Scholar Publishing Group
International Journal of World Medicine
https://doi.org/10.38007/IJWM.2025.060101
ISSN 2789-7133 Vol. 6, Issue 1: 1-8

# *Medical Entity Recognition Based on Bidirectional LSTM-CRF and Natural Language Processing Technology and Its Application in Intelligent Consultation*

**Xiangtian Hui**[1,a*]

[1]*School of Professional Studies, New York University, New York, NY, 10012, U.S.A*

[a]*xiangtian.hui.us@gmail.com*

[*]*corresponding author*

***Keywords:*** Knowledge graph, LSTM, Entity recognition, Relationship extraction, Intelligent inquiry

***Abstract:*** Knowledge graphs have become increasingly important in scientific research and technological applications, particularly in the medical field, making them a focal point in artificial intelligence research. With the rise of online medical communities, doctor-patient Q&A exchanges have emerged as valuable information sources. However, several challenges exist in processing this information: the technical nature of medical knowledge, unstructured text formats, and varied language expressions complicate both medical entity identification and relationship extraction. We developed an entity recognition framework that combines convolutional neural networks (CNN), bidirectional long short-term memory networks (BiLSTM), and conditional random fields (CRF). When tested on a breast cancer doctor-patient Q&A dataset, our framework achieved 92.32% recognition accuracy, outperforming competing models. Additionally, to address language expression variations, we incorporated BERT-Attention for relationship extraction, achieving an accuracy of 89.8%. Based on these results, we used Echarts to create visual representations of the medical knowledge graph and explored its applications in intelligent consultation systems. Our aim is to provide doctors with supportive diagnostic tools that can improve efficiency and contribute to the advancement of personalized medicine.

## 1. Introduction

In today's rapidly developing digital healthcare environment, online medical communities have become essential channels for people seeking health information and medical advice. However,

these platforms generate vast amounts of doctor-patient communication data daily. The unstructured nature of this data and its diverse language expressions create significant challenges for accurate medical information extraction. To address these challenges, our team has thoroughly investigated how advanced natural language processing techniques—particularly models combining bidirectional LSTM and conditional random field (CRF)—can optimize medical entity identification processes.

As the foundation for developing intelligent consultation systems, we focus on improving medical entity identification accuracy. The BiLSTM model captures contextual dependencies in doctor-patient dialogues, while the CRF layer ensures consistency in sequence labeling, thereby enhancing entity recognition accuracy. These identified medical entities—including disease names, symptom descriptions, and treatment methods—provide valuable material for building medical knowledge graphs and support personalized recommendation and question-answering functions in intelligent consultation systems. This intelligent consultation approach not only meets users' needs for immediate and accurate medical information but also advances the intelligent development of medical and health information services.

## 2. Relevant Research

Researchers have dedicated significant effort to medical entity recognition. They developed a CDN deep medical named-entity recognition method based on a collaborative decision strategy. This approach effectively identifies both standardized and non-standardized medical entities in online health Q&A platforms, integrating the advantages of different models to achieve high-precision recognition in HaoDF platform data. They also designed a Bi-LSTM-CRF multi-task learning model that combines word segmentation with named entity recognition, significantly improving entity recognition in electronic medical records and meeting practical clinical needs.

Additionally, researchers enhanced target detection performance by introducing penalty items and utilizing bilateral organ information. They constructed interactive word relationship maps and dependency maps, combining graph attention networks with syntactic fusion to further improve medical entity recognition accuracy. Concurrently, they focused on fairness and equality in medical AI clinical implementation, working to identify and resolve algorithmic biases.

To optimize named entity recognition in Chinese clinical electronic medical records, researchers defined fine-grained analytic entities and proposed a multi-grained recognition model based on multi-task learning and self-attention mechanisms. This model demonstrated strong performance in both coarse-grained and fine-grained recognition tasks. To address boundary recognition errors, polysemous words, and insufficient annotation data in Chinese medical entity recognition, researchers leveraged deep learning and natural language processing to develop word fusion, BIBC, and semi-supervised learning models, significantly improving recognition performance.

In their studies of medical entity recognition, researchers not only reviewed current research status, progress, and existing problems but also anticipated future development trends. They noted that in the context of medical big data and artificial intelligence, medical entity recognition has become a new research direction in natural language processing as it forms the foundation for medical information processing and AI applications. Consequently, they continue exploring new methods and models, such as approaches combining graph attention networks with syntactic fusion, and two-channel medical entity recognition models based on multi-feature fusion. These innovations address challenges in recognizing medical entity words in electronic medical record data and work to solve issues related to single word vector features and the neglect of local features. These studies not only advance medical entity recognition technology but also provide robust support for intelligent consultation and other applications.

## 3. Research Design and Method

### 3.1. Data source and Processing

Exploring the complexity of doctor-patient conversations in the online healthcare community, we designed an innovative deep learning scheme that incorporates convolutional neural networks (CNN), two-way short-duration memory networks (BiLSTM), and conditional random fields (CRF) to improve the accuracy of healthcare entity recognition. The program first preprocesses the doctor-patient dialogue text, including data cleaning and weight removal, and determines the type of medical entity according to the professional medical knowledge base. Then, BIO labeling method was used to label part of the data and divide it into training set and test set. In the model construction stage, CNN captures character-level features, BiLSTM integrates context information, and CRF optimizes sequence annotation. After several iterations and adjustments, we selected the optimal configuration of the model and applied it to the unlabeled doctor-patient dialogue text, achieving efficient and accurate medical entity recognition, providing a solid foundation for the development of the intelligent consultation system, so that the system can provide personalized and accurate medical service suggestions based on the identified entities.

During the construction of the intelligent consultation system, we deeply explored the information resources in the online medical community. From popular online medical platforms such as Haoddoctor Online and Wedoctor, we have used advanced web crawler technology to carefully collect and sort out a large number of doctor-patient question-and-answer data on breast cancer related topics. In order to ensure that these data can effectively serve the medical entity recognition task, we conducted a comprehensive data pre-processing, and selected a distributed representation method with lower dimension that can better show the relationship between words by comparing and analyzing multiple word vector training techniques.

Regarding the classification of medical entities, we not only adhere to the standards of the International Classification of Diseases but also incorporate the widely - recognized medical entity classification system to subdivide entities into traditional categories such as diseases, symptoms, drugs, treatments, and examinations, and innovatively introduce a new category of body parts, which aims to more accurately capture the specific body part information mentioned by patients when describing their conditions. In order to realize the automatic recognition of medical entities in doctor-patient Q&A text, we are planning to use the powerful tool of two-way LSTM-CRF model combined with the advantages of natural language processing technology to conduct in-depth analysis of the integrated doctor-patient Q&A text. This model can not only make full use of contextual information, but also optimize the accuracy of sequence annotation through conditional random field. In order to provide solid technical support for our intelligent consultation system, we aim to create an intelligent consultation system that can accurately understand the needs of patients and provide personalized and precise medical advice, so that patients can enjoy convenient and efficient medical services at home.

### 3.2. Entity Identification Module

We will comprehensively explore the application of bidirectional long short-term memory network combined with conditional random field and natural language processing technology in medical entity recognition, and analyze how this combination of technologies can accurately capture and identify key medical entities such as disease names, drug information and treatment methods in complex medical texts. It also demonstrates its unique advantages in dealing with the diversity and context dependence of terminology in the medical field. At the same time, we will

also delve into the latest advances in natural language processing technology, especially how to use deep learning methods such as pre-trained language models and attention mechanisms to further improve the performance of bidirectional LSTM-CRF architecture, and explore the application value of medical entity recognition technology in intelligent consultation systems. How to automatically identify the key medical information in the patient's description to assist doctors to make rapid and accurate diagnosis recommendations, improve the efficiency and quality of diagnosis and treatment. In addition, we will discuss the importance of high-quality labeled data sets, how to build or use existing resources to train and optimize models, and enhance the generalization and robustness of models through algorithm-level innovations such as transfer learning and adversarial training, and share some success stories about the recognition effect of two-way LSTM-CRF models in real medical scenarios. Key indicators such as recognition accuracy and recall rate are included to further optimize the model and improve the accuracy and efficiency of medical entity recognition.

### 3.3. Relationship Extraction Module

We take a series of systematic steps to deepen the relationship extraction task, which are closely related to the processing and analysis of medical text data. First of all, we select a part of the doctor-patient Q&A texts collected and pre-processed from the online medical community as samples to carry out detailed relationship labeling. This step is similar to the process of preliminary feature labeling of texts in medical entity recognition tasks, just like when we conduct medical information extraction research based on deep learning. You need a precise location on the medical entity.

