

Research Article

Medical Knowledge Graph in Chinese Using Deep Semantic Mobile Computation Based on IoT and WoT

Wanheng Liu,¹ Ling Yin²,³ Cong Wang,¹ Fulin Liu,³ and Zhiyu Ni³

¹Beijing University of Posts and Telecommunications, China

²National Population Health Data Center, China

³Affiliated Hospital of Hebei University, China

Correspondence should be addressed to Ling Yin; lwahlcf@163.com

Received 4 February 2021; Accepted 13 May 2021; Published 28 May 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Wanheng Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, a novel medical knowledge graph in Chinese approach applied in smart healthcare based on IoT and WoT is presented, using deep neural networks combined with self-attention to generate medical knowledge graph to make it more convenient for performing disease diagnosis and providing treatment advisement. Although great success has been made in the medical knowledge graph in recent studies, the issue of comprehensive medical knowledge graph in Chinese appropriate for telemedicine or mobile devices have been ignored. In our study, it is a working theory which is based on semantic mobile computing and deep learning. When several experiments have been carried out, it is demonstrated that it has better performance in generating various types of medical knowledge graph in Chinese, which is similar to that of the state-of-the-art. Also, it works well in the accuracy and comprehensive, which is much higher and highly consisted with the predictions of the theoretical model. It proves to be inspiring and encouraging that our work involving studies of medical knowledge graph in Chinese, which can stimulate the smart healthcare development.

1. Introduction

In recent years, healthcare was attached much importance in people's daily life, and semantic mobile computing applied in healthcare has become a hot topic and received much attention from researchers and academia from all over the world recently. To make it more convenient and easy for both doctors and patients, a comprehensive medical knowledge graph in Chinese using the deep learning method on a mobile device based on IoT and WoT is of great significance. Also, it can serve for smart healthcare which can be applied in remote consultation platform and mobile devices in hospital to implement disease classification and treatment recommendations. It is more time-saving, economic, and efficient.

It is well known that technological revolutions have hit the industrial world and semantic mobile computing in Internet of Things (IoT) and Web of Things (WoT) has emerged from intuitive, which can make it more convenient for people to get the information more faster, accurately, and intelligently from mass amount of Internet information. The

knowledge graph was firstly proposed by Google in 2012 [1], which is the knowledge base that Google utilized to enhance its search engine. There are mainly three eras in the knowledge graph, including the ontology era, semantic web, and knowledge graph era. In previous studies, knowledge graphs in the ontology era were constructed manually by experts in relevant fields, including Cyc [2] in 1984, WordNet [3] proposed by Miller G A. in 1985, and Chinese dictionary HowNet. When the World Wide Web was set up in 1989, it resolves the previous restriction that the query tool could only look up information step by step in a specific path. To address the limitations in application, the Semantic Web [4] era occurred when Linked Data [5] was presented by Berners-Lee T. in 2006, and DBpedia [6, 7] was the representative one. Then in 2001, Wikipedia was launched to provide a free encyclopedia for people, for example, DBpedia and Yago [8]. Then, Freebase [9] was built in 2007, which can be manually and actively imported. However, to conduct effective search and analysis of knowledge, the knowledge graph era emerged. The typical examples are Knowledge

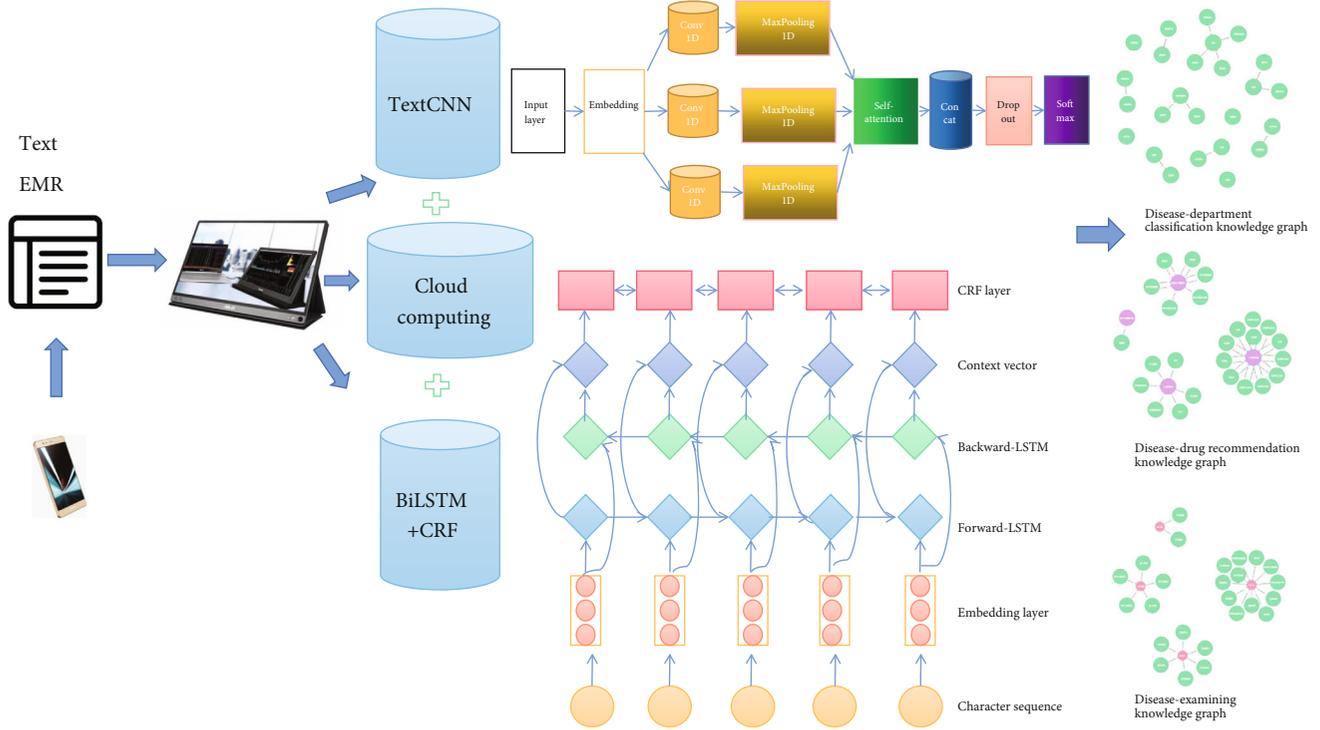


FIGURE 1: The architecture of our medical knowledge graph in the Chinese system.

