

## Research Article

# A BERT-BiGRU-CRF Model for Entity Recognition of Chinese Electronic Medical Records

Qiuli Qin <sup>1</sup>, Shuang Zhao <sup>1</sup>, and Chunmei Liu <sup>2</sup>

<sup>1</sup>School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup>Beijing Jiaotong University Health Service Center, Beijing 100044, China

Correspondence should be addressed to Qiuli Qin; qlqin@bjtu.edu.cn

Received 10 December 2020; Revised 4 January 2021; Accepted 9 January 2021; Published 28 January 2021

Academic Editor: Abd E.I.-Baset Hassanien

Copyright © 2021 Qiuli Qin 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.

Because of difficulty processing the electronic medical record data of patients with cerebrovascular disease, there is little mature recognition technology capable of identifying the named entity of cerebrovascular disease. Excellent research results have been achieved in the field of named entity recognition (NER), but there are several problems in the pre processing of Chinese named entities that have multiple meanings, of which neglecting the combination of contextual information is one. Therefore, to extract five categories of key entity information for diseases, symptoms, body parts, medical examinations, and treatment in electronic medical records, this paper proposes the use of a BERT-BiGRU-CRF named entity recognition method, which is applied to the field of cerebrovascular diseases. The BERT layer first converts the electronic medical record text into a low-dimensional vector, then uses this vector as the input to the BiGRU layer to capture contextual features, and finally uses conditional random fields (CRFs) to capture the dependency between adjacent tags. The experimental results show that the F1 score of the model reaches 90.38%.

## 1. Introduction

Named entity recognition is to extract entities with actual meaning from massive unstructured text data [1, 2]. In the medical field, medical entities mainly include symptoms, examinations, diseases, drugs, treatments, operations, body parts etc., and are an important part of the establishment of a medical knowledge base. Chinese electronic medical record (CMR) [3] is a combination of structured and unstructured texts, which generally include not only patient information, but also a large amount of medical knowledge, but it is difficult to process. With the development of deep learning technology, entity recognition algorithms have applied research in many fields, but they lack applications in the field of cerebrovascular diseases (CVD) [4].

Cerebrovascular diseases have become one of the most threatening diseases to human health in the world due to the four characteristics [5, 6]. The treatment of cerebrovascular diseases highly depends on the doctor's experience. With the increase in the number of patients with CVD, there is a

greater demand for cerebrovascular disease physicians. Since the training cycle of professional doctors is relatively long [7, 8], it will cause an imbalance in supply and demand of "more patients and fewer doctors". With the introduction of the concept of "AI + Medical," the use of machine learning technology to assist diagnosis and treatment, that is, through the construction of a complex model, the feedback mechanism is used to continuously optimize the parameters of the model then, use the existing clinical data and neuroimaging data in the hospital to diagnose and treat cerebrovascular diseases or predict recurrence. On the one hand, auxiliary diagnosis and treatment decision-making is helpful to improve the professional level of doctors and improve the quality of CVD medical services. On the other hand, it can optimize the uneven distribution of medical resources [9]. At present, the scientific research of machine learning in the field of CVD mainly focuses on two aspects: diagnosis and prognosis prediction of cerebrovascular disease: (1) From the perspective of CVD diagnosis, most scholars use structured data to nest machine learning models to complete

disease diagnosis. The literature [10–12] established a joint diagnosis model based on logistic regression method and XGBoost machine learning method by collecting clinical data of demographic characteristics. (2) From the perspective of prognosis prediction, the use of machine learning methods for risk prediction has gradually become the trend of disease prediction, while machine learning methods such as random forest, decision tree, SVM, and other machine learning methods have achieved certain research results in the prediction of cerebrovascular diseases. Literature [13–15] constructed logistic regression, k-NN, random forest, decision tree, and SVM machine learning models based on follow-up data, and verified the advantages of machine learning models in cerebrovascular disease risk prediction. It shows that the effect of neural network learning score is better. In short, from the analysis of clinical data sources, CVD medical data includes cerebrovascular disease imaging data, follow-up data, electronic medical records, and other data. The focus of most scholars is still on structured data such as follow-up data and neuroimaging data, while the focus on electronic medical record data in the field of CVD is slightly lower. At present, the increase in the number of CVD patients is accompanied by the ever-increasing number of electronic medical records for CVD patients. Electronic medical records can provide scholars with more data resources. For the processing of unstructured text information in electronic medical records, named entity recognition (NER) is a key step, and there are relatively few researches dedicated to named entity recognition in the field of cerebrovascular.

The current research on named entity recognition focuses on three aspects: (1) From the perspective of traditional entity recognition methods, traditional methods include methods based on dictionaries and rules [16–23]. This method relies heavily on domain dictionaries and domain experts. The selection of features is done manually, and subjectivity and labor costs are relatively large. With the development of machine learning technology [19, 20], more and more scholars are paying attention to models such as conditional random field (CRF), Hidden Markov Model (HMM), Support Vector Machine (SVM). However, NER based on traditional machine learning technology has higher requirements for feature selection [21–23], and the quality of feature selection directly affects the effect of entity recognition. (2) From the perspective of deep learning methods: with the development of deep learning technology, literature [24–27] confirmed the advantages of deep neural network technology by comparing traditional CRF models, that is, deep neural network technology has less artificial feature intervention than traditional methods and can obtain higher accuracy and recall rate. Deep learning can automatically extract word features, reduce the subjectivity of feature selection, and help further improve the accuracy of recognition results. Therefore, it is better than traditional statistical algorithms such as CRF and HMM. The common single-entity recognition neural network generally only considers the sample input and lacks in-depth thinking about the output relationship. Based on the idea of model fusion, most scholars usually use LSTM-CRF as the main

framework to solve the deficiencies of neural network models. The available literature [28–32] uses traditional word2vec, Glove, and other word vector methods, uses a BiLSTM-CRF model as the core, and adds a CNN model, attention mechanism, RNN model, etc., to the core framework. Furthermore, the word vector undergoes continuous fine-tuning of the parameters, resulting in the final recognition achieving more accurate recognition. The process of parameter tuning is employed to set the hyper-parameters of the model. However, for the BiLSTM-CRF model, there are more parameter settings, and the model training time is longer. Literature [26–31, 33, 34] proposed the BGRU neural model, which has simple results and high computational efficiency, can make full use of context information to eliminate entity ambiguity, and has some good effects in the field of entity recognition. (3) From the perspective of pre-training models, the above-described pre-processing models all use traditional word vector methods such as word2vec and Glove. This method focuses on the feature extraction between words and often ignores word context information. In order to improve this problem, as the Google BERT pre-training model is proposed, the literature studies [35–38] combine the BERT word embedding model on the basis of the traