Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 2489754, 10 pages
https://doi.org/10.1155/2021/2489754

WILEY | Hindawi

Research Article

Research on Named Entity Recognition of Electronic Medical Records Based on RoBERTa and Radical-Level Feature

Yue Wu, Jie Huang, Caie Xu, Huilin Zheng, Lei Zhang, and Jian Wan

School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Zhejiang Hangzhou 310023, China

Correspondence should be addressed to Jian Wan; wanjian@zust.edu.cn

Received 11 May 2021; Revised 6 June 2021; Accepted 16 June 2021; Published 28 June 2021

Academic Editor: Honghao Gao

Clinical named entity recognition (CNER) identifies entities from unstructured medical records and classifies them into predefined categories. It is of great significance for follow-up clinical studies. Most of the existing CNER methods fail to give enough thought to Chinese radical-level characteristics and the specialty of the Chinese field. This paper proposes the Ra-RC model, which combines radical features and a deep learning structure to fix this problem. A bidirectional encoder representation of transformer (RoBERTa) is utilized to learn medical features thoroughly. Simultaneously, we use the bidirectional long short-term memory (BiLSTM) network to extract radical-level information to capture the internal relevance of characteristics and stitch the eigenvectors generated by RoBERTa. In addition, the relationship between labels is considered to obtain the optimal tag sequence by applying conditional random field (CRF). The experimental results demonstrate that the proposed Ra-RC model achieves F1 score 93.26% and 82.87% on the CCKS2017 and CCKS2019 datasets, respectively.

1. Introduction

Named entity recognition (NER) refers to the extraction of specific entities from unstructured texts, which plays a vital role in subsequent tasks, such as constructing knowledge graphs and personalized recommendation systems [1–3]. In recent years, with the rapid development of medical information technology, textual data of electronic medical records (EMRs) keep on increasing. As a fundamental Chinese medical information extraction task, named entity recognition of Chinese clinical EMRs has attracted extensive attention [4].

NER of clinical EMRs relates to the automatic discovery of all kinds of named entities closely associated with patients' health from EMRs, such as disease, drugs, or symptoms. Early researches in the CNER tasks mainly use lexicons-based and rule-based approaches [5, 6]. And then, a lot of statistical models are used for CNER [7, 8]. With the substantial increase in hardware computing power, the deep learning method has been successfully applied to CNER. At present, many research approaches have focused on exploring a generic domain model for migration. Traditional bidirectional long short-term memory networks [9, 10] and unsupervised pretraining of language models [11–14] are widely migrated to the CNER field. Both neural network algorithms have accomplished state-of-art achievement on the regular named recognition field. However, these models also have a room for improvement. First, the generality of the LSTM network leads to the model has no adequate capacity to extract features, where the extracted features are limited by the correctness of the dataset annotation and the context information. Second, the released versions of the pretraining model are more suitable for the general Chinese entity extraction. Both of them do not adapt to the characteristics of the EMR dataset, which underperforms on the task of medical entity extraction.

Moreover, the identification of Chinese clinical named entity recognition has been problematic. Firstly, many clinical named entities are multiword, and some of them are even being very long. It is not easy to distinguish the word boundaries of medical multiword in Chinese. What is more, the identical word and phrase can be divided into different kinds of named entities, for example, stroke can be delegated a modifier, and it can be additionally classified into particular disease and disease class and so on [15]. In addition, some

specific types of medical entities often have characteristics different from the general ones, especially in the radical-level characteristics of the entity. For instance, many characters of disease entities tend to have "疒" radicals, such as "病," and "痛." In ancient Chinese characters, "月" is related to human organs and flesh. Furthermore, many entities that consist of body parts often have "月" radicals, such as "脏," "脑," and "骨." These radical-level characteristics also have a significant reference value in determining labels, especially in complex medical entities consisting of multiple categories, such as the disease entity of "body parts and symptoms" format. However, this information has not been fully utilized by the regular named entity recognition model.

To address these issues, we propose a Ra-RC model which combines radical information with a deep learning structure. Above all, we adopt BiLSTM to encode radical characteristics. Simultaneously, RoBERTa is utilized to capture the characteristics of medical texts and generate characteristic representations. After that, we concatenate radical representations and characteristic representations and then use CRF to get predictive label sequences. Our proposed method has extensively evaluated its feasibility and utility on the CCKS2017 dataset and CCKS2019 dataset.

The main contribution can be summarized as follows:

(1) Considering the particularity of the medical entity and the underutilization of the radical-level information, we use BiLSTM to extract the radical characteristics

(2) Integrate radical-level information and deep learning model to solve the poor extraction performance of medical entities caused by migrating the general deep learning model

(3) The experimental results show that the Ra-RC model has a good performance on both datasets

2. Related Work

2.1. Clinical Named Entity Recognition. NER of EMRs has not only crucial practical significance but also high academic research value. Academics have done much research on it. At present, there are a large number of researches on NER in English clinical EMRs [16, 17]. For example, aiming at the lack of enough annotated data, Yang et al. [18] and Peters et al. [19] used transfer learning and semisupervised learning to extract entities, which could significantly improve the performance. There are many high-quality annotated data in the field of English CNER, such as JNLPBA, BC2GM, and NCBI. Due to the lack of high-quality EMRs and many nonstandard abbreviations, the Chinese CNER domain NER task is difficult [20].

An end-to-end deep learning method can be used to explore deeper features. The main network structure of this method is BiLSTM combined with CRF [21]. Li et al. [22] proposed a conditional random field algorithm that integrated characters, speech, and dictionary features based on establishing a medical dictionary. The experiments showed that these features were conducive to improving the CNER effect. Wang et al. [23] integrated dictionary features into the BiLSTM-CRF, and the results showed that prior knowledge helped improve the performance of the BiLSTM-CRF. Liu et al. [24] compared the CRF model requiring manual features with the LSTM-CRF without manual features and found that the F1 score of the LSTM-CRF on the i2b2 2010, 2012, and 2014 corpora was better than that of CRF.

However, the above CNER methods based on a deep neural network could not model the polysemy of words. That is, they could not solve the problem of polysemy. Therefore, Devlin et al. proposed a bidirectional encoder representation from transformer pretrained language model (BERT), which used bidirectional transformer encoders to capture potential semantic relations and generated a pretrained language model. Based on BERT, Liu et al. put forward the RoBERTa model to enhance the performance of BERT. And then, Lan et al.'s ALBERT model, a lightweight BERT model, was put forward for using two strategies to reduce the size of BERT. Dai et al. [25] compared the model performance after Word2vec and BERT were fused with BiLSTM-CRF, and the experiment showed that the model performance would be better if BERT was fused with the traditional BiLSTM-CRF model. However, these models failed to consider the characteristics of medical datasets thoroughly, and the performance on medical entity extraction was not highly effective.

2.2. Radical-Level Information. The specialization of the medical field leads to the particular linguistic structure of medical texts. Many experts have investigated on this characteristic. Peng et al. [26] put forward two types of Chinese radical-level hierarchical embeddings, and experimental results showed that radical-level semantics and sentiments on the sentence-level classification of emotions were better than char embeddings and word embeddings. A new deep learning technology referred to as "Radical Embedding" was proposed, and Shi et al. [27] conducted three experiments to verify its effectiveness. The results showed that the effect of radical embeddings was the same as competing methods and sometimes even better. Yin et al. [28] proposed BiLSTM-CRF based on radical features and used self-attention to capture character dependence. A new strategy was proposed to integrate dictionary information with characteristic presentation from BERT, and the F1 value of this method reached 91.60% and 89.56% on CCKS2017 and CCKS2018, respectively. However, in the existing researches, the information of radical has not been fully utilized.

3. Methods

3.1. Radical Characteristics. The Chinese electronic medical record datasets are different from the other datasets. In the CCKS2017 and CCKS2019 datasets, the frequency of radical-level feature is shown in Figure 1.

As the introduction mentioned, the radical "月" is often associated with the human organ, the radical "疒" is often related with the disease, and the radical "口" frequently appears in symptom entities. As shown in Figure 1, the Chinese five elements "metal, wood, water, fire, and earth" are

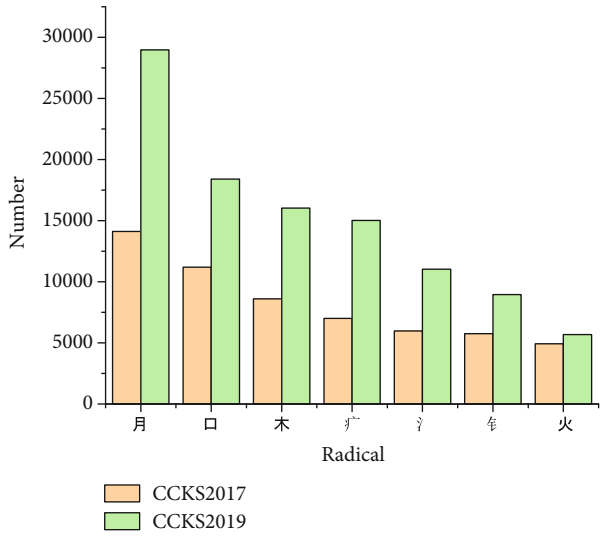

Figure 1: The occurrence frequency of each radical feature.

often included in medical entities. For example, "钅" correlates with microelement and drug names such as "钙" and "铁." "木" is related to "查体" and "脑血栓" and the name of the Chinese patent medicine. "氵" is associated with body fluids (plasma, tissue fluid, and lymphatic fluid) and symptoms such as "渗" and "溶." "火" has a relationship with inflammation-related entities such as "病灶" and "骨髓炎." "土" relates to modification words of a body part such as "壁" and "型." These radical features play an essential role in identifying medical entities [29].

The sources of the radicals include two parts: local dictionaries and Baidu Chinese dictionaries (https://hanyu.baidu.com). The local dictionary is created by crawling the familiar words of Xinhua Dictionary (http://xh.5156edu.com/). Thus, it generates a dictionary of key-value pairs in the form of "chars-radicals."

3.2. Design of Architecture. The proposed Ra-RC framework for the clinical named entity recognition task is shown in Figure 2. The framework mainly includes BiLSTM for radical-level representation, sequence modeling, and label inference layer. We train RoBERTa on both datasets where radical representations are extracting from BiLSTM. After that, we concatenate the char representations and radical-level representations and then feed them into CRF to decode.

3.2.1. BiLSTM for Radical-Level Representation. To make the most of the radical information, it needs to be extracted by a deep learning framework. From the perspective of theoretical and practical effects, both BiLSTM and RoBERTa are more suitable for feature extraction tasks, and RoBERTa enhances the performance based on BERT to have better expression ability. Therefore, this paper chooses these technologies to get contextual semantic information.

Figure 3 shows an overview of the radical-BiLSTM model. Formally, the inputs contain two parts: word embedding and radical embedding. Firstly, each word finds its corresponding radicals using a mapping dictionary which was constructed. Secondly, both words and radicals pass through the same trainable matrix of the lookup layer. Afterward, for the preliminary representations of radical messages, we concatenate both embeddings recorded as X_i, and then feed X_i into the BiLSTM network to extract the feature.

As shown in Figure 3, the radical-level representation $X_i = (x_1, x_2, \cdots, x_n)$ is taken as an input to the BiLSTM network. The BiLSTM network has two kinds of LSTM cells [30] that extract the feature in the forward $(\overrightarrow{h})$ and backward $(\overleftarrow{h})$ directions. The i^{th}characters $H_{Li} = (h_{l1}, h_{l2}, \cdots, h_{ld})$ are the output of hidden state in the backward direction, and the $H_{Ri} = (h_{r1}, h_{r2}, \cdots, h_{rd})$ is obtained after the forward LSTM. Afterwards, we can get the complete output hidden state $C_i = [(H_{Li}, H_{Ri})]$ of each position i by concatenating H_{Li} and H_{Ri}. An LSTM cell is made up of three gates which are used to select semantic information. The implement of LSTM cell is

$$i_t = \sigma(W_{xi}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}), \tag{1}$$

$$f_t = \sigma\left(W_{xf}x_t + b_{\text{if}} + W_{hf}h_{t-1} + b_{hf}\right), \tag{2}$$

$$o_t = \sigma(W_{xo}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}), \tag{3}$$

$$\tilde{c}_t = \tanh\left(W_{xc}x_t + b_{ic} + W_{hc}h_{t-1} + b_{hc}\right), \tag{4}$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t, \tag{5}$$

$$h_t = o_t \tanh\left(c_t\right), \tag{6}$$

where $\sigma(\cdot)$ denotes element-wise Sigmoid function and $\tanh(\cdot)$ denotes hyperbolic tangent functions. w is a weight matrix, and b is bias. i_t, O_t, and f_t are called input gate, output gate, and forget gate, respectively.

The output of the BiLSTM network is referred to as C_i, and characteristic representations, which are called P_i, are extracted from RoBERTa. The final representationsO_i splice C_i and P_i together.

3.2.2. Sequence Modeling. We use the famous architecture of RoBERTa, which consists of the bidirectional transformer encoder for feature extraction and sentence modeling. As an autocoding language model, the model can introduce noise data to reconstruct the original data. It randomly selects some words to be predicted through the Mask language model mechanism and shields them with the [MASK] symbol. The training process is shown in Figure 4. Firstly, input sentences are segmented and annotated according to character level.

Secondly, the sentence is processed as a distributed representation $Y = (Y_1, Y_2, \cdots, Y_t, \cdots, Y_n)$, consisting of token embedding, segment embedding, and position embedding. Y_t indicates the input status of each character:

$$Y_t = Y_{\text{token\_emb}} + Y_{\text{seg\_emb}} + Y_{\text{pos\_emb}}. \tag{7}$$

The transformer encoder is the most core component of the RoBERTa pretraining model, where multiheaded attention is the most critical module of the transformer unit.

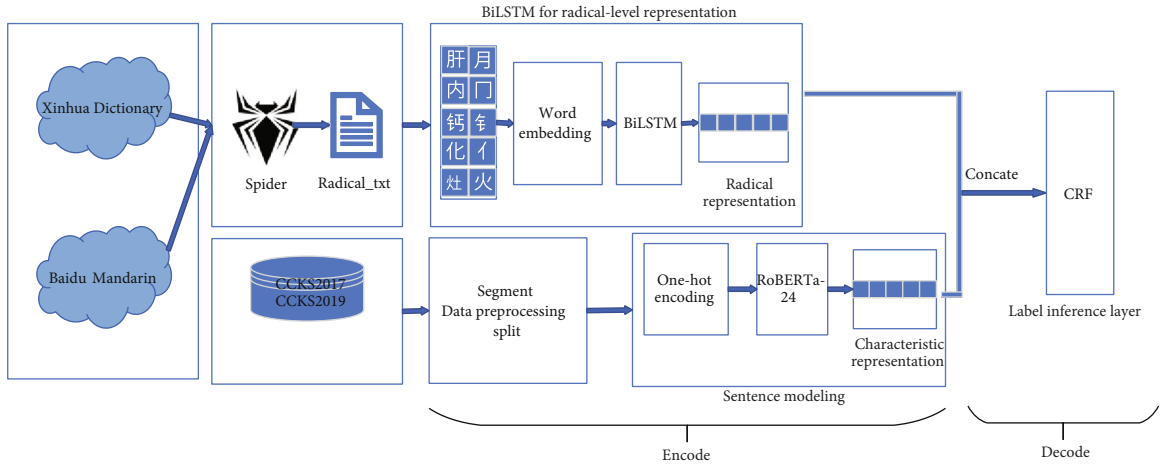

Figure 2: Ra-RC framework of CNER.

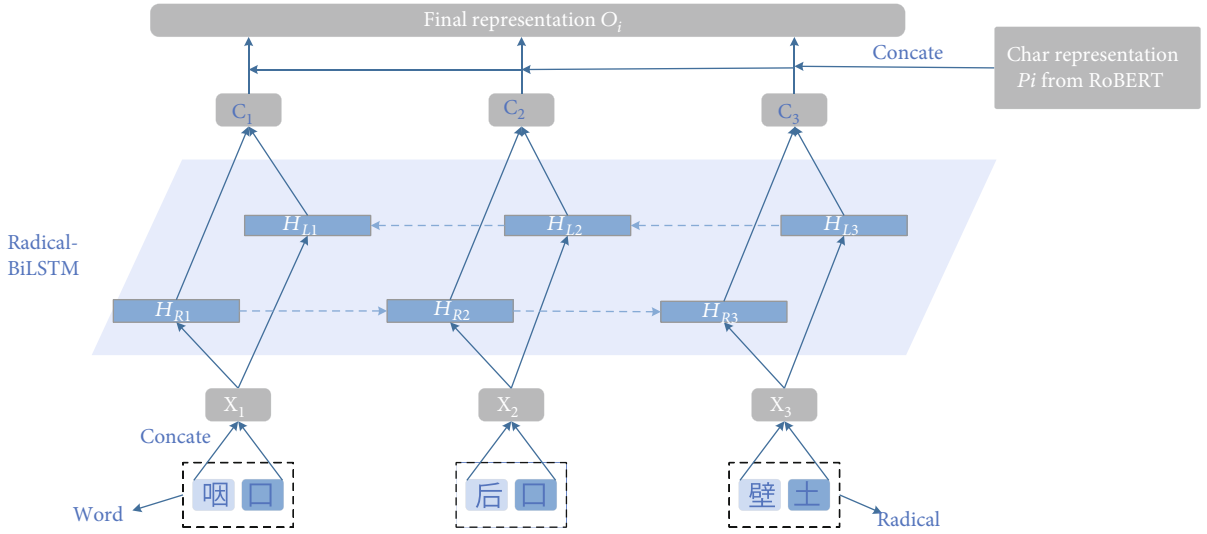

Figure 3: The model of radical-BiLSTM.

The multiheaded attention mechanism is utilized to capture character dependencies. The calculation of the single-head attention mechanism is shown in equation (8).

$$\text{head}_i = \text{Attention}\left(Y_t W_i^Q, Y_t W_i^K, Y_t W_i^V\right). \tag{8}$$

where W_i^Q, W_i^K, and W_i^V are the weight parameters for i^{th} calculation, respectively.

Then, the results of i^{th} calculations are stitched together. Moreover, we linearly transformed once more to obtain the results of the multiheaded attention calculation. The specific formula is shown in equation (9), where W^o is the weight parameter.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \cdots, \text{head}_n) W^o. \tag{9}$$

3.2.3. Label Inference Layer. Finally, we use a sequential CRF [31] layer to infer the correct tag sequence. $S = \{w_1, w_2, w_3, \cdots, w_n\}$ represents the tag sequence, $y = \{y_1, y_2, y_3, \cdots, y_n\}$ score corresponding to the $S = \{w_1, w_2, w_3, \cdots, w_n\}$ sequence.

$$s(S, y) = \sum_{i=1}^{n} E_{i,y_i} + \sum_{i=1}^{n+1} T_{y_{i-1}, y_i}, \tag{10}$$

where E is the emission matrix output by the RoBERTa layer, and $E_{i,j}$ represents the probability that the i^{th} word is classified into the j^{th} label; T is the transition matrix, and $T_{i-1,i}$ refers to the score transferred from label i-1 to i; and $s(S, y)$ refers to the score of the label prediction sequence y generated by the input sequence S.

In a given input sequence S, the CRF model is trained using the maximized log-likelihood function. The formulas

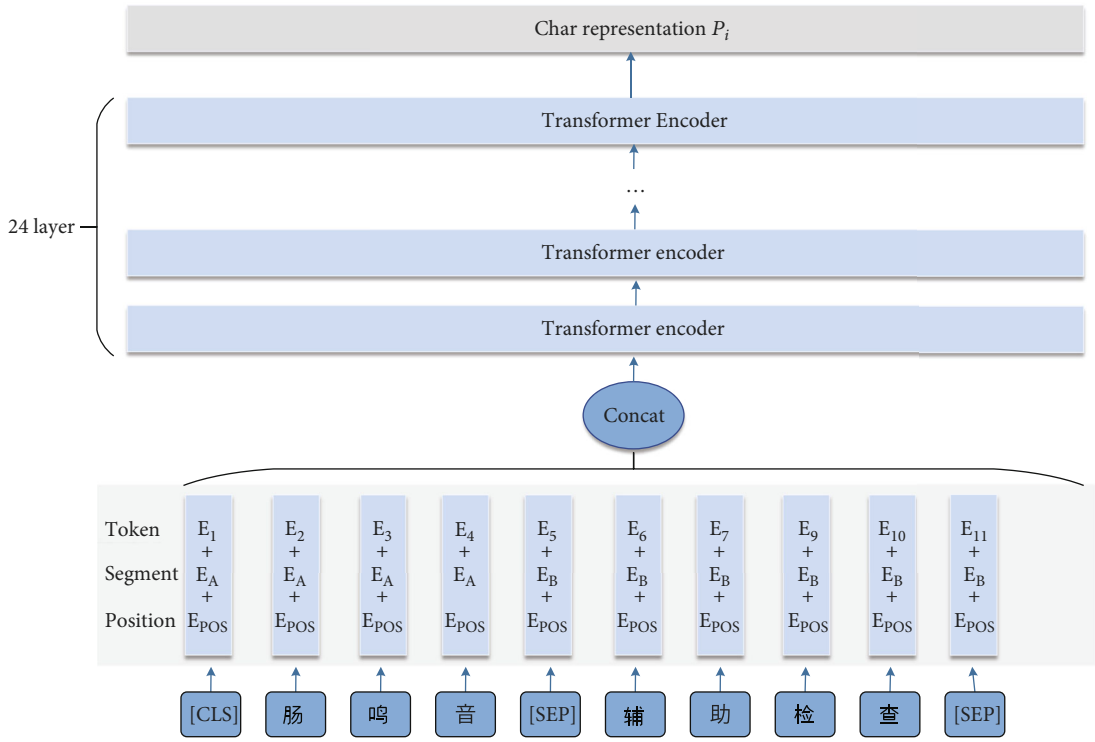

Figure 4: The RoBERTa for sentence modelling.

are shown in equations.

$$p(y \mid S) = \frac{e^{s(S,y)}}{\sum_{\tilde{y} \in Y_X} e^{s(S,\tilde{y})}}, \tag{11}$$

$$\log\left(p(y \mid S)\right) = s(S, y) - \log\left(\sum_{\tilde{y} \in Y_X} e^{s(S,\tilde{y})}\right). \tag{12}$$

The higher the $s(S, y)$ score, the greater the probability. Besides, Y_x is the sequence of all the possible tags for a given sentence S, and $\log\left(p(y \mid S)\right)$ is the defined loss function.

In the decoding process, the Viterbi algorithm is used to solve the CRF global optimal sequence label. The formula is given below, where $y*$ is the sequence in which the score function achieves the maximum value.

$$y^* = \arg\max_{\tilde{y} \in Y_X} s(X, \tilde{y}). \tag{13}$$

4. Experiments

4.1. Datasets. In this study, the CCKS2017 and CCKS2019 datasets are utilized to conduct experiments. The datasets contain actual EMR data, and a professional medical team manually annotated all EMR corpora. As we did not participate in the competition, the CCKS2017 dataset is incomplete. The numbers of the various types of medical entities are given in Figure 5.

The Beginning, Inside, Outside (BIO) sequence labeling system, a standard labeling strategy in the NER field, is adopted in this study. Note that "B" means the starting position of the medical entity, "I" represents the middle position of the medical entity, and "O" indicates that it is not a medical entity, such as "B-X," "I-X," and "O", where X represents the type of medical entity.

4.2. Evaluation. In this experiment, accuracy (P), recall rate (R), and F1 score are used as the comprehensive evaluation indexes of NER. The specific formulas are shown as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\%, \tag{14}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\%, \tag{15}$$

$$F_1 \text{ Measure} = \frac{2PR}{P + R} \times 100\%, \tag{16}$$

where TP is the number of correctly identified medical entities, FP is the number of unrelated medical entities identified, and FN is the number of unknown medical entities.

4.3. Environment. In this experiment, the NER model is based on the TensorFlow framework. Besides, the hardware and software environments are listed in Table 1.

4.4. Result and Discussion. In the experiment, the ratio of the training set to the test set is 5 : 1. The relevant parameters are set as follows: Radical embedding is initialized from a uniform distribution, where the embedding dim is set to 128. The hidden size of BiLSTM is 128. Char embedding is initialized via a pretrained model. The parameters of pretrained are

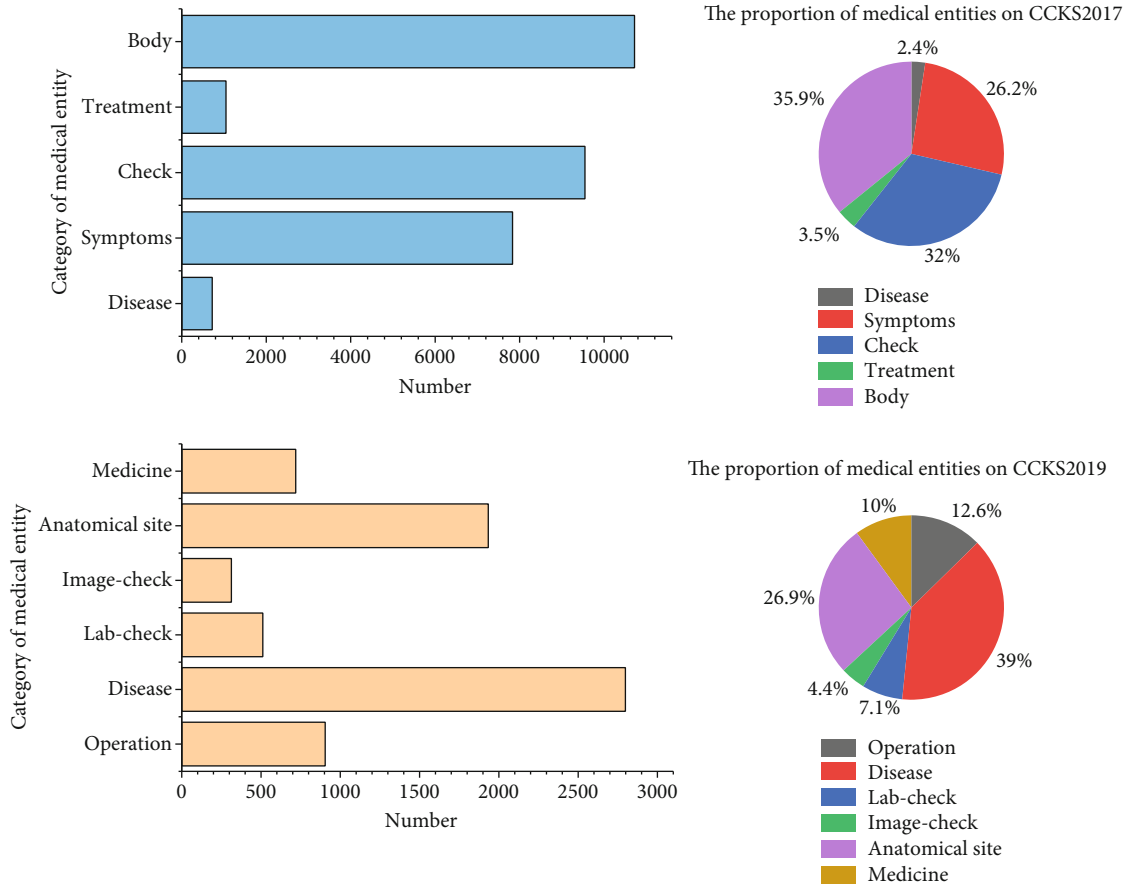

Figure 5: The numbers and proportion of both datasets.

Table 1: Experimental environment settings.

Item	Environment
Operating system	Ubuntu 18.04.5 LTS
CPU	i7-8700 @3.20GHz
GPU	NVIDIA GeForce RTX 2080Ti
Memory	31G
Python version	3.6
TensorFlow version	TensorFlow1.13.0
Keras version	2.2.4

all default parameters. Adam optimization algorithm [32] is adopted to optimize the model and the learning rate of 1*E*-5.

4.4.1. Compare Three Pretraining Models. To better integrate with the radical-level information, three pretraining models are trained and tested on the extraction of medical entities, and then, the best one was selected as our baseline model.

As observed in Tables 2 and 3, RoBERTa has the best effect of extracting entities. This reason is that RoBERTa has more data, more steps, and a large batch than BERT. Moreover, the RoBERTa-wwm-ext-large model has a 24-tier transformer to get a more robust capability of feature extraction.

4.4.2. Ablation Experiments. We take RoBERTa-CRF as the baseline model, and the comparison after adding the radical information is shown in Tables 4 and 5. RC stands for RoBERTa-CRF, and Ra-RC means adding radical information.

The extraction results of medical entities of CCKS2017 are shown in Table 4. The F1 values of the "Symptom" and "Check" categories are the highest, which are "96.53" and "96.36," respectively. Nevertheless, the recognition effect on "Treatment" is lacking, indicating that this type of entity is difficult to recognize. According to Figure 1, the sample size of this type of entity is small, accounting for only 3.60%. Hence, the neural network does not have enough samples to learn features, and the structure of entities is like the "Disease" entity, which is prone to classification errors. For example, consider "输卵管结扎术" and "输卵管结扎术后," they belong to different entity classes, where the former belongs to the "disease and diagnosis" entity class and the latter belongs to the "Treatment" entity class. On the whole, we can observe that our Ra-RC model based on radical-level information that BiLSTM extracts achieves the best performance with the F1 value of 93.26%, the precision of 94.14%, and the recall of 92.39% on the CCKS2017 dataset. The F1 value of the RA-RC is 1.2% higher than that of RC. In view of entity categories, the F1 values of all categories are higher than RC except for "Disease" and "Treatment."

Table 2: Comparison of three pretraining models on CCKS2017.

Model	Precision	Recall	F1 score
BERT+CRF	89.78	91.64	90.70
ALBERT+CRF	89.12	91.38	90.24
RoBERTa-wwm-ext+CRF	92.62	90.35	91.47
RoBERTa-wwm-ext-large+CRF	92.79	91.34	92.06

Table 3: Comparison of three pretraining models on CCKS2019.