Vault [10], Google Knowledge graph, Facebook graph search, and Baidu Knowledge.

In recent years, as the maturity of technology and more emphasis on healthcare, it has become more and more popular to apply knowledge graph in the medical field and has attracted much attention from researchers in computer and medical to combine these two fields [11–26]. Rotmensch et al. [11] in 2017 proposed learning a health knowledge graph from electronic medical records by using probabilistic. Longxiang et al. [12] presented a framework that can retrieve knowledge in a knowledge graph automatically in 2017. Wen et al. [14] proposed an attention-aware path-based relation extraction for medical knowledge graph. Sun et al. [21] in 2018 incorporate description embeddings into medical knowledge graphs representation learning. Kai et al. [22] presented a novel semantic similarity measure in biomedical knowledge graphs in 2018. Wang et al. [23] based on knowledge graph design and implement personal health record systems using semantic integration methods in 2018. Wang et al. [25] made use of a knowledge graph to process medical archives in 2019. In addition, Song et al. [13] put forward the Chinese pediatric medical knowledge graph in 2019. Ying et al. [15] proposed a system for drug-drug similarity measure based on a knowledge graph in 2019. Xiao et al. [26] in 2019 based on medical knowledge graph to enhance the fraud, waste, and abuse detection on claim data using the deep learning-based method to extract the entities and relationships from the knowledge sources. Li et al. [27] used an imperfect knowledge graph to classify rare diseases in 2019. Chen et al. [16] in 2020 proposed robustly extracting medical knowledge from EHRs, in which nonlinear functions are adopted in building the causal graph to better understand

existing model assumptions. Zheng et al. [17] put forward a dedicated knowledge graph benchmark for biomedical data mining in 2020. Also, medical knowledge graph can be applied in specific departments, for example, Xie et al. [24] in 2018 presented a data-driven traditional Chinese medicine knowledge discovery method based on a knowledge graph. Li et al. [18] came up constructing a knowledge graph of knee osteoarthritis in Chinese using an automatic approach in 2020. Chai et al. [19] in 2020 proposed the diagnosis method of thyroid disease combining knowledge graph and deep learning. Alawad et al. [20] utilize medical knowledge graph combined with a deep learning method for cancer phenotyping in 2021. Great progress has been made in the knowledge graph applied in the medical field above; however, it still suffers from problems as follows: (1) owing to data lacking and inadequate, it is hard for the model of knowledge graph to achieve the best performance. (2) On account of limitations of the medical field, the tools for building knowledge graph in the healthcare industry are immature and domain knowledge to be integrated in the same medical system has some limitations. (3) There are still some difficulties in updating medical knowledge. (4) Due to the most medical knowledge graph focusing on a single disease, various disease classifications and treatment recommendations have been overlooked. (5) Because of complex medical knowledge, it is hard to find the balance in the process of the knowledge graph structure.

To overcome the limitations above, a novel Chinese healthcare knowledge graph using deep learning method which can be utilized in the mobile device based on IoT and WoT is proposed in this paper. In order to address the lack of data, 600 thousand pieces of data were processed in our work to support the training of the model for good

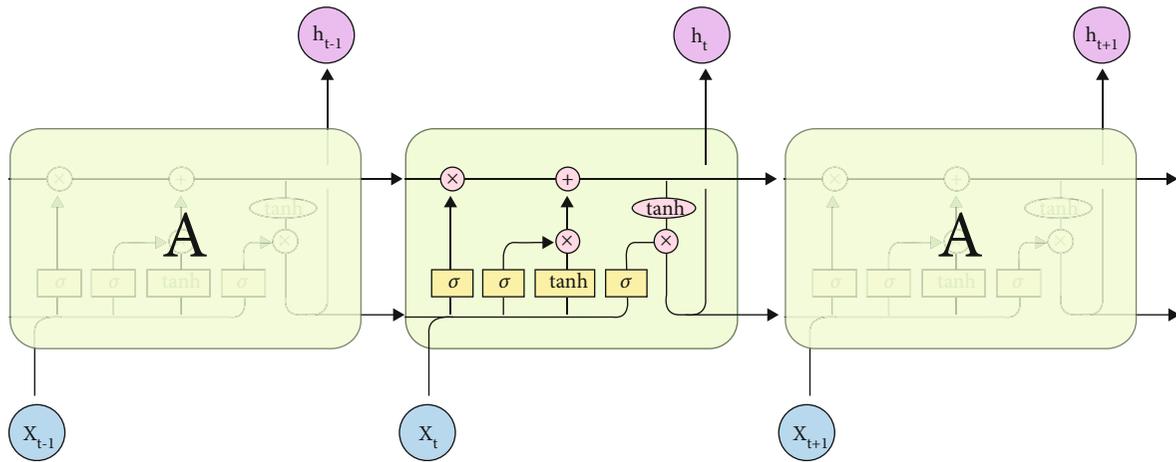


FIGURE 2: The structure of LSTM.

