*2025 International Conference on Intelligent Computing*
*July 26-29, Ningbo, China*
https://www.ic-icc.cn/2025/index.php

# CvdKG: Cardiovascular Disease Knowledge Graph Construction with Cascading Pointer Networks

Yu Song[1], Bohan Yu[1], Dezhi Kong[1], Pengcheng Wu[1], Shuai Zhang[1], Xia Liu[1], Kejun Wu[1], Kunli zhang[1]*
1 Zhengzhou University, No. 100, Science Avenue, Zhengzhou, Henan, China
ybh18338462402023@163.com

**Abstract.** Cardiovascular disease is a common major chronic disease characterized by high mortality rate, and high difficulty in rehabilitation. This paper constructs a Cardiovascular diseases Knowledge Graph (CvdKG). According to the medical vocabulary and medical knowledge base, CvdKG has determined 15 entity types and 74 relationship types with diseases and operations as the core. The Cascading pointer network ($\mathrm{CASREL}_{\mathrm{pclBERT}}$) model suitable for the medical field is used to automatically extract knowledge from medical texts and manually proofread them. Knowledge fusion is carried out based on multi similarity weighting. The constructed CvdKG includes 217 core cardiovascular diseases, 8,845 related diseases, 433 surgeries, and 68,316 triples. CvdKG can provide data support for intelligent question answering and auxiliary diagnosis of cardiovascular diseases.

**Keywords:** Knowledge graph; Cardiovascular disease; Cascading pointer network

## 1 Introduction

Cardiovascular diseases are chronic diseases with high incidence rates and low control rates. In the disease mortality composition ratio of urban and rural residents in China, in 2020, they accounted for 48.00% and 45.86% of deaths in rural and urban areas, respectively, having a significant impact on the health and safety of the Chinese people. However, the public's awareness of this disease is only 51.5% [1], which is one of the main reasons for the high mortality rate of cardiovascular diseases. For the general public, understanding the knowledge of cardiovascular diseases helps in timely and effective treatment. On one hand, cardiovascular disease-related information is highly specialized; on the other hand, there is a lack of data to support intelligent applications for cardiovascular diseases.

The concept of the knowledge graph was formally proposed by Google in 2012 [2]. A knowledge graph organizes unstructured data resources into structured knowledge through techniques such as knowledge extraction, processing, and fusion. It consists of head entities, tail entities, and relationships. Depending on the domain of knowledge described in the graph, knowledge graphs can be divided into general knowledge graphs and domain-specific knowledge graphs. General knowledge graphs mainly focus on common knowledge and emphasize breadth, with typical examples being FreeBase [3], DBPedia [4], and Wikidata [5]. Domain-specific knowledge graphs are oriented toward knowledge in a particular field and emphasize depth, usually requiring the inclusion of fine-grained concepts for description. In the medical field, knowledge granularity needs to be controlled at levels such as diseases, drugs, symptoms, and other relevant hierarchical relationships.

The medical field is complex, with text data that is highly specialized, has complex relationships, and includes numerous types. Many scholars have worked on constructing medical knowledge graphs. Existing medical knowledge graphs have significantly improved standardization, formalization, and systematization compared to medical resources. However, due to the complexity of cardiovascular diseases, with treatment methods primarily based on surgery, current data cannot accurately describe these diseases.

This paper takes cardiovascular diseases and surgeries as the core and establishes a knowledge description framework. It uses a cascading pointer network-based automatic medical knowledge extraction model, combined with iterative extraction methods, to extract cardiovascular disease-related knowledge and construct a Cardiovascular Diseases Knowledge Graph (CvdKG).

## 2 Related Work

In the medical field, thanks to the rapid development of information technology and the widespread adoption of medical information systems, massive medical term sets and knowledge resources have accumulated in medical databases. Typical examples include ICD-10, ATC, and MeSH [6]. Among them, ICD-10 is the international disease classification code maintained by the World Health Organization (WHO); ATC is the anatomical, therapeutic, and chemical classification system maintained by the WHO, which includes 14 major anatomical concepts corresponding to drug components and their usage and dosage standards; MeSH is a medical subject heading list compiled by the U.S. National Library of Medicine, which includes 15 major categories. Medical term sets and other resources lay the foundation for constructing the knowledge representation system of knowledge graphs.

