traditional clinical pathways are in natural language, we employed a semantic-based approach to segment and process the clinical pathways, then used the BGE model to vectorize them and store them in the FAISS vector library, thereby completing the construction of the clinical path knowledge base. The construction of the clinical status pathway knowledge base differs slightly, as it is in a structured form, specifically a tree-like data structure. Therefore, we traversed the clinical status pathway (CSP) using depth-first traversal, segmenting different paths into distinct chunks. The process of vectorization and storage is the same as that for clinical pathways. When using the knowledge base, we vectorize the query, calculate the cosine similarity with each chunk in the vector library, and select the five chunks with the closest similarity, along with the query, as inputs to the LLM.

B.8 Medical Q&A

Figure 9 shows the system prompt we use for Medical Q&A. Specifically, our query originates from Huatuo-26M, after which we consult the knowledge base for pertinent information serving as external knowledge according to this query; subsequently, we incorporate both elements into our prompt template.

You are now an experienced specialist doctor in breast cancer. Please think carefully and respond to the {Problem}. You may refer to the following information relevant to the question {CSP}.

Figure 9: System prompt for Medical Q&A

B.9 Assisted Decision-making

Figure 10 illustrates the Prompt used during the execution of the assisted decision-making task, whereas another Figure 11 depicts the Prompt employed for evaluating this task. Specifically, in the course of executing the task, it is necessary to extract the patient's information from the EMR, subsequently retrieve relevant knowledge pertaining to the patient from the knowledge base, and ultimately integrate both sets of information into the Prompt template. During the evaluation phase, responses from various knowledge bases were scored across four criteria. The EMR served as the baseline for these scores, against which the relative scores of the other knowledge bases were calculated.

You are now an experienced breast cancer doctor. Please predict the subsequent treatment process for the patient based on the patient's personal information {Text} and the clinical status pathway {CSP} corresponding to the patient. The treatment process should primarily include timelines and actions.

Figure 10: System prompt for Assisted Decision-making

You are an experienced breast cancer specialist. You have been provided with a patient's information as follows {Text}. For this patient, there are two proposed treatment pathways:
•**Pathway 1**: {}
•**Pathway 2**: {}
Please evaluate these two pathways based on four criteria: Accuracy, Completeness, Reasonableness, and Safety. Each criterion should be scored on a scale of 1 to 10 (with 10 being the highest). Here are the definitions for each criterion:
•**Accuracy**: Refers to the precision and reliability of the diagnostic tests or treatments administered for the specific condition and diagnosis of the patient, providing solid data support for subsequent medical decisions.
•**Completeness**: Emphasizes the seamless integration of all stages from admission to discharge, ensuring continuity and systematic coherence in the treatment process.
•**Rationality**: Focuses on achieving the expected therapeutic outcomes while balancing cost-effectiveness and efficiency, effectively controlling the growth of healthcare costs.
•**Safety**: Requires considering the patient's underlying conditions and special physical conditions during treatment, ensuring that the selected treatment plan maintains potential risks at a manageable level.
The final evaluation should be returned in a dictionary format as follows:
{ "Pathway 1": { "Accuracy": 8, "Completeness": 9, " Rationality ": 7, "Safety": 10 }, "Pathway 2": { "Accuracy": 7, "Completeness": 8, " Rationality ": 9, "Safety": 8 } }

Figure 11: System prompt for Assessment of Assisted Decision-making

```python

class ClinicalNode:
    def __init__(self, feature, diagnose, operation, order, days):
        self.feature = feature
        self.diagnose = diagnose
        self.operation = operation
        self.order = order
        self.days = days

def merge_clinical_paths(nodes):
    merged_nodes = recursive_merge(
        nodes = nodes,
        merge_condition = same_diagnose_order_diff_feature,
        merge_action = merge_features
    )

    if not agent_validation(merged_nodes, "Feature"):
        return handle_validation_failure(nodes, "Feature")
```

```python

    merged_nodes = recursive_merge(
        nodes = merged_nodes,
        merge_condition = same_feature_diagnose_order_diff_operation,
        merge_action = merge_operations
    )

    if not agent_validation(merged_nodes, "Operation"):
        return handle_validation_failure(merged_nodes, "Operation")

    final_nodes = merge_by_days(merged_nodes)

    if not agent_validation(final_nodes, "Final"):
        return handle_validation_failure(final_nodes, "Final")

    return final_nodes

def recursive_merge(nodes, merge_condition, merge_action):
    while True:
        found = False
        for i in sorted(nodes, key=lambda x: x.order):
            for j in sorted(nodes, key=lambda x: x.order):
                if i == j: continue
                if merge_condition(i, j):
                    new_node = merge_action(i, j)
                    nodes.remove(i)
                    nodes.remove(j)
                    nodes.append(new_node)
                    found = True
                    break
            if found: break
        if not found: break
    return nodes

def same_diagnose_order_diff_feature(node1, node2):
    return (node1.diagnose == node2.diagnose and
            node1.order == node2.order and
            node1.feature != node2.feature)

def same_feature_diagnose_order_diff_operation(node1, node2):
    return (node1.feature == node2.feature and
            node1.diagnose == node2.diagnose and
            node1.order == node2.order and
            node1.operation != node2.operation)

def merge_features(node1, node2):
    return ClinicalNode(
        feature = f"{node1.feature}+{node2.feature}",
        diagnose = node1.diagnose,
        operation = node1.operation,
        order = node1.order,
        days = min(node1.days, node2.days)
    )

def merge_operations(node1, node2):
    return ClinicalNode(
        feature = node1.feature,
        diagnose = node1.diagnose,
        operation = f"{node1.operation}+{node2.operation}",
        order = node1.order,
        days = min(node1.days, node2.days)
    )

def merge_by_days(nodes):
    return sorted(nodes, key=lambda x: x.days)

def agent_validation(nodes, stage):
    for _ in range(3):
        if llm_validate(nodes, stage):
            return True
```

```
        nodes = retry_merge(nodes, stage)
    return False

def handle_validation_failure(nodes, stage):
    error_report = {
        "error_stage": stage,
        "original_nodes": nodes,
        "correct_knowledge": llm_generate_correction(nodes)
    }
    log_error(error_report)
    return nodes
```

Listing 1: Pseudo-code for the merge algorithm