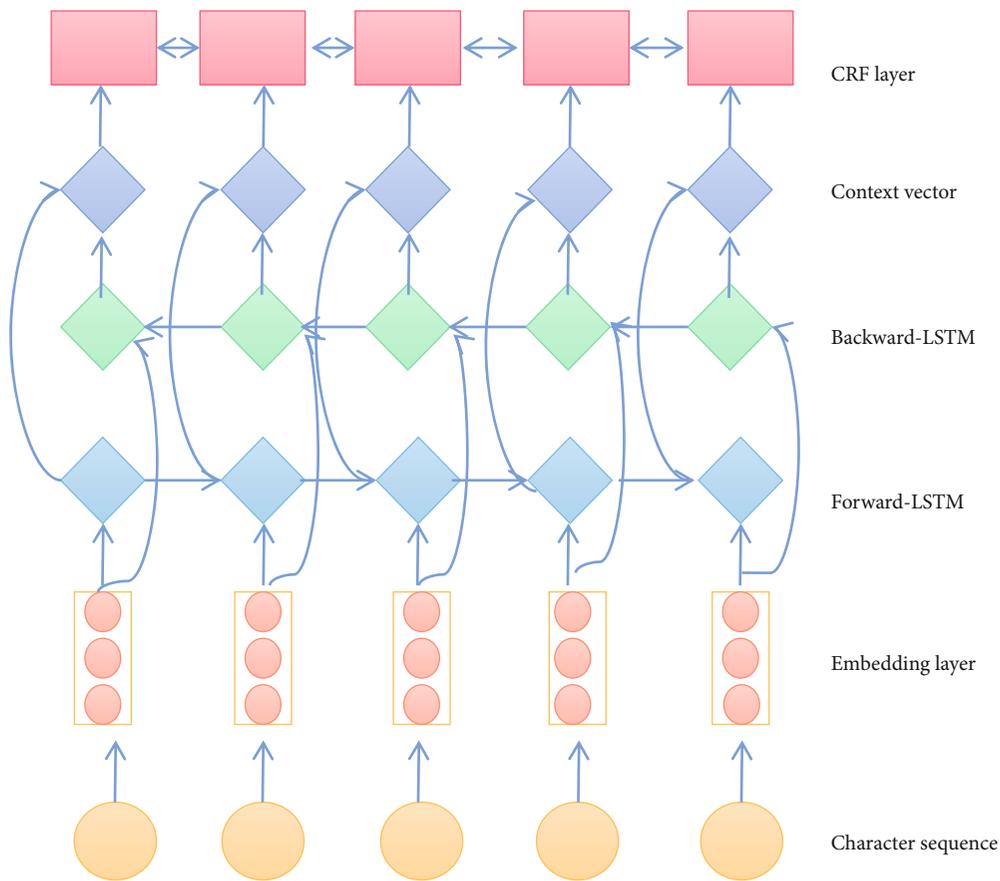


FIGURE 3: The architecture of BiLSTM+CRF.

performance achieved. Furthermore, the deep learning method instead of the traditional method is adopted in our study to realize an intelligent medical knowledge graph. And the self-attention is used in our deep learning network to reduce the amount of computation cost. Besides, cloud computing is also utilized to make it in practice which is convenient for both doctors and patients through a mobile device based on IoT and WoT. New data would be computed and stored on the cloud, in which text and image

computing, processing, and storage capabilities are involved. It counts that not only can the diseases be classified but also treatment recommendations can be made in the intelligent knowledge graph we presented. What is more, our medical knowledge system can be applied in practice in remote consultation platform served for both patients and doctors. And the subject is concerned chiefly with the study of a knowledge graph. The contributions in our study are the following:

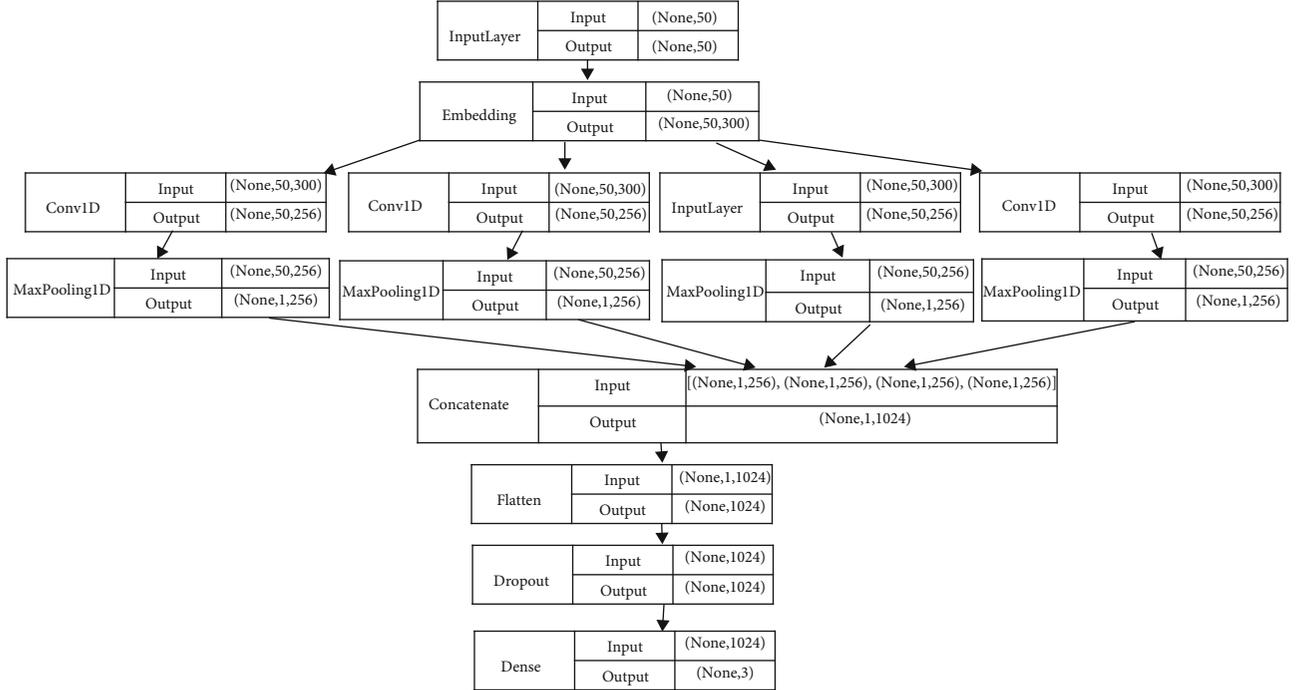


FIGURE 4: The architecture of Text-CNN.

- (i) A novel and more comprehensive medical knowledge graph in Chinese based on deep learning method is proposed, which can be served for doctors and patients
- (ii) The self-attention is utilized instead of the traditional attention mechanism in our deep learning network, which is more simple and the computational cost is relatively low
- (iii) A new dataset is labelled and processed to train and obtain a well model
- (iv) To make it can be in practice based on IoT and WoT, the cloud computing is combined to support our study

In this paper, it is divided into 4 sections as follows: in Section 2, it introduces the materials and methods in which the system and network architecture contained. And in Section 3, we show the results and discussion of our work in this part. Finally, it is concluded in this paper in Section 4.