Currently, scholars around the world have established vast medical knowledge bases. For example, Aodema et al. [7] constructed a Chinese Medical Knowledge Graph (CMeKG), which includes 6,310 diseases and over a million medical concept relationship instances. Cheng et al. [8] constructed a stroke medical knowledge graph using a similarity-based method linked to existing knowledge graphs. Zhang et al. [9] built a Chinese pediatric epilepsy knowledge graph (CPeKG). Chen et al. [10] constructed a knee osteoarthritis knowledge graph. Zhang et al. created a personalized treatment knowledge graph for depression [11]. In these studies, most use a disease-centric knowledge representation system, while some scholars have also employed symptom- and drug-centric systems. Zan Hongying et al. [12] constructed a Chinese symptom knowledge base (CSKB) centered on symptoms. Current research is mainly focused on building knowledge graphs with disease, symptoms, drugs, etc., as the core. However, cardiovascular diseases involve a significant number of surgical treatments, requiring the introduction of a surgery-centric knowledge graph.

The main processes in knowledge graph construction are knowledge extraction, knowledge processing, and knowledge fusion, among which knowledge extraction is the key step that determines the efficiency and quality of graph construction. With the rise of deep learning, automatic extraction based on deep learning has become a hot topic in knowledge extraction research. Tan C et al. [13] proposed a boundary-aware neural network model to predict the category information of entities. Yuan Qi et al. [14] proposed a semi-automated knowledge graph construction method from semi-structured and unstructured data. Wei Z et al. [15] proposed a framework that models relationships as functions that map subjects to objects in sentences, thereby handling overlapping issues more accurately. Experimental results show that this framework is suitable for constructing cardiovascular knowledge graphs.

In summary, current medical knowledge graphs rarely adopt a surgery-centric knowledge representation system. Existing medical knowledge graphs are insufficient to describe the complex knowledge of cardiovascular diseases. The entity-relationship joint extraction model can improve the efficiency of knowledge extraction. Therefore, this paper focuses on the research of knowledge representation systems and knowledge extraction methods based on cascading pointer networks, and constructs a cardiovascular disease knowledge graph.

## 3 CvdKG Construction Process

The construction process of CvdKG is divided into four key steps: data collection and preprocessing, knowledge representation system construction, knowledge extraction, and knowledge fusion. The overall framework is shown in Figure 1.

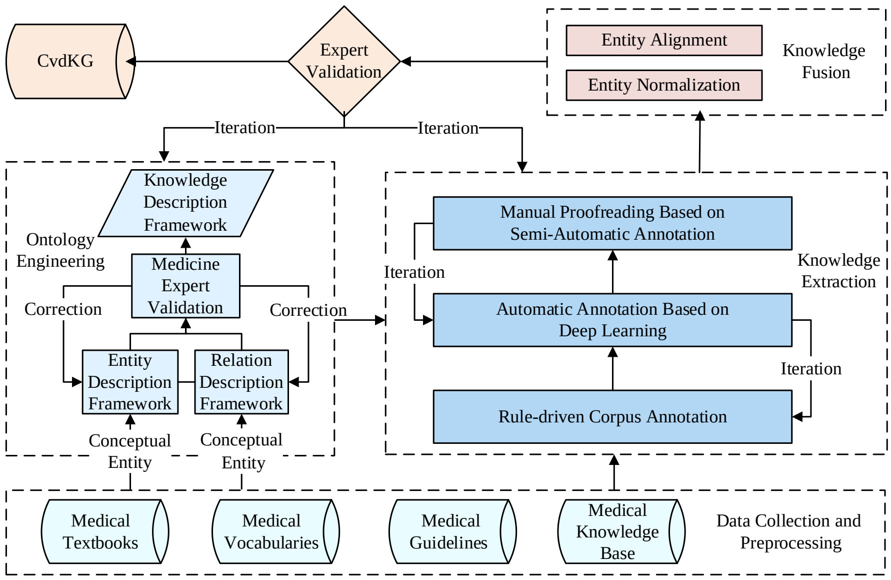


**Fig. 1.** CvdKG constructs the overall framework

Based on the knowledge representation system of CMeKG, a knowledge representation system is established by incorporating the characteristics of cardiovascular diseases and surgeries. A rule-based method is used to label training data, and deep learning techniques are employed to automatically extract knowledge from medical texts. Manual proofreading is performed in a semi-automatic annotation platform. Knowledge from multiple sources is fused based on similarity and visualized.

### 3.1 Entity Classification Framework

The professionalism and authority of the entity classification framework are crucial to the quality of medical knowledge graphs. Through the analysis of medical texts in the cardiovascular field and with guidance from professional doctors, an entity classification framework centered on diseases has been developed. This framework includes nine categories of entities: diseases, symptoms, epidemiology, examinations, sociology, drugs, anatomical sites, prognosis, and adjuvant therapies. According to "The 2021 White Paper on Cardiovascular Surgery and Extracorporeal Circulation Data in China", the number of cardiovascular surgeries exceeded 280,000[16], highlighting surgery as a critical treatment modality for cardiovascular diseases. Based on this, we propose an additional entity classification framework centered on surgical procedures, which includes six categories of entities: surgery, terminology, medical devices, explanations, surgical treatment, and surgical methods (steps).