Model	Precision	Recall	F1 score
BERT+CRF	81.82	79.33	80.56
ALBERT+CRF	81.91	77.68	79.74
RoBERTa-wwm-ext+CRF	81.83	79.33	80.55
RoBERTa-wwm-ext-large+CRF	82.29	79.69	80.97

Table 4: Comparison of RC and Ra-RC on CCKS2017.

		RC	Ra-RC
Disease	*P*	90.29	89.44
	R	89.76	89.17
	F1	90.02	89.31
Symptoms	*P*	94.08	95.00
	R	98.62	98.11
	F1	96.30	96.53
Check	*P*	95.52	95.73
	R	96.04	97.00
	F1	95.78	96.36
Treatment	*P*	60.97	61.25
	R	70.42	69.01
	F1	65.35	64.90
Body	*P*	89.14	89.29
	R	90.67	90.79
	F1	89.90	90.03
Total	*P*	92.79	94.14
	R	91.34	92.39
	F1	92.06	93.26

At the same time, the extraction results of medical entities of CCKS2019 are shown in Table 5. It can be seen from Table 5 that the Ra-RC model combining radical-level information has an improvement of 1.9% in terms of F1 value compared with the RC model, which is without radical-level information on the CCSK2019 dataset. All types of entities have increased except for the "disease" entity. "Medicine" has the best recognition performance of all entities where the F1 score reaches 92.77%. However, the recognition effect on the entity class of "Lab-Check" is insufficient, because some entities of "lab-check" in which the composition is complex often cause an error in boundary judgment. For instance,

Table 5: Comparison of RC and Ra-RC on CCKS2019.

		RC	Ra-RC
Image-check	*P*	78.48	81.23
	R	84.09	85.71
	F1	81.19	83.41
Operation	*P*	81.25	79.61
	R	75.00	77.56
	F1	78.00	78.57
Medicine	*P*	87.79	93.27
	R	89.23	92.27
	F1	88.50	92.77
Disease	*P*	80.02	78.74
	R	78.59	79.00
	F1	79.30	78.87
Lab-check	*P*	72.46	74.28
	R	66.28	69.96
	F1	69.23	72.06
Anatomical site	*P*	78.90	82.67
	R	86.57	85.23
	F1	82.55	83.93
Total	*P*	82.29	83.31
	R	79.69	82.44
	F1	80.97	82.87

these entities are always composed of "letters and other characters," such as "CEA," "F/T," "T–PSA," and "CA125." Moreover, some image-check entities are made up of letters, such as "OR" and "CT". Due to the similarity of "image-check" and "lab-check" structures, the model cannot analyze the boundary between two kinds of entity classes.

In order to further compare the performance of RC and RA-RC, we also calculate the F1 value, recall, and precision of different methods, as shown in Figure 6. The RC (17) and RA-RC (17) represent that experiments conducting on the CCKS2017 dataset, RC (19) and RA-RC (19) are evaluated on the CCKS2019 dataset. In Figure 6, the F1 score of the Ra-RC model is higher than RC on both datasets.

4.4.3. Comparative Experiment with Existing Research Work. In addition to the basic model described above, several researchers have conducted CNER studies on both datasets. For example, Li et al. [33] use a BiLSTM-CRF model combined with specialized word embeddings for CNER tasks. They use health domain datasets to create more prosperous and robust word embeddings. In addition to this, external health domain vocabulary is used to improve entity recognition results. Ouyang et al. [34] use the BiLSTM-CRF model combining the *n*-gram algorithm to the CNER tasks. At the same time, they introduce three types of external information as inputs to the model. Qiu et al. [35] use Chinese characters and dictionary features as input and then feed them into the

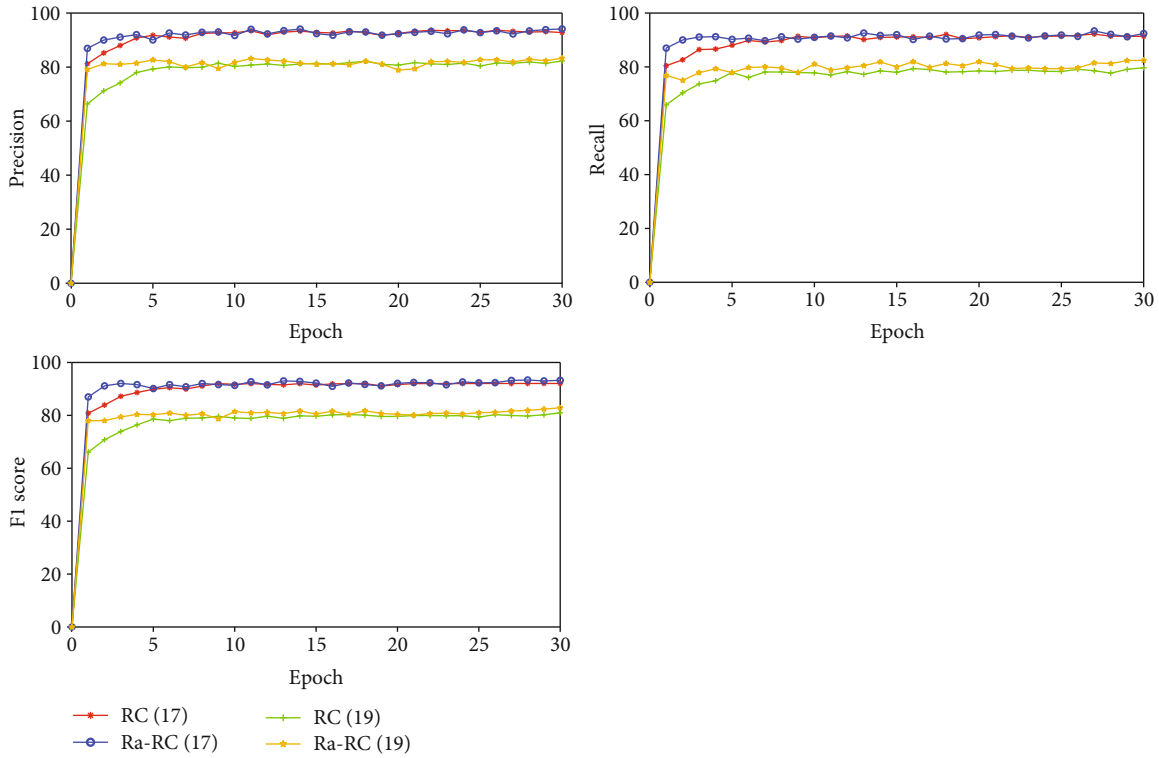

Figure 6: Evaluation results of RC and Ra-RC on the CCKS2017 and CCKS2019 datasets.