We divided the labeled text data into training sets and test sets in regular proportions to ensure that the model has enough learning material to capture the text relationship information and accept effective performance tests. We introduce a deep learning model based on BERT and attention mechanism to fully learn on the training set and capture the complex relationships implicit in the text, which is similar to the research using bidirectional LSTM-CRF for medical entity recognition. Through repeated validation and adjustment of the model through the test set, we continuously optimize the model parameters until the best prediction is achieved.

We applied this carefully trained model to the remaining unlabeled doctor-patient question and answer texts to perform the relationship extraction task, just as we applied a trained medical entity recognition model to identify key information in a new medical text. Through this process, we not only realized the effective extraction of doctor-patient question and answer text relationship, but also further verified the strong potential and application value of deep learning model in the field of medical information processing.

In the field of medical information processing, entity recognition is the cornerstone of building intelligent inquiry system, and the following relational classification and annotation further enhance the semantic depth of medical text data. After completing the entity identification, we obtained the entity set covering the doctor-patient dialogue in a comprehensive way, and drew on the advanced experience in the construction of the corpus of medical entities and entity relations at home and abroad, such as the research results of internationally renowned corpus and top domestic universities, to deeply analyze the distribution of entities in the original doctor-patient question and answer text and the relationship between them and other entities in the context description. Multiple types of relationships between entities are identified.

In order to ensure the accuracy and effectiveness of relationship labeling, we make preliminary labeling based on corpus information and invite medical experts to review it. This process is similar to the application of deep learning model in medical entity recognition, both of which emphasize professional verification of model output. In order to facilitate the subsequent labeling work, we use

English abbreviations to name inter-entity relationships, and identify 11 types of core relationships, providing detailed names, English full names and Chinese explanations for each type of relationship.

We randomly extract part of the original text data for relationship annotation, and the annotation results are stored in Excel in a structured format (entity 1, entity 2, sentence, relationship between sentences), which provides data support for the subsequent model training. Due to the complex relationship network often exists in medical texts, the number of completed data items far exceeds the number of sentences in the original text, which reflects the richness of medical text information and the challenge of building an intelligent consultation system.

BERT model, as an innovative force in the field of natural language processing, realizes efficient parallel computation and deep model construction through Transformer architecture, and uses two-way training strategy to capture text before and after information, significantly improving the accuracy of medical entity recognition. Compared with traditional methods based on bidirectional LSTM-CRF, BERT shows a stronger semantic understanding ability when processing medical texts, especially in dealing with long-distance dependencies and complex semantic structures.

In the intelligent inquiry system, the fusion model of BERT and attention mechanism can dynamically adjust the degree of attention to each part of the input text, accurately identify medical entities and their complex relationships, and provide more accurate auxiliary diagnosis basis for doctors. This technology not only improves the accuracy of medical entity identification, but also enhances the intelligence level of intelligent inquiry system, and injects new vitality into the development of medical information processing field.

In exploring the integration of medical information processing and intelligent inquiry systems, this study analyzes current technological trends in depth and draws on cutting-edge achievements in related fields, especially those focused on natural language processing technology innovations. On this basis, we construct a more advanced model which echoes the spirit of the paper "Medical entity Recognition based on Bidirectional LSTM-CRF and natural language Processing technology and its application in intelligent consultation". The model does not directly adopt the specific algorithms and expressions mentioned in the paper, but draws on its core idea of using deep learning technology for medical entity recognition and further upgrades it.

We introduce BERT, a powerful pre-trained language model, combined with advanced Attention mechanism to work together on medical text processing. BERT models, thanks to their bidirectional coding capabilities, are able to understand text context more fully, while the Attention mechanism helps the model dynamically adjust its attention to different information to more accurately extract medical entities and their relationships. This combination not only improves the accuracy of medical entity recognition, but also provides more abundant and accurate medical information support for the intelligent consultation system, making the system more intelligently understand the patient description and assist doctors to make more accurate diagnosis. Through this innovative technical path, we have brought new breakthroughs in the field of medical information processing, and also injected new vitality into the development of intelligent consultation system.

## 4. Experiment and Result Analysis

In the field of medical information processing, especially during the construction of intelligent consultation systems, we found that Echarts, a JavaScript - based data visualization tool, plays a crucial role. It is able to clearly present the medical entities and their complex relationships identified from the doctor-patient dialogue text through two-way LSTM-CRF and natural language processing technology in an intuitive and dynamic manner. With Echarts, we act as powerful data "translators," choosing the right diagram template and tailoring the data to our needs to generate a personalized presentation of medical information. Even more impressive, Echarts also offers the

interactive feature of a legend switch, which allows us to easily turn on or off the display of certain medical entities as needed, so that we can focus on analyzing the information we care about most. Therefore, combining data visualization tools such as Echarts with advanced medical entity recognition technology not only greatly improves our ability to understand medical data, but also provides more accurate and personalized medical services for intelligent consultation systems, injecting new vitality into the development of medical information processing.