2. Materials and Methods

The architecture of the medical knowledge graph in Chinese based on deep neural networks applied in telemedicine as can be seen in Figure 1. The cloud computing is adopted to process and store the data, which makes it in practice based on IoT and WoT when using a mobile device. When new data is acquired, it can be computed by the cloud, and then, new data would be stored on the cloud to improve the medical knowledge graph, which is more efficient and convenient. Firstly, text information which can be recognized and

TABLE 1: The structure of TextCNN.

Embedding	Conv(3)	Conv(4)	Conv(5)
Bn		Bn	Bn
Relu		Relu	Relu
Conv(3)		Conv(4)	Conv(5)
Bn		Bn	Bn
Relu		Relu	Relu
Max-pool		Max-pool	Max-pool
Concat			
Dropout			
Softmax			

EMR on the mobile device of the patients is input to the mobile device of the doctor. When the EMR and text information can be recognized is obtained, it would be transmitted to our model to compute, which BiLSTM, CRF, TextCNN, and self-attention contained to accomplish the feature representation and sequence labeling. What is more, our model is based on 600 thousand of training data. Therefore, it has better performance in accuracy and comprehensive of our medical knowledge graph generated. It can not only perform department classification according to the disease and also can recommend food and drug for the patients. In addition, it can advise the patients to examine the major project of the disease to reduce the time and economy consuming.

2.1. Network Architecture. The neural network we adopted in our study is BiLSTM (Bidirectional Long Short-Term

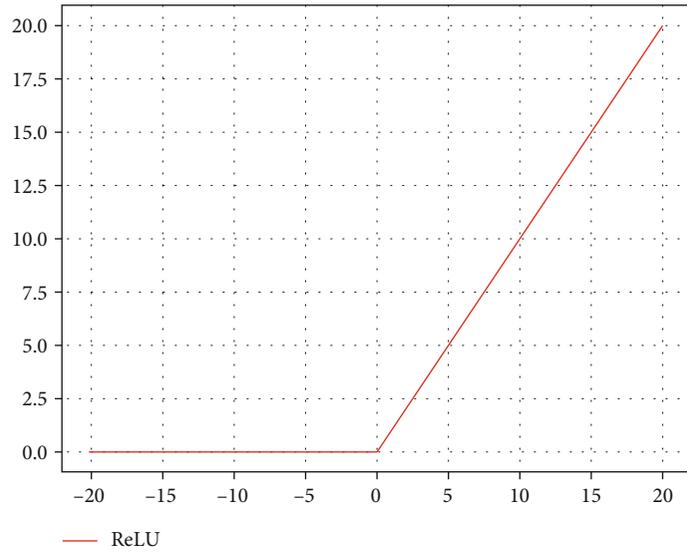


FIGURE 5: Representation of Relu Activation Function.

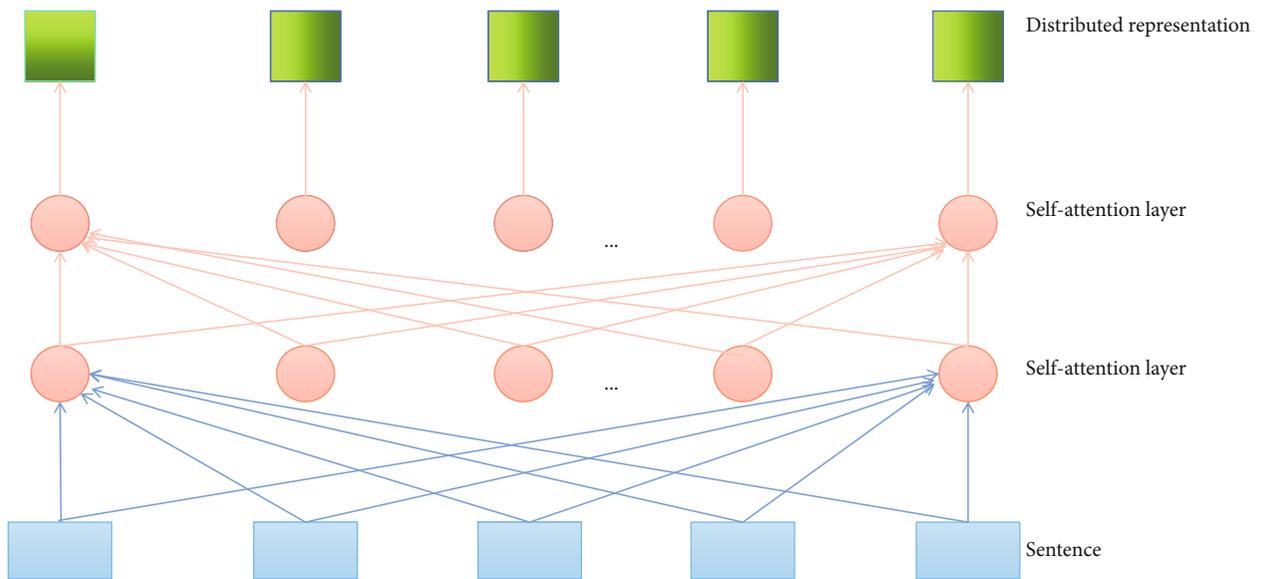


FIGURE 6: Self-attention.

Memory), which consisted of the forward LSTM and the backward LSTM, and it is often used to model contextual information in natural language processing tasks. It is a fact that the LSTM is a special type of RNNs. And RNN is a special neural network structure made up of input layer, hidden layer, and output layer three layers, which is proposed according to the view that human cognition is based on past experience and memory. Not only the input from the previous moment is taken into account but also the memory function of the network for the previous content is also added in RNN. It works well for time series problems; however, it can result in a small memory value when costing too much time. Therefore, the LSTM is proposed to address the problem of gradient vanishing and gradient explosion in long sequence training, which is the extension of the traditional

model and has a better performance in a long sequence. As shown in Figure 2, it can be seen that in LSTM proposed by Sepp Hochreiter in 1997 [28], which is not a single layer of neural networks, but there are four layers and a lot of gates added that interact in a specific way. In LSTM, the first step is to decide what information should be discarded from the cellular state, which was made by the sigmoid in the forget-gate layer. Then, to decide what new information ought to be stored in the cell state, where the sigmoid layer in the input gate determines which values to be updated and the tanh layer creates a new candidate value vector to be added to the state. And a status will be updated when the above steps are combined, and the old cell state will be updated to a new cell state. Finally, it needs to decide what to output, in which the sigmoid layer determines to output which parts

of the cell state. In conclusion, under the control of the gate, the state vector of an LSTM unit and the output can be obtained. Based on LSTM, in which the state vector of each cell is input to the cell at the next moment and the process is unidirectional, BiLSTM can be considered as two layers of LSTM. It is found that the state inputs of the two layers are in opposite directions, which are forward and backward propagation, respectively. In BiLSTM, it has better performance and better capture bidirectional semantic dependencies which can solve the problem of unable to encode information from back to front in LSTM.