### 3.2 Relationship Classification Framework

To provide a more fine-grained description of cardiovascular disease knowledge, this paper establishes a set of relationship descriptions tailored to entities within the cardiovascular domain. Examples include relationships such as <Disease, Symptom, Clinical Manifestation>, <Surgery, Method (Steps), Surgical Method (Steps)>, and others. The Cardiovascular Disease Knowledge Graph (CvdKG) defines a total of 73 types of relationships. Among these, several are particularly noteworthy.

Complications in cardiovascular diseases are complex [17] and can be categorized into three main types: disease-related complications, which are directly caused by the disease itself; surgery-related complications, which arise from surgical procedures and may involve damage, loss, or dysfunction of other tissues or organs; and drug-related complications, which are caused by the use or prolonged use of certain medications, leading to adverse effects on other tissues or organs. Under the guidance of professional physicians, these complications have been classified into three categories: Complication, Complication (Drug), and Complication (Postoperative).

Cardiovascular disease patients typically undergo a variety of medical examinations during diagnosis, such as physical examination, CT scans, electrocardiograms (ECG), and others. With input from medical professionals, these examinations have been categorized into six types: auxiliary examination, laboratory examination, imaging examination, pathological examination, endoscopic examination, and screening.

Cardiovascular surgeries are inherently complex. Medical guidelines contain extensive clinical recommendations for surgical procedures. Through analysis of these guidelines, it is evident that different types of patients (based on gender, age, disease stage, disease subtype, etc.) require tailored surgical approaches. Moreover, preoperative and postoperative examinations, as well as precautions, vary significantly [17]. Guided by professional physicians, the following relationships related to surgeries have been defined: method steps, which describe detailed procedural steps for performing the surgery; indications, which specify conditions under which the surgery is recommended; contraindications, which outline conditions under which the surgery should be avoided; preoperative preparation, which includes necessary preparations before the surgery; postoperative care, which involves follow-up care required after the surgery; and suitable patient types, which identify patient profiles for whom the surgery is appropriate. Figure 2 illustrates the knowledge description framework centered on diseases and surgeries, using "Acute Coronary Syndrome" and "Radiofrequency Ablation" as examples.

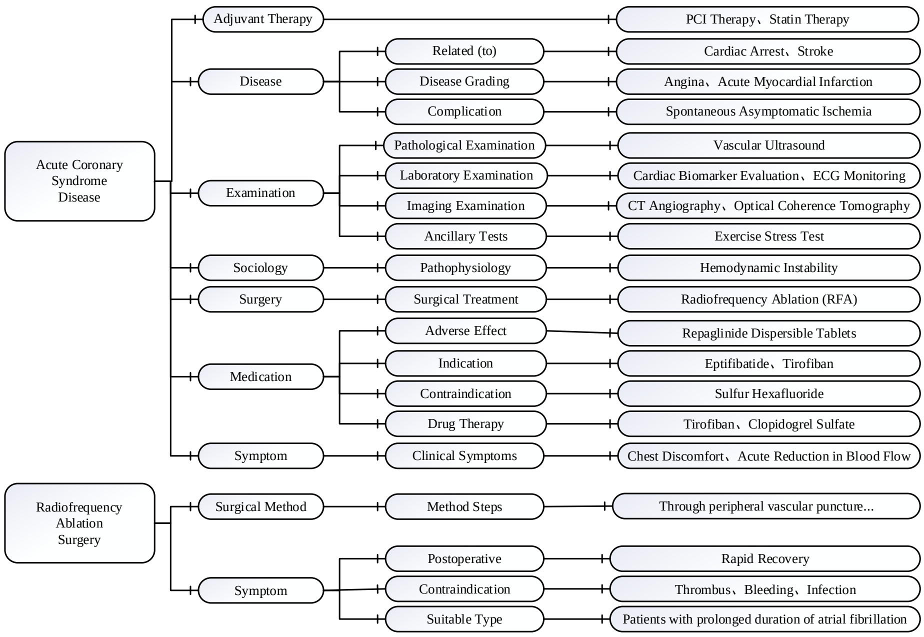


**Fig. 2.** Example of CvdKG relationship description

# 4 CvdKG Data Processing

## 4.1 Data Collection and Preprocessing

The construction of CvdKG uses four types of data: medical textbooks, medical vocabularies, medical guidelines, and medical knowledge bases. The characteristics and scale of the corpus used for CvdKG construction are shown in Table 1.