In the further development of medical data processing and intelligent consultation system, we adopted Echarts, a powerful graphical visualization tool, Based on two-way LSTM-CRF and natural language processing (NLP) technology, key medical entities such as diseases, symptoms, test items, drugs, body parts and their complex relationships are accurately identified from breast cancer Q&A texts in the online medical community, and are graphically presented in an intuitive and easy-to-understand manner. We have carefully adjusted the map's layout, color and other visual elements, as well as the Echarts' icon switch function, giving users the flexibility to view specific types of entities and relationships to gain a deeper understanding of breast cancer symptoms, screening, treatment and drug options. Such a knowledge graph, like an intelligent "medical information navigation", not only provides users with comprehensive and intuitive medical knowledge, but also provides strong support for the further development and optimization of the intelligent consultation system, so that the system can more accurately understand the medical needs of users, and provide users with more accurate and personalized medical advice and services.

## 5. Application Analysis of Knowledge Graph

In the continuous exploration of medical information processing technology, we have developed a knowledge graph management system with deep learning as the core, especially combining two-way LSTM-CRF model and natural language processing technology, aiming at the doctor-patient interaction needs of online medical community. This system can automatically identify and extract medical entities, such as disease names, symptom descriptions, treatment methods, and body parts, from doctor - patient question - answer texts, and build an intuitive and easy-to-understand knowledge network based on these entities and their associations.

This knowledge graph management system not only provides users with a new intelligent consultation platform, so that doctors can quickly grasp the patient's condition, accurately judge the relationship between symptoms and diseases, so as to give more personalized diagnosis and treatment recommendations, but also greatly improve the patient's medical experience. Through graphical display, patients can easily obtain medical information and treatment recommendations that are highly relevant to their own conditions, making the transmission of medical knowledge more intuitive and vivid. The application of the system also significantly improves the efficiency and accuracy of intelligent consultation, which can provide doctors with instant information reference, reduce the risk of misdiagnosis and missed diagnosis, and also allow patients to get more accurate responses in a shorter time. In short, the application of this knowledge graph management system, which combines two-way LSTM-CRF and natural language processing technology, in intelligent consultation not only optimizes the processing process of medical information, but also brings more convenient and accurate medical experience to both doctors and patients.

## 6. Conclusion and Prospect

In the growing convergence of modern healthcare and internet technology, we recognized the significant research value of doctor-patient communications on online medical platforms. We selected breast cancer-related doctor-patient Q&As from an online medical community as our

research material. Using an intelligent algorithm combining bidirectional LSTM-CRF with natural language processing, we conducted in-depth analysis of the medical information in these exchanges. We successfully identified and extracted medical entities from the text—including disease names, symptom descriptions, and treatment methods—with notable accuracy.

Building on this foundation, we clarified the relationships among these entities and visually presented a detailed medical knowledge system in graphical form. This innovative approach demonstrates the promising potential of combining online medical platforms with intelligent technology and expands the application scope of medical knowledge graphs. We discovered that this technology extends beyond traditional electronic medical records or medical dictionaries to the online communities and forums that people commonly use today. More significantly, this technology offers a new approach to intelligent consultation, allowing quick matching and provision of professional disease information and solutions based on patient descriptions. Through graphical display, the entire consultation process becomes more intuitive and understandable, substantially improving the patient experience.

Our research has several areas for improvement. Currently, we've focused solely on breast cancer, but future work could expand to include more diseases, creating a more comprehensive knowledge system. The classification of medical entities could be more detailed—for example, differentiating drugs into traditional Chinese medicine and Western medicine categories would increase the knowledge system's precision. We could also attempt to combine the entity identification and relationship extraction steps to reduce errors and information redundancy, further improving the technology's efficiency and accuracy. Overall, this study reveals the significant potential of medical entity recognition based on bidirectional LSTM-CRF and natural language processing for intelligent consultation systems, while also indicating directions for future research and optimization.