2.2. BiLSTM+CRF. CRF (Conditional Random Field) [29] is often utilized to achieve better performance in sequence labeling tasks, which the whole sequence can be taken into account. In BiLSTM, due to no conditional constraint of label transitions, a completely wrong sequence of annotations may be output. Therefore, in our architecture, BiLSTM was combined with CRF is adopted in our architecture. Due to the independent of each other in the output of the softmax layer of the LSTM, it cannot affect each other in output when the contextual information learned by BiLSTM. Nevertheless, the label output of the maximum probability value can be picked at each step. And the transition feature function is contained in CRF which the context of the label can be learned. Then CRF is used to do the output layer of BiLSTM when the order between output labels considered. It can regard the output of BiLSTM as the probability of each tag's value, and the function of CRF as the results adjusted according to the relationship between labels. And the architecture of BiLSTM combined with CRF is shown in Figure 3, which strengthens the relevance of information between texts and the optimal sequence result can be obtained through the probability of the feature to the tag.

2.3. TextCNN. The TextCNN is added in our model to extract local features of the current word. On account of the capacity of the model, some important information may discard when long sentences are learned by BiLSTM. Also, for the word segmentation task, the label of the current word is basically only associated with the previous several words. As shown in Figure 4, when a sample of some sentences input and needs to be classified, the convolution operation will be conducted by CNN. Each word in the sentence is composed of n -dimensional word sentence, so the size of the input matrix is $m \times n$, where m is the sentence length. It is counted that for text data, the filter moves down no longer swipes horizontally. Then the max-pooling operation of each vector would be carried out and the pooling value be stitched. Finally, the feature representation of the sentence is obtained and the sentence vector is classified by the classifier.

There are four parts in Text-CNN, embedding layer, convolution layer, pooling layer, and fully connected layer. In the embedding layer, the words can be mapped into corresponding vectors. And the main purpose of this layer is to encode the natural language input into distributed representation. In the convolution layer, the one-dimensional convolution operation is performed for the word mapped features. And it is mainly used to extract different N -gram features through

TABLE 2: The comparison of F1 scores (state-of-the-art) of different models for NER.

System	Accuracy
CRF	90.9
Conv-CRF	89.59
BiLSTM-CRF	90.1
BiLSTM-CNNs	90.77
Attention-BiLSTM-CRF	90.84
Self-attention-BiLSTM-CRF+TextCNN	90.93

TABLE 3: The comparison of accuracy(state-of-the art) in different models for our dataset.

Model	Accuracy
BiLSTM-CRF	84.97
Attention-BiLSTM-CRF	85.1
Self-attention-BiLSTM-CRF+TextCNN	85.5

convolution. The input statement or text can be converted to a two-dimensional matrix when passed through the embedding layer. It is noted that in TextCNN, various types of kernels should be utilized at the same time, and each size of kernels is also multiple. Then, in the pooling layer, the pooling operation is conducted on the convolutioned results, which aims at extracting the features with the highest activation degree to N -gram features extracted by convolution. The maximum value is taken when a number of one-dimensional vectors have been convolutioned; then, they are pieced together as output in the max-pooling layer. Finally, in a fully connected layer, softmax is added to output the probabilities for each category.

It is clear that the basic structure of the TextCNN is shown in Table 1. The batch normalization is added in the structure. When the natural language is normalized, it can reduce randomness and bring it closer to predefined standards. What is more, it can contribute to reducing the amount of different information that would be processed and increase efficiency. In other words, if batch normalization is performed, the text distribution is more closer to normal distribution.

2.4. Loss Function. The sequence tagging at the sentence level can be executed in the CRF layer. And the parameter in the CRF layer can be treated as a matrix A , which is $(k+2) \times (k+2)$. As shown in Formula (1), it means that for the whole sequence x , the score of the whole sequence annotation is equal to the sum of the scores of all positions, which is divided into two parts. The first part in the formula can be affected by the π , which is the output of BiLSTM. And the second part was determined by the CRF transfer matrix A , where $A_{y_{i-1} y_i}$ and y_i represent the transfer score from the y_{i-1} tag to the y_i tag. In conclusion, front-and-back tag dependency constraints were taken into account in CRF, and the mark state transition probability was used as the score.