Under the guidance of professional doctors, cardiovascular knowledge and hierarchical relationships from the medical vocabulary are extracted and used as the knowledge definition standards and the foundational knowledge description framework. The knowledge obtained from the medical vocabulary is used as the standard for knowledge extraction. Structured triples and corpus with entity-relationship annotations are extracted from medical knowledge bases. Medical textbooks and guidelines are downloaded from medical websites, and the acquired text is processed with OCR, data cleaning, manual proofreading, and preprocessing to obtain formatted medical text.

**Table 1.** CvdKG basic corpus statistics

| Text Type | Data Characteristics | Data Scale |
|---|---|---|
| Medical Textbooks | Rich in knowledge about diseases, symptoms, drugs, surgeries, etc., with high authority. | 6,001,580 words |
| Medical Vocabulary | Highly authoritative, with publicly reliable data. | 11,457 terms |
| Medical Guidelines | Standardized protocols for evaluation or treatment, characterized by reliability, repeatability, clarity, and clinical applicability. | 247,358 words |
| Medical Knowledge Base | Structured text | 10,694 triples |

### 4.2 Rule-Based Training Corpus Annotation

In this paper, the existing cardiovascular disease-related triples from the medical knowledge graph are extracted, and their accuracy is confirmed by professional doctors. The extracted triples are then annotated in the medical text using methods such as Chinese word segmentation and text matching. The specific process is shown in Figure 3. A total of 12,350 training samples were obtained in this stage.

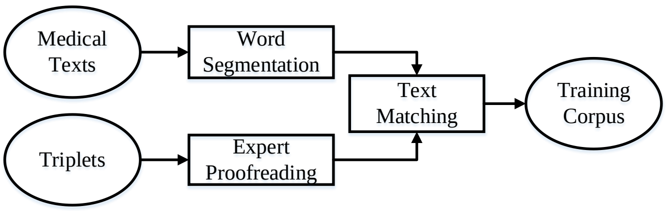


**Fig. 3.** Training corpus construction process

### 4.3 Cascading pointer network

Medical texts often involve multiple relationships centered around the same core entity (disease, surgery, drug, etc.). Therefore, the knowledge extraction task can be regarded as a "one-to-many" extraction and classification task. CvdKG uses CASREL as the knowledge extraction model, where the input is the text, and the output is all the triples <subject, predicate, object; s, p, o> in the text. The model structure is shown in Figure 4. The model first identifies all possible subjects (head entities); then, under the given category relationships, it identifies the objects (tail entities) related to the subject.

The shared encoding layer of CASREL uses the pre-trained BERT model [18]. The Pengcheng Laboratory has launched a pre-trained model suitable for Chinese medical texts, ness of pclBERT. To obtain a more accurate joint entity-relationship extraction model for cardiovascular diseases, this paper uses ness of pclBERT as the shared encoding layer of the CASREL model to capture contextual semantic information and represent words and characters. Additionally, ness of pclBERT's pre-trained encoding is used in the perception representation stage of the subject (s).

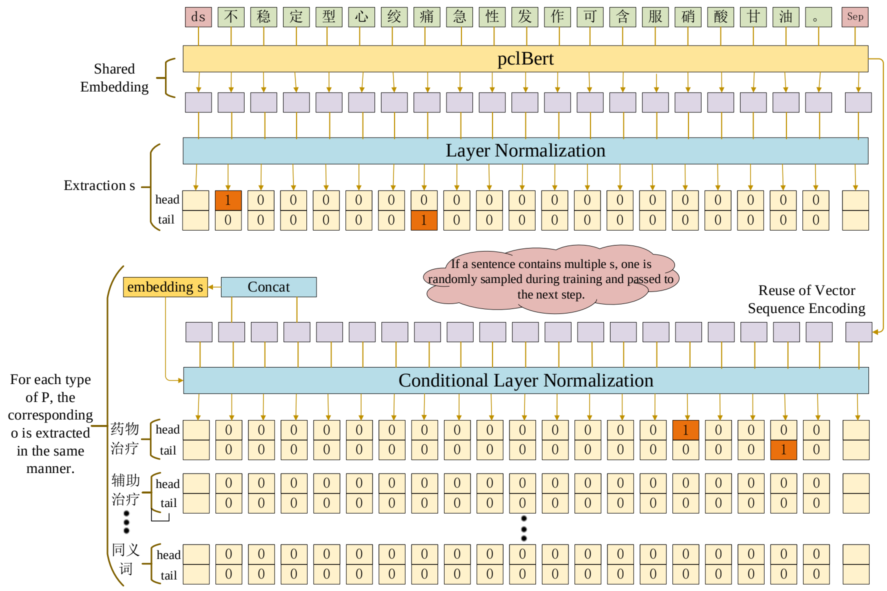


**Fig. 4.** CASREL entity relationship joint extraction model

In the head entity recognition stage, the input text is encoded by the shared encoding layer, and the resulting sequence is denoted as H. H is then passed into the Layer Normalization layer, where a "half-pointer, half-annotation" structure is used to predict the start and end positions of the head entity s. The calculation formulas are shown in Equations (1) and (2).