Table 6: Comparison of different methods on CCKS2017.

Method	Precision	Recall	F1 score
Li et al. [33]	—	—	87.95
Ouyang et al. [34]	—	—	88.85
Qiu et al. [35]	90.63	92.02	91.32
Tang et al. [36]	88.60	94.25	91.34
Wang et al. [23]	90.83	91.64	91.24
Luo et al. [4]	—	—	91.75
Yin et al. [28]	92.30	93.28	92.79
Yin et al. [28] *	92.27	93.73	93.00
Ours	94.14	92.39	93.26

residual dilated convolutional neural network combined with CRF to identify the entity. Tang et al. [36] propose a method that combines language model and multihead attention. Firstly, the sentence vectors are fed into BiGRU and the pretrained model. After that, this paper concatenates the output of them. Moreover, the output is given to the block of BiGRU and multihead attention. Wang et al. [23] construct a medical domain dictionary using relevant medical resources and then integrate the dictionary features and word vectors into the BiLSTM-CRF model to identify entities. Luo et al. [4] propose a CNER method that is based on ELMo and multitask learning. The ELMo is trained by adding the stroke features as input. Simultaneously, multitask learning is used to make full use of existing data to improve the model's performance. Yin et al. [28] propose the AR-CCNER model. The radical feature is extracted by the convolutional neural network (CNN). At the same time, this paper uses BiLSTM-attention to capture contextual features and the dependency between characters.

The experimental results are shown in Table 6. However, although all experiments are based on the CCKS2017 dataset, our dataset may not be the same as those of above researchers because we did not participate in the competition.

The results show that the Ra-RC achieves better precision and F1 score on the CCKS2017 dataset. Li et al.'s [33] model performs the worst because their approach is based on word segmentation, which causes the model to fail to identify word boundaries well. The latter is much larger than the former compared to the character set and the word set. This means that the corpus is not sufficient for the model to learn word embedding information effectively. What is more, the results show that the model of Yin et al. [25], which combines radical information and performs well on CCKS2017. The F1 score of the model has achieved 92.79%. This also proves that the radical feature is helpful for entity extraction. Moreover, they use self-attention to capture intercharacter dependencies, enhancing the extraction ability of entity, and the F1 score of the model has achieved 93.00%. However, this method is not compared with the pretraining model, which is the mainstream model in the CNER field. We compare the entity extraction effects of three pretraining models (BERT/ALBERT/RoBERTa) on the CCKS2017 dataset and combine them with another mainstream CNER technique (BiLSTM) to improve performance. The results show that pretraining the model helps to improve the performance of the model.

Table 7: Comparison of different methods on CCKS2019.

Method	Precision	Recall	F1 score
BERT-IDCNN-MHA-CRF (Liang et al. [37])	82.63	82.23	82.43
BiLSTM-CRF (baseline)	81.49	80.52	81.00
IDCNN-CRF (baseline)	80.56	81.47	81.01
Ours	83.31	82.44	82.87

The comparison of the CCKS2019 dataset is shown in Table 7. The experimental results show that the proposed model is superior to the baseline model in P, R, and F1 values. The F1 value of our model is slightly higher than the value of the model that Liang et al. [37] proposed, which indicates that the recognition ability of both is similar.

5. Conclusions

Aiming at the problem of insufficient medical entity extraction effect caused by the migration of the generic algorithm, we propose the Ra-RC model, which combines radical information extracted by BiLSTM with characteristic capturing by the pretrained model. To achieve a better entity extraction effect, we train three pretrained models for comparison. In addition, we introduce the radical feature, which can be seen as morphological information to enhance semantic information. After that, we concatenate both vectors and then feed them into CRF to get the corresponding label sequences. The experimental results on both datasets show that the Ra-RC method in this paper is superior to the baseline model.

A follow-up study will focus on how to distinguish entities more accurately with similar text structures. In addition, we will use this method in the following tasks, such as medical relation extraction and medical knowledge graph construction.

Data Availability

We have used the CCKS2017 and CCKS2019 datasets for our experiments. And datasets can be downloaded through the following link: https://github.com/baiyewww/Data.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 61972358) and Zhejiang Provincial Key Research and Development Program Project (Grant 2020C03071).

References

[1] Y. Yin, Q. Huang, H. Gao, and Y. Xu, "Personalized APIs recommendation with cognitive knowledge mining for industrial systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6153–6161, 2020.

[2] Y. Yin, Z. Cao, Y. Xu, H. Gao, R. Li, and Z. Mai, "QoS prediction for service recommendation with features learning in mobile edge computing environment," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 4, pp. 1136–1145, 2020.

[3] X. X. Yang, S. J. Zhou, and M. Cao, "An approach to alleviate the sparsity problem of hybrid collaborative filtering based recommendations: the product-attribute perspective from user reviews," *ACM/Springer Mobile Networks and Applications (MONET)*, vol. 25, no. 2, pp. 376–390, 2020.

[4] L. Luo, Z. H. Yang, Y. W. Song, and N. L. H. F. Lin, "Chinese clinical named entity recognition based on stroke ELMo and multi-task learning," *Chinese Journal of Computers*, vol. 43, no. 10, pp. 1943–1957, 2020.

[5] M. Song, H. Yu, and W. S. Han, "Developing a hybrid dictionary-based bio-entity recognition technique," *BMC medical informatics and decision making*, vol. 15, no. 1, pp. 1–8, 2015.

[6] A. Coden, G. Savova, I. Sominsky et al., "Automatically extracting cancer disease characteristics from pathology reports into a disease knowledge representation model," *Journal of Biomedical Informatics*, vol. 42, no. 5, pp. 937–949, 2009.

[7] D. Li, G. Savova, and K. Kipper-Schuler, "Conditional random fields and support vector machines for disorder named entity recognition in clinical texts," *Workshop on Current Trends in Biomedical Natural Language Processing*, pp. 94-95, 2008.

[8] Y. Feng, C. Ying-Ying, Z. Gen-Gui, L. H. Wen, and L. Ying, "Intelligent recognition of named entity in electronic medical records," *Chinese Journal of Biomedical Engineering*, vol. 30, no. 2, pp. 256–262, 2011.

[9] M. Miwa and M. Bansal, "End-to-end relation extraction using LSTMs on sequences and tree structures," *54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 2, pp. 1105–1116, 2016.

[10] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, https://arxiv.org/abs/1508.01991.

[11] X. P. Qiu, T. X. Sun, Y. G. Xu, Y. F. Shao, N. Dai, and X. J. Huang, "Pre-trained models for natural language processing: a survey," *Science China Technological Sciences*, vol. 63, no. 10, pp. 1872–1897, 2020.

[12] Z. Z. Lan, M. Chen, S. Goodman, G. Kevin, S. Piyush, and R. Soricut, "ALBERT: a lite BERT for self-supervised learning of language representations," *ICLR*, pp. 1–16, 2020.

[13] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," pp. 4171–4186, 2018, https://arxiv.org/abs/1810.04805.

[14] Y. H. Liu, M. Ott, N. Goyal et al., "RoBERTa: a robustly optimized BERT pretraining approach," 2019, https://arxiv.org/abs/1907.11692.

[15] P. D. Soomro, S. Kumar, A. A. Banbhrani, A. A. Shaikh, and H. Raj, "Bio-NER: biomedical named entity recognition using rule-based and statistical learners," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 12, pp. 163–170, 2017.

[16] G. H. Xu, C. Y. Wang, and X. F. He, "Improving clinical named entity recognition with global neural attention," in *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, vol. 11642, Cham, 2019.

[17] R. Chalapathy, E. Z. Borzeshi, and M. Piccardi, "Bidirectional LSTM-CRF for clinical concept extraction," pp. 7–12, 2016, https://arxiv.org/abs/1611.08373.

[18] Z. Yang, R. Salakhutdinov, and W. W. Cohen, "Transfer learning for sequence tagging with hierarchical recurrent networks," pp. 1–10, 2017, https://arxiv.org/abs/1703.06345.

[19] M. E. Peters, W. Ammar, C. Bhagavatula, and R. Power, "Semi-supervised sequence tagging with bidirectional language models," *55th Annual Meeting of the Association for Computational Linguistics*, pp. 1756–1765, 2017.

[20] J. Qiu, Q. Wang, Y. Zhou, T. Ruan, and J. Gao, "Fast and accurate recognition of Chinese clinical named entities with residual dilated convolutions," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 935–942, Madrid, Spain, 2019.

[21] G. Wu, G. Tang, Z. Wang, Z. Zhang, and Z. Wang, "An attention-based BiLSTM-CRF model for Chinese clinic named entity recognition," *IEEE Access*, vol. 7, pp. 113942–113949, 2019.

[22] X. Li, H. Zhang, and X. H. Zhou, "Chinese clinical named entity recognition with variant neural structures based on BERT methods," *Journal of Biomedical Informatics*, vol. 107, p. 103422, 2020.

[23] Q. Wang, Y. Zhou, T. Ruan, D. Gao, Y. Xia, and P. He, "Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition," *Journal of biomedical informatics*, vol. 92, article 103133, 2019.

[24] Z. Liu, M. Yang, X. L. Wang et al., "Entity recognition from clinical texts via recurrent neural network," *BMC medical informatics and decision making*, vol. 17, Suppl 2, p. 67, 2017.

[25] Z. Dai, X. Wang, P. Ni, Y. Li, G. Li, and X. Bai, "Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records," in *2019 12th international congress on image and signal processing, biomedical engineering and informatics (cisp-bmei)*, Suzhou, China, 2019.

[26] H. Peng, E. Cambria, and X. Zou, "Radical-based hierarchical embeddings for Chinese sentiment analysis at sentence level," in *The Thirtieth International Flairs Conference*, pp. 347–352, Marco Island, Florida, 2017.

[27] X. Shi, J. Zhai, X. Yang, Z. Xie, and C. Liu, "Radical embedding:delving deeper to Chinese radicals," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, vol. 10, pp. 594–598, Beijing, China, 2015.

[28] M. Yin, C. Mou, K. Xiong, and J. Ren, "Chinese clinical named entity recognition with radical-level feature and self-attention mechanism," *Journal of biomedical informatics*, vol. 98, article 103289, 2019.

[29] J. Z. Wang, "The generation, development and evolution of radicals and their types," *Traditional Chinese Medicine Culture*, vol. 1, no. 5–8, 1990.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 282–289, Williamstown, MA, USA, 2001.

[32] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005.

[33] Z. Li, Q. Zhang, Y. Liu, D. Feng, and Z. Huang, "Recurrent neural networks with specialized word embedding for Chinese clinical named entity recognition," *CEUR Workshop Proceedings*, , pp. 55–60, 2017.

[34] E. Ouyang, Y. Li, L. Jin, Z. Li, and X. Zhang, "Exploring n-gram character presentation in bidirectional RNN-CRF for Chinese clinical named entity recognition," *CEUR Workshop Proceedings*, , pp. 37–42, 2017.

[35] J. Qiu, Y. Zhou, Q. Wang, T. Ruan, and J. Gao, "Chinese clinical named entity recognition using residual dilated convolutional neural network with conditional random field," *IEEE Transactions on Nanobioscience*, vol. 18, no. 3, pp. 306–315, 2019.

[36] G. Q. Tang, D. Q. Gao, T. Ruan, Q. Ye, and Q. Wang, "Clinical electronic medical record named entity recognition incorporating language model and attention mechanism," *Computer Science*, vol. 47, no. 3, pp. 211–216, 2020.

[37] W. Liang, Y. H. Zhu, F. Zhan, and X. B. Ji, "Named entity recognition of electronic medical records based on BERT," *Journal of Hunan Univerity of Technology*, vol. 34, no. 4, pp. 54–62, 2020.