## References


*[1] Chen, Junyu. "Research on Intelligent Data Mining Technology Based on Geographic Information System." Journal of Computer Science and Artificial Intelligence 2.2 (2025): 12-16.*

*[2] Xu, Yue. "Research on Maiustream Web Database Development Techncl ogy." Journal of Computer Science and Artificial Intelligence 2.2 (2025): 29-32.*

*[3] Shi, Chongwei. "Ovarian Hereditary Diseases: Progress in Prevention and Treatment and Research on Prenatal Diagnosis." Scientific Journal of Technology 7.2 (2025): 125-131.*

*[4] Gu, Yiting. "Practical Approaches to Develo**High-performance Web Applications Based on React." Frontiers in Science and Engineering 5.2 (2025): 99-105.*

*[5] Shi, Chongwei. "Research on Gene Identification Algorithms Based on Signal Processing Techniques." 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2024.*

*[6] Zhao, Fengyi. "Risk Assessment Model and Empirical Study of in Vitro Diagnostic Reagent Project Based on Analytic Hierarchy Process." International Journal of New Developments in Engineering and Society 8.5 (2024), 76-82.*

*[7] Yang J. Research on the Strategy of MedKGGPT Model in Improving the Interpretability and Security of Large Language Models in the Medical Field[J]. Academic Journal of Medicine & Health Sciences, 5(9): 40-45.*

*[8] Yang J. Research on the Application of Medical Text Matching Technology Combined with Twin Network and Knowledge Distillation in Online Consultation[J].*

*[9] Cao, Y., Cao, P., Chen, H., Kochendorfer, K. M., Trotter, A. B., Galanter, W. L., ... & Iyer, R. K. (2022). Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-*

*based model. In Multimodal AI in healthcare: A paradigm shift in health intelligence (pp. 245-256). Cham: Springer International Publishing.*

*[10] Chen, H., Wang, Z., & Han, A. (2024).* Guiding Ultrasound Breast Tumor *Classification with Human-Specified Regions of Interest: A Differentiable Class Activation Map Approach. In 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS) (pp. 1-4). IEEE.*

*[11] Yang, Jinzhu "Integrated Application of LLM Model and Knowledge Graph in Medical Text Mining and Knowledge Extraction."Social Medicine and Health Management (2024), 5(2): 56-62*

*[12] Varatharajah, Y., Chen, H., Trotter, A., & Iyer, R. K. (2020). A Dynamic Human-in-the-loop Recommender System for Evidence-based Clinical Staging of COVID-19. In HealthRecSys@ RecSys (pp. 21-22).*

*[13] Varatharajah, Y., Chen, H., Trotter, A., & Iyer, R. K. (2020). A Dynamic Human-in-the-loop Recommender System for Evidence-based Clinical Staging of COVID-19. In HealthRecSys@ RecSys (pp. 21-22).*

*[14] Zhao, Fengyi "Development Design and Signal Processing Algorithm Optimization of Traditional Chinese Medicine Pulse Acquisition System Based on CP301 Sensor." Advances in Computer, Signals and Systems (2024), 8(6): 106-111*

*[15] Wang, Yuxin "Research on Intelligent Macro Image Recognition Algorithm of Oil Pipe Failure Based on Deep Learning." Journal of Image Processing Theory and Applications (2025), 8(1): 1-7*

*[16] Guo H , Yan J .The Study of Named Entity Identification in Chinese Electronic Medical Records Based on Multi-tasking[C]//International Conference on Knowledge Science, Engineering and Management.Springer, Singapore, 2024.DOI:10.1007/978-981-97-5501-1_22.*

*[17] Liu Z , Huang Y , Yu X ,et al.DeID-GPT: Zero-shot Medical Text De-Identification by GPT-4[J]. 2023.*

*[18] Heath M , Torpy D J , Rushworth R L .An analysis of the utilisation of medical identification jewellery among children and young adults with type 1 diabetes mellitus in Australia[J].Endocrine (1355008X), 2023, 79(1).DOI:10.1007/s12020-022-03224-3.*

*[19] Shoshan Y , Ratner V .MEDICAL OBJECT DETECTION AND IDENTIFICATION VIA MACHINE LEARNING:US202016932115[P].US2022020143A1[2025-04-01].*

*[20] Zhang, Jinshuo "Research on Real Time Condition Monitoring and Fault Warning System for Construction Machinery under Multi Source Heterogeneous Data Fusion." Journal of Engineering Mechanics and Machinery (2024), 9(2): 139-144*