$$p_i^{start} = \sigma(W_{start}X_i + b_{start}) \tag{1}$$

$$p_i^{end} = \sigma(W_{end}X_i + b_{end}) \tag{2}$$

Here, $p_i^{start}$ and $p_i^{end}$ represent the probabilities that the i-th word is the start and end of the entity, respectively. $W_{(.)}$ and $b_{(.)}$ represent the trainable parameters and bias. If $p_i^{(.)}$ exceeds the threshold, it is set to 1; otherwise, it is set to 0. When the input is "Unstable angina acute attack can be treated with nitroglycerin." the model identifies the disease entity as "Unstable angina" The entity start 'Un' and entity end 'angina' are marked with 1, and the others are marked with 0. If multiple entities are predicted at this stage, the nearest matching principle is used to pair the identified start and end to obtain the candidate head entity set.

In the tail entity recognition stage, s is re-encoded through the shared encoding layer, and a sequence vector H′ of the same length as H is obtained. Both H and H′ are passed into the Conditional Layer Normalization for the perception representation of the subject $s$, where a "half-pointer, half-annotation" structure is constructed for each relation ppp to predict the corresponding start and end positions of the tail entity $o$. The calculation formulas are shown in Equations (3) and (4).

$$p_i^{start_o} = \sigma\left(W_{start_r}\left(X_i + V_s^{\,k}\right) + b_{start_r}\right) \tag{3}$$

$$\mathrm{p}_i^{end_o} = \sigma\left(W_{end_r}\left(\mathrm{X}_i + \mathrm{V}_s^{\,k}\right) + b_{end_r}\right) \tag{4}$$

Here, $p_i^{start\_o}$ and $p_i^{end\_o}$ represent the probabilities that the $i$-th word is the start and end of the tail entity under the relation encoding $r$, respectively. $V_s^{\,k}$ represents the encoding vector of the subject entity $s$. Similar to head entity extraction, $p_i^{(.)}$ is set to 1 if it exceeds the threshold; otherwise, it is set to 0. In the example sentence in Figure 4, the relationship "drug treatment involves the entity nitroglycerin", while no relevant entities exist for other relations such as "adjunctive treatment", "synonyms", etc. Therefore, the final output of the model is the triple <Unstable angina, Drug treatment, Nitroglycerin>.

There may be multiple entities in the text, i.e., $S = \{s_1, s_2, \ldots, s_n\}$. During training, one labeled entity $s$ is randomly sampled for perception representation. During prediction, all $s$ are traversed one by one, completing the perception representation for n entities.

### 4.4 Knowledge Proofreading and Iterative Extraction

In this study, the trained model is first applied to a small-scale unstructured text. After prediction, the extracted results are manually proofread. Once the accuracy of the data is confirmed, the results are merged with the original data and the model is retrained. Through multiple iterations, the model’s accuracy is gradually improved. During the manual proofreading phase, to enhance the accuracy of proofreading, this paper uses the entity-relationship annotation platform developed by Zhang et al. [19]. The triples extracted by the machine are pre-annotated in the text using text-matching techniques, improving the proofreading efficiency.

### 4.5 Experiment

This paper designs two sets of experiments: the performance experiment of the Cascading Pointer Network model and the consistency check between the automatically extracted results from random sampling during the iterative extraction process and the manually annotated results. The evaluation metrics used for the experiments are Precision (P), Recall (R), and F-Score (F).

Three datasets are used for the experiments, including the publicly available datasets NYT [14], WebNLG [14], and the self-constructed dataset CvdKG. The NYT dataset was originally generated using distant supervision, containing 1.18 million sentences and 24 predefined relation types. The WebNLG dataset was initially created for the NLG task, containing 5,019 sentences and 246 predefined relation types. CvdKG is the cardiovascular disease entity-relationship joint extraction dataset constructed in this paper, containing 12,350 sentences and 12 predefined relation types, such as diseases, drugs, symptoms, etc.

