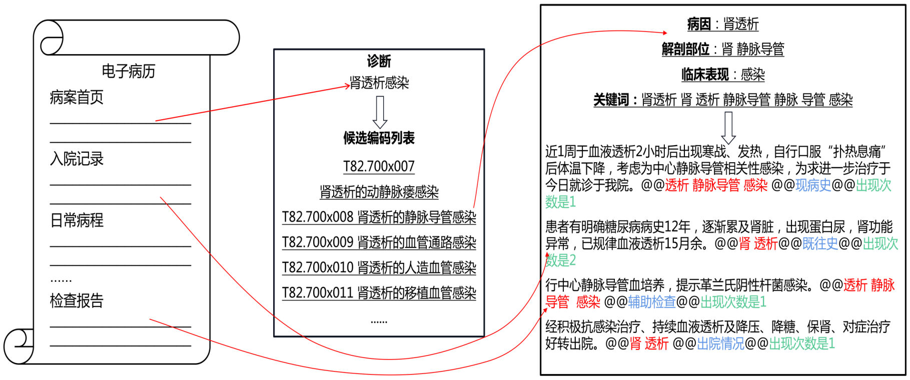
A Appendix

A.1 The Chinese version of Figure 3 of the paper


电子病历
病案首页
入院记录
日常病程
……
检查报告
诊断
肾透析感染
候选编码列表
T82.700x007
肾透析的动静脉瘘感染
T82.700x008 肾透析的静脉导管感染
T82.700x009 肾透析的血管通路感染
T82.700x010 肾透析的人造血管感染
T82.700x011 肾透析的移植血管感染
……
病因：肾透析
解剖部位：肾 静脉导管
临床表现：感染
关键词：肾透析 肾 透析 静脉导管 静脉 导管 感染

近1周于血液透析2小时后出现寒战、发热，自行口服“扑热息痛”后体温下降，考虑为中心静脉导管相关性感染，为求进一步治疗于今日就诊于我院。@@透析 静脉导管 感染 @@现病史@@出现次数是1

患者有明确糖尿病病史12年，逐渐累及肾脏，出现蛋白尿，肾功能异常，已规律血液透析15月余。@@肾 透析@@既往史@@出现次数是2

行中心静脉导管血培养，提示革兰氏阴性杆菌感染。@@透析 静脉导管 感染 @@辅助检查@@出现次数是1

经积极抗感染治疗、持续血液透析及降压、降糖、保肾、对症治疗好转出院。@@肾 透析 @@出院情况@@出现次数是1

Figure 1: A real example to show the candidate codes of one diagnosis and the reliable evidence under multi-axes for one of the candidate codes (also the correct ICD code), the example is from a real Chinese EMR. Each piece of evidence is followed by a few “@@” symbols for more information helpful for inference, including keywords, the appearing field in the medical record, and times of appearance.

A.2 Detailed Description of Evidence Retrieval

The evidence retrieval module consists of five steps:

1. Extraction of the diagnosis list: Firstly, we extract the diagnosis list from the Chinese electronic medical records. The diagnosis list in Chinese electronic medical records is well-structured and easily extractable.

2. Identification of candidate codes: For each diagnosis, we identify the top N candidate codes. We employ the edit distance, tf-idf, and word2vec-based similarity computation methods to capture the top N candidate codes among all the ICD codes, which are related to each diagnosis at the character, discrimination, and semantic levels, respectively. For the similarity-based method, we use word2vec to train the Chinese EMRs collected and obtain token representations, which are then combined to form the whole representation of the diagnosis and the ICD code. During the experiment, we set the number of candidate codes for each diagnosis to 50. The recall rate of the candidate ICD codes compared to the correct ICD codes for the training set, development set, and test set is 96.58%, 95.43%, and 95.55%, respectively. The recall rate increases with the increase in the number of candidate codes. For example, if the number is increased to 100, the recall rate can exceed 98%. Therefore, during replication, you can adjust the number of candidate codes based on the specific requirements.

3. Determination of retrieval keywords: We utilize a multi-axial parser to categorize the knowledge under each candidate code. Additionally, considering the effectiveness of rich synonym descriptions in ICD coding (as verified by the MSMN method, we incorporate synonyms for these keywords as supplementary information.

4. Greedy evidence retrieval based on keywords and prior knowledge: It is evident that evidence containing more keywords is considered to be more representative and reliable. Based on this, we sort all possible combinations of keywords in descending order according to the number of keywords and retrieve the sentences containing them from likely locations recorded in the prior knowledge. We replace sentences with fewer keywords with more evidence based on a greedy strategy. Once the supporting evidence for a keyword has been retrieved, the retrieval process does not continue for that particular keyword. We stop retrieving when all keywords have been traversed.

5. If you encounter a situation where the evidence retrieval for a specific keyword is unsuccessful in practice, you may consider expanding the retrieval scope to the entire medical record. This is due to

the varying quality of actual EMRs, which may not strictly adhere to standardized writing requirements. The feasibility of this approach depends on the quality of the Chinese electronic medical records available to you. In our actual experiments, such cases were relatively rare, accounting for approximately 15%.

A.3 Abbreviated Table of Prior Information for Potential Locations of Clinical Supporting Evidence for ICD Codes

Table 1: This is a small section of a complete prior knowledge table, listing the possible locations in medical records where clinically supporting evidence for three-digit code groups of interest may be found.

Code Range and name	Possible locations for evidence appearance in Chinese EMRs			
A00-A09 Intestinal infectious diseases	Testing (and results)	Symptom presentation	History of exposure to endemic or epidemic areas	
A15-A19 Tuberculosis	Symptom presentation	Examination (and results)	Chest examination results	Vaccination and Infectious Disease History
A20-A28 Certain zoonotic bacterial diseases	Testing (and results)	Occupation and working conditions		
A30-A49 Other bacterial diseases	Skin and mucous membrane examination results	Testing (and results)	Neurological reflex examination results	
A50-A64 Sexually transmitted infections	Skin and mucous membrane examination results	External genital examination results	Testing (and results)	
A65-A69 Other spirochetal diseases	Skin and mucous membrane examination results	Testing (and results)	History of exposure to endemic or epidemic areas	
A70-A74 Other diseases caused by chlamydia	Chest examination results	Head and facial examination results	Testing (and results)	
A75-A79 Rickettsioses	Testing (and results)	History of exposure to endemic or epidemic areas		
...	...	...	...	...