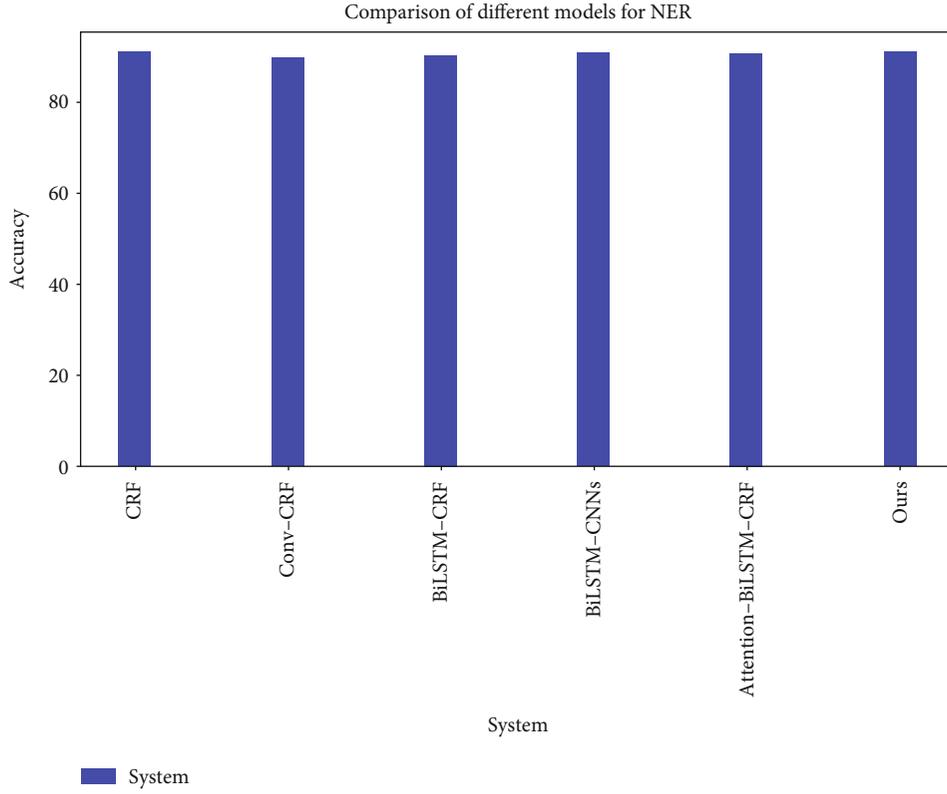


FIGURE 7: The accuracy in different models for NER.

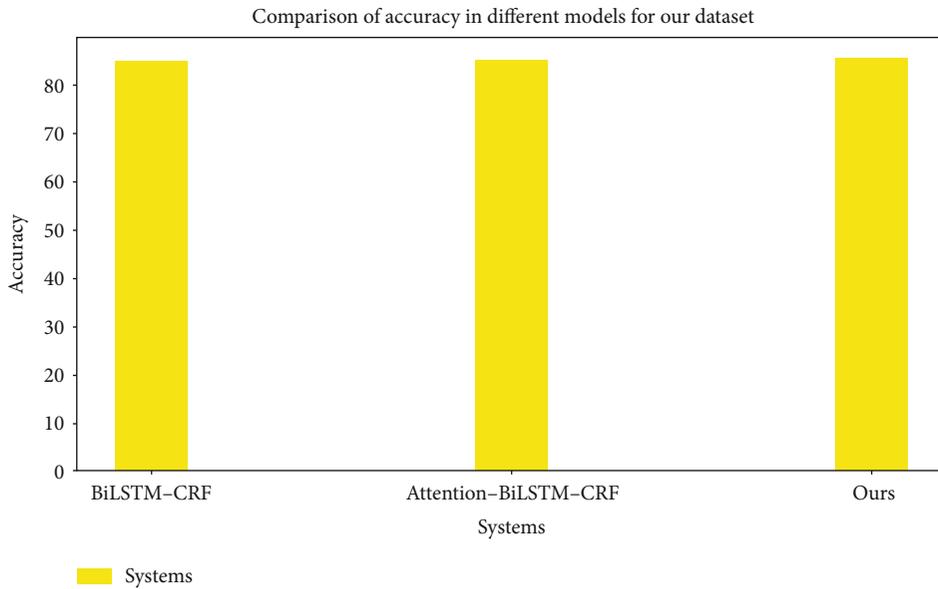


FIGURE 8: The accuracy in different models for our dataset.

$$\text{score}(x, y) = \sum_{i=1}^n P_i, y_i + \sum_{i=1}^{n+1} A y_i - 1, y_i. \quad (1)$$

$$P(y | x) = \frac{\exp(\text{score}(x, y))}{\sum_{y'} \exp(\text{score}(x, y'))}. \quad (2)$$

Then, the normalized probability can be acquired by the softmax, which can be presented as in (Formula) (2).

When model training, the maximizing logarithm likelihood function would be utilized. It can be defined in Formula

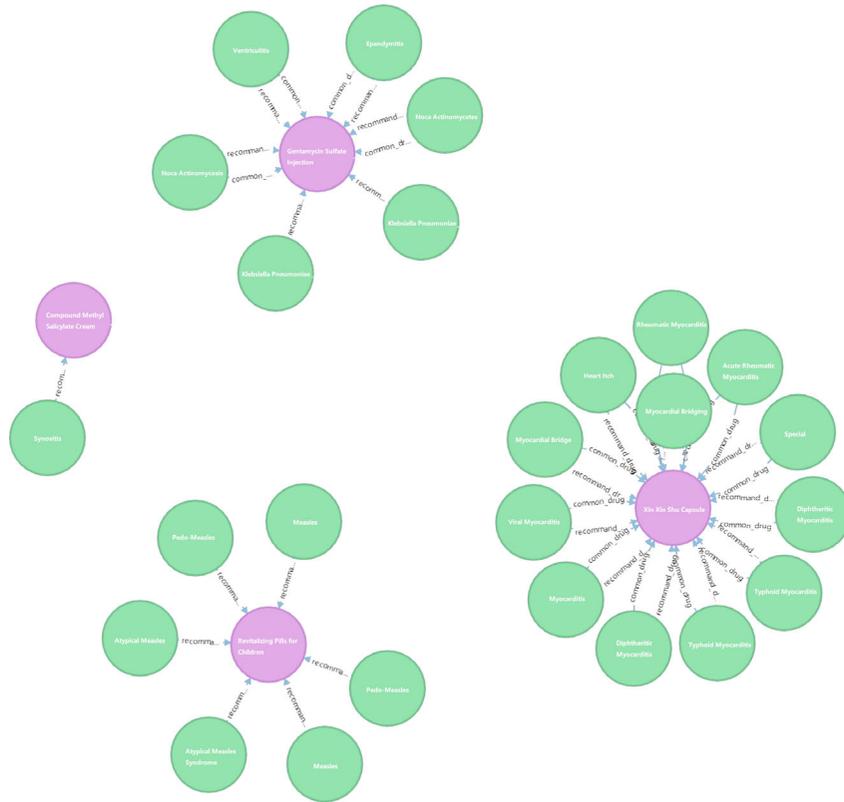


FIGURE 11: Disease-food not recommended.

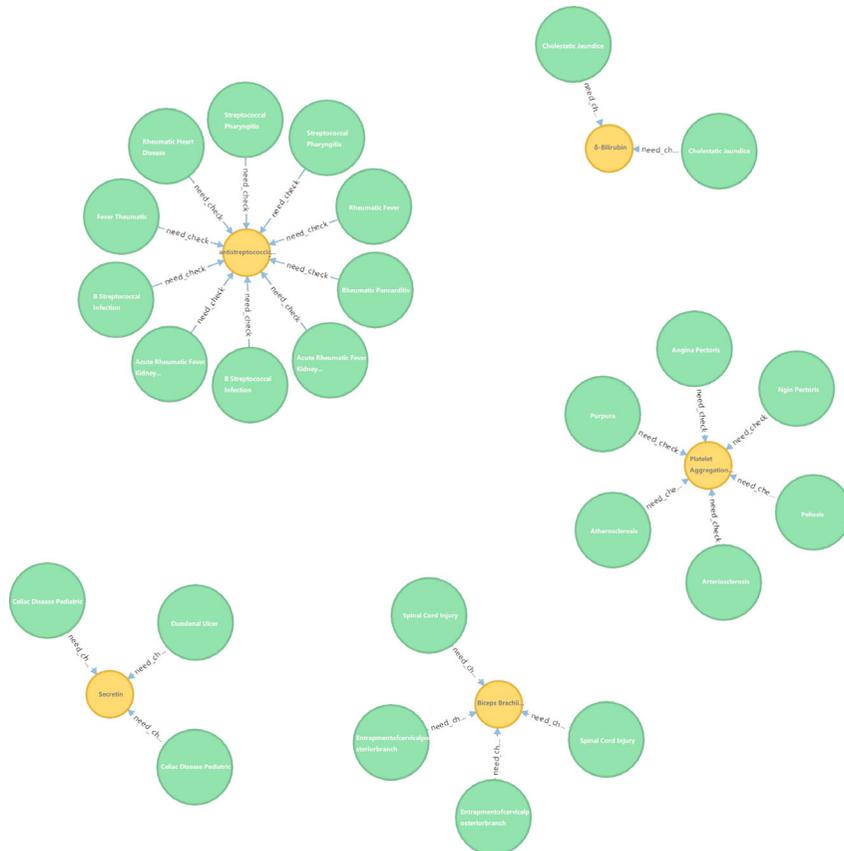


FIGURE 12: Disease drug recommendation knowledge graph.

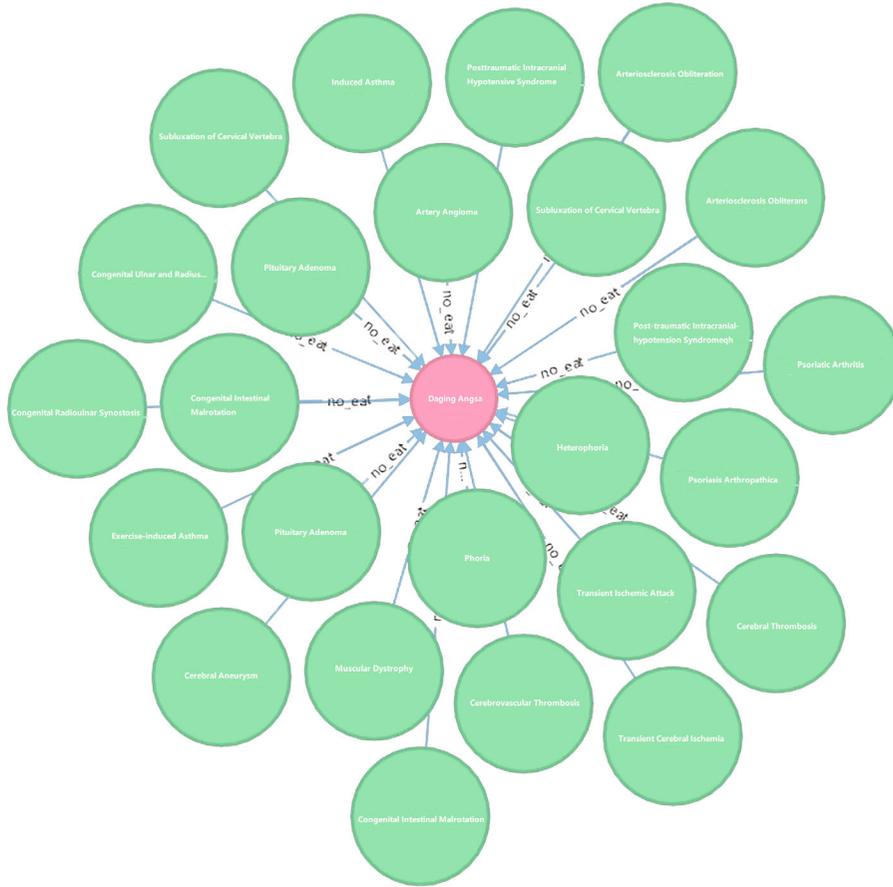


FIGURE 13: Drug-producer.

2.7. *Cloud Computing.* To make our model in practice, cloud computing is also adopted in our system. Cloud computing is an Internet-based computing in which the shared hardware and software resources and information are made available to computers and other devices on demand. When it comes to our system, to make it more efficient, the amount of data that patients and doctors are working with is not stored locally, but in the medical data center. When new data is acquired, it can be computed by the cloud, and then, new data would be stored on the cloud, in which text and image computing, processing, and storage capabilities are involved. What is more, the cloud computing is low-price and more convenient. Due to its automated centralized management, it can free engineers in the hospital from the increasing cost of data center management and enjoy the advantages of low cost. In addition, the terminal equipment requirements are very low in cloud computing, so it is no need to purchase expensive terminal equipment. Furthermore, it makes it easy to share medical text data and applications between different devices when cloud computing is applied. Therefore, it can help doctors with remote diagnosis and remote treatment.

3. Results and Discussion

Our experimental hardware device is based on Windows 10 operating system and intel i7 CPU processor, and the train-

ing and testing process is based on python, tensorflow, and neo4j. In Table 2, it shows the state-of-the-art F1 scores of different models for NER. And it can be seen that our system has a better performance than common systems for NER. When testing on our dataset, our model still has higher accuracy and performs better in Table 3. And the standard evaluation method we used is based on the macromean F1 value. Then, the bar chart shows in Figure 7 and in Figure 8 to make it more intuitive and clear. Meanwhile, it is shown the results of the medical knowledge graph in Chinese obtained from the experiments of our model. Firstly, it demonstrated that the disease department classification can be formed a knowledge graph in Figure 9, which the disease can be recommended to the correct and corresponding disease department. Secondly, the recommendation food of the disease can also be generated knowledge graph in Figure 10, which the reasonable diet can be advised according to the corresponding disease while the unsuitable food is not advised to the disease which is shown in Figure 11. Also, according to the diseases, the drug recommendation knowledge graph is produced as shown in Figure 12, in which the reasonable and effective medicine can be suggested to the corresponding diseases for the patients. And the drug producer is shown in Figure 13, which can be seen in the medicine factory. What is more, our model can suggest the patients to check the main factor of the disease in

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the study of this work and publication of this paper.

Acknowledgments

Our work is supported by the Key R&D Program of Hebei Province, People's Livelihood Science and Technology Project, Project No.: 20277716D, Project name: Study on the pathological characteristics and pathogenic mechanism of Covid-19.

References

- [1] A. Singhal, "Introducing the Knowledgegraph: Things, Not Strings," *Official google blog*, vol. 5, p. 16, 2012.
- [2] D. B. Lenat, M. Prakash, and M. Shepherd, "CYC: using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks," *AI Magazine*, vol. 6, no. 4, p. 65, 1985.
- [3] G. A. Miller, "Wordnet: Dictionary Browser' in Information in Data," *Proceedings of the First Conference of the UW Centre for the New Oxford Dictionary*, University of Waterloo, Waterloo, Canada, 1985.
- [4] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 28–37, 2001.
- [5] T. Berners-Lee, "Linked data-design issues," 2006, <http://www.w3.org/DesignIssues/LinkedData.html>.
- [6] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: a nucleus for a web of open data," *The semantic web*, pp. 722–735, 2007.
- [7] R. R. Swick, *Resource Description Framework (RDF) Model and Syntax Specification W3C Recommendation*, W3c Recommendation World Wide Web Consortium, 1999.
- [8] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge," in *Proceedings of the 16th international conference on World Wide Web*, pp. 697–706, Banff, Alberta, Canada, 2007.
- [9] K. Bollacker, R. Cook, and P. Tufts, "Freebase: a shared database of structured general human knowledge," *AAAI*, vol. 7, pp. 1962–1963, 2007.
- [10] X. Dong, E. Gabrielovich, G. Heitz et al., "Knowledge vault: a web-scale approach to probabilistic knowledge fusion," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 601–610, New York, USA, 2014.
- [11] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, "Learning a health knowledge graph from electronic medical records," *Scientific Reports*, vol. 7, no. 1, p. 5994, 2017.
- [12] L. Shi, S. Li, X. Yang, J. Qi, G. Pan, and B. Zhou, "Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services," *BioMed Research International*, vol. 2017, 12 pages, 2017.
- [13] Y. Song, L. Cai, K. Zhang, H. Zan, T. Liu, and X. Ren, *Construction of Chinese Pediatric Medical Knowledge Graph [C]//Joint International Semantic Technology Conference*, Springer, Singapore, 2019.
- [14] D. Wen, Y. Liu, K. Yuan, S. Si, and Y. Shen, "Attention-aware path-based relation extraction for medical knowledge graph," in *International Conference on Smart Computing and Communication*, pp. 321–331, Springer, Cham, 2017.
- [15] Y. Shen, K. Yuan, J. Dai, B. Tang, M. Yang, and K. Lei, "KGDDS: a system for drug-drug similarity measure in therapeutic substitution based on knowledge graph curation," *Journal of Medical Systems*, vol. 43, no. 4, pp. 1–9, 2019.
- [16] I. Y. Chen, M. Agrawal, S. Horng, and D. Sontag, *Robustly Extracting Medical Knowledge from EHRs: A Case Study of Learning a Health Knowledge Graph*, Pacific Symposium on Biocomputing, 2020.
- [17] S. Zheng, J. Rao, Y. Song et al., "PharmKG: a dedicated knowledge graph benchmark for biomedical data mining," *Briefings in Bioinformatics*, 2020.
- [18] X. Li, H. Liu, X. Zhao, G. Zhang, and C. Xing, "Automatic approach for constructing a knowledge graph of knee osteoarthritis in Chinese," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1–8, 2020.
- [19] X. Chai, "Diagnosis method of thyroid disease combining knowledge graph and deep learning," *IEEE Access*, vol. 8, pp. 149787–149795, 2020.
- [20] M. Alawad, S. Gao, M. C. Shekar et al., *Integration of Domain Knowledge Using Medical Knowledge Graph Deep Learning for Cancer Phenotyping*, 2021, arXiv:2101.01337.
- [21] X. Sun, Y. Man, Y. Zhao, J. He, and N. Liu, "Incorporating description embeddings into medical knowledge graphs representation learning," in *International Conference on Human Centered Computing*, pp. 188–194, Springer, Cham, 2018.
- [22] K. Lei, K. Yuan, Q. Zhang, and Y. Shen, "MedSim: a novel semantic similarity measure in bio-medical knowledge graphs," in *International Conference on Knowledge Science, Engineering and Management*, Springer, Cham, 2018.
- [23] H. Wang, X. Miao, and P. Yang, *Design and Implementation of Personal Health Record Systems Based on Knowledge Graph [C]//2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, IEEE (Institute of Electrical and Electronics Engineers), 2018.
- [24] Y. Xie, C. Yan, and D. Zhang, *Personalized Diagnostic Modal Discovery of Traditional Chinese Medicine Knowledge Graph [C]// 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, IEEE, 2018.
- [25] X. Wang, R. Wang, Z. Bao, J. Liang, and W. Lu, "Effective medical archives processing using knowledge graphs," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1141–1144, 2019.
- [26] J. Xiao, H. Sun, W. Zhu et al., "Building a medical knowledge graph to enhance the fraud, waste and abuse detection on claim data (preprint)," *JMIR Medical Informatics*, vol. 8, no. 7, 2019.
- [27] X. Li, Y. Wang, D. Wang, W. Yuan, D. Peng, and Q. Mei, "Improving rare disease classification using imperfect knowledge graph," *BMC Medical Informatics and Decision Making*, vol. 19, no. 5, pp. 1–10, 2019.
- [28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [29] J. Lafferty, A. McCallum, and F. Pereira, *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, Proceedings of ICML, 2001.