To verify the effectiveness of pclBERT, three different encoding layer settings of the Cascading Pointer Network were selected for comparison experiments. These include random initialization of all parameters in the BERT model, an LSTM-based pre-trained model framework, and a BERT-based pre-trained model framework. The experimental results are shown in Table 2. From the table, it can be seen that in CvdKG, the F-score of $\mathrm{CASREL}_{\mathrm{pclBERT}}$ increases by 3.3% compared to $\mathrm{CASREL}_{\mathrm{pclBERT}}$, indicating that $\mathrm{CASREL}_{\mathrm{pclBERT}}$ performs better in medical texts. The results in the public datasets also show that the model used can achieve good performance in entity-relationship joint extraction in the public domain.

**Table 2.** Experimental results of CASREL in NYT, WebNLG and CvdKG

| Model | NYT | | | WebNLG | | | CvdKG | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F | P | R | F | P | R | F |
| $\mathrm{CASREL}_{\mathrm{random}}$ | 81.5 | 75.7 | 78.5 | 84.7 | 79.5 | 82.0 | - | - | - |
| $\mathrm{CASREL}_{\mathrm{LSTM}}$ | 84.2 | 83.0 | 83.6 | 86.9 | 80.6 | 83.7 | - | - | - |
| $\mathrm{CASREL}_{\mathrm{BERT}}$ | 89.7 | 89.5 | 89.6 | 93.4 | 90.1 | 91.8 | 70.4 | 69.3 | 69.8 |
| $\mathrm{CASREL}_{\mathrm{pclBERT}}$ | 88.4 | 88.1 | 87.9 | 90.1 | 89.6 | 90.3 | **73.4** | **72.8** | **73.1** |

**Table 3.** Iterative extraction manual proofreading consistency comparison

| Number | Machine | Manual | Consistent Count | Consistency (%) |
|---|---|---|---|---|
| 1 | 162 | 172 | 157 | 91.28 |
| 2 | 106 | 113 | 106 | 93.81 |
| 3 | 128 | 126 | 124 | 96.88 |
| 4 | 220 | 223 | 218 | 97.76 |
| 5 | 145 | 147 | 144 | 97.79 |

### 4.6 Knowledge Fusion

This paper uses entity alignment and entity normalization for fusion. The triples obtained during the knowledge extraction phase are fused by calculating entity similarity using text similarity methods, and the final entity form is selected based on the medical vocabulary. The text similarity methods mainly include attribute similarity functions, structural similarity functions, and similarity functions based on edit distance.

In the actual knowledge base entity alignment process, the similarity between two entities $e_i$ and $e_j$ is defined as shown in Equation (5). $sim(e_i, e_j)$ represents the attribute similarity function, and $sim_{NB}(e_i, e_j)$ represents the structural similarity function. $M_{i,j}$ is the similarity function based on edit distance, which is calculated by the minimum number of editing operations required to convert between two strings. The basic operations include swapping positions, insertion, replacement, etc. This similarity function can effectively handle sensitive issues. A commonly used similarity function based on edit distance is the Levenshtein distance [21].

$$SIM = \theta_1 sim(e_i, e_j) + \theta_2 sim_{NB}(e_i, e_j) + \theta_3 M_{i,j} \tag{5}$$

$$\mathrm{M}_{i,j}=\begin{cases}\mathrm{M}_{i-1,j-1}, \\ \text{if } \mathrm{s}_{1,i}=\mathrm{s}_{2,j}; \\ 1+\min\{\mathrm{M}_{i-1,j},\mathrm{M}_{i,j-1},\mathrm{M}_{i-1,j-1}\}, \\ \text{otherwise}\end{cases} \tag{6}$$

The Levenshtein distance between two strings $S_1$ and $S_2$ is computed using dynamic programming: The algorithm initializes a matrix M of size $(|S_1|+1)\times(|S_2|+1)$, where the element at the i-th row and j-th column of M is denoted as $M_{i,j}$. The values of $M_{i,j}$ for $0 \ll i \ll |S_1|$ and $0 \ll j \ll |S_2|$ can be computed using Equation (6).

This paper uses the aforementioned three methods to calculate the similarity scores between two entities. Through multiple experiments, the weights $\theta_1$, $\theta_2$, and $\theta_3$ for the three methods are determined to be 0.02, 0.49, and 0.49, respectively, with an entity similarity threshold of 0.85. Examples of entity alignment and normalization are shown in Table 4.

**Table 4.** Example of entity alignment and normalization

| Standard Entity | Similar Entities |
|---|---|
| Hemoglobin elevation | Hemoglobin elevation \| Hemoglobin increase |
| Electrophysiological examination | Electrophysiological examination \| Electrophysiological test |
| Left atrial enlargement | Left atrial enlargement \| Left heart enlargement |
| Heart failure | Heart failure \| Cardiac failure |

## 5 Knowledge Graph Analysis and Visualization

This study constructs a knowledge graph for cardiovascular diseases, which includes 15 types of entities centered on diseases and surgeries, along with 73 types of relationships. Detailed information about the entities is presented in Table 5. According to the statistical results, CvdKG contains 9,062 diseases and 433 surgical procedures. Among these, 217 diseases are classified as core cardiovascular diseases based on the ICD-10 standard. The distribution of relationships within CvdKG is shown in Figure 5, with the most frequent relationship types including adverse reactions, clinical manifestations, drug treatments, and etiology. These findings align with the characteristics of cardiovascular diseases, which are diverse and involve complex features such as symptoms, causes, diagnostic tests, and drug treatments.

**Table 5.** CvdKG entity quantity statistics

| Entity Type | Number | Entity Type | Number |
|---|---|---|---|
| Disease | 9,062(217) | Terminology | 785 |
| Symptoms | 7,353 | Adjuvant Therapies | 621 |
| Drugs | 6,514 | Explanations | 502 |
| Sociology | 4,771 | Surgical Treatments | 490 |
| Examination | 2,140 | Other Categories | 434 |
| Anatomical Sites | 1,734 | Prognosis | 165 |
| Other Treatments | 1,139 | Medical Devices | 138 |
| Surgery | 433 | Surgical Methods (Steps) | 97 |

Through the analysis of entities and relationships in CvdKG, experts have confirmed that the knowledge graph provides a meaningful and structured description of cardiovascular disease knowledge. This structured representation has significant potential to support clinical decision-making, research, and education in the field.

To visualize the construction results of CvdKG, this study employs Echarts, primarily showcasing its network structure. For each entity, the visualization highlights triples where the entity serves as the subject. Nodes connected to the same entity are color-coded to represent identical semantic relationships. An example of this visualization is provided in Figure 6. This approach enables an intuitive understanding of the relationships between entities and their semantic connections, further enhancing the usability and interpretability of the knowledge graph.

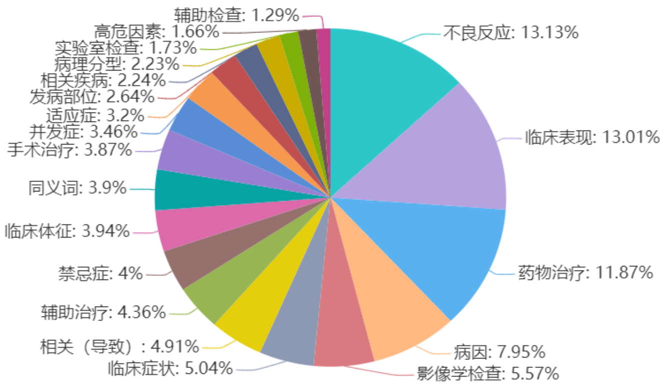


**Fig. 5.** CvdKG relation distribution

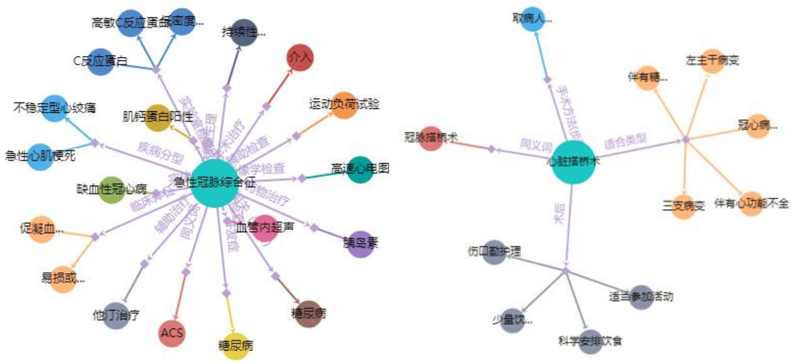


**Fig. 6.** CvdKG sample -Acute coronary syndrome

## 6 Conclusion

Based on the existing framework for medical knowledge graph descriptions, this study develops a knowledge description framework centered on diseases and surgeries and constructs the CvdKG. A medical text automatic extraction model, $\mathrm{CASREL}_{\mathrm{pclBERT}}$, is proposed based on the cascaded pointer network for joint entity-relation extraction. Experiments are designed to validate the effectiveness of this method in extracting knowledge within the cardiovascular disease domain. Through iterative extraction and knowledge fusion, the construction and visualization of CvdKG are completed. CvdKG can provide data support for applications such as intelligent question-answering systems and computer-aided diagnosis for cardiovascular diseases. Moreover, the methodology used in constructing the knowledge graph in this study is transferable to other diseases. In future research, efforts will continue to improve the automated construction methods for medical knowledge graphs and enhance the accuracy of the automatic extraction model.

## References

1. Report on Cardiovascular Health and Diseases in China 2022: an Updated Summary [J]. Chinese Circulation Journal, 2023,38(06):583-612.
2. Pujara J, Miao H, Getoor L, et al. Ontology-aware partitioning for knowledge graph identification[C].Proceedings of the 2013 workshop on Automated knowledge base construction. 2013: 19-24.
3. K. Bollacker, C. Evans, P. Paritosh et al., "Freebase: a collaboratively created graph database for structuring human knowledge, "in Proceedings of the of the 2008 ACM SIGMOD Int'l Conf. on Management of Data, pp. 1247–1250, ACM Press, Vancouver, BC, Canada, June 2008.
4. C. Bizer, J. Lehmann, G. Kobilarov et al., "DBpedia - a crystallization point for the Web of Data, " [J].vol. 7, no. 3, pp. 154–165, 2009.
5. Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85.
6. Nelson S J, Johnston W D, Humphreys B L. Relationships in medical subject headings (MeSH)[M]. Relationships in the Organization of Knowledge. Springer, Dordrecht, 2001: 171-184.
7. Odmaa B, Yang yunfei, SUI Zhifang, et al. Preliminary Study in the Construction of Chinese Medical Knowledge Graph [J]. JOURNAL OF CHINESE INFORMATION PROCESSING, 2019, 33(10): 1-7.
8. Cheng B, Zhang J, Liu H, et al. Research on medical knowledge graph for stroke[J]. Journal of Healthcare Engineering, 2021, 2021.
9. CAO Mingyu, LI Qinging, YANG Zhihao, et al. A Question Answering System for Primary Liver Cancer Based on Knowledge Graph [J]. JOURNAL OF CHINESE INFORMATION PROCESSING, 2019, 33(6): 88-93.
10. Chen Y, Chen X, Guo Y, et al. Knee Osteoarthritis Knowl-edge Graph: Driving Real-World Evidence-Based ClinicalDecision Support[J]. IEEE Journal of Biomedical and Hea-lth Informatics,2020,24(2):471-480

11. Zhang Y, Ding Y, Song Y, et al. Construction of depression-specific knowledge graph for discovering personalized depression treatment[J].Journal of biomedical informatics, 2020.102: 103358
12. ZAN Hongying, HAN Yangchao, FAN Yaxin, et al. Construction and Analysis of Symptom Knowledge Base in Chinese [J]. JOURNAL OF CHINESE INFORMATION PROCESSING, 2020, 34(4): 30-37.
13. Tan C, Qiu W, Chen M, et al. Boundary enhanced neural span classification for nested named entity recognition[C].Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(05): 9016-9023.
14. Yuan Q, Liu Y, Xie ZP, et al. Semi-automatic Construction Method of Pet Knowledge Graph[J]. Application Research of Computers, 2020, 37(01): 178-182.
15. Wei Z, Su J, Wang Y, et al. A novel cascade binary tagging framework for relational triple extraction[J]. arXiv preprint arXiv:1909.03227, 2019.
16. ZHAO Jv, HEI Feilong, et al. White book of Chinese cardiovascular surgery and extracorporeal circulation in 2021 [J]. CHINESE JOURNAL OF EXTRACORPOREAL CIRCULATION, 2022, 20(4):4.
17. QU Xiaona, Serious Complications Analysis After Interventional Therapy for Cardiovascular Diseases [J]. Smart Healthcare, 2022(021):008.
18. Sun F, Liu J, Wu J, et al. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer[C]. Proceedings of the 28th ACM international conference on information and knowledge management. 2019: 1441-1450.
19. ZHANG Kunli, ZHAO Xu, GUAN Tongfeng, et al. A Platform for Entity and Entity Relationship Labeling in Medical Texts [J]. JOURNAL OF CHINESE INFORMATION PROCESSING, 2020, 34(6): 117-125.
20. Artstein R, Poesio M. Inter-coder agreement for computational linguistics[J]. Computational linguistics, 2008, 34(4): 555-596.
21. Bordes A,Usunier N,Garcia-Duran A,et al.Translating embeddings for modeling multi-relational data[C].Neural Inform ation Processing Systems (NIPS).2013:1-9.
22. Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., et~al.: Qwen technical report. arXiv preprint arXiv:2309.16609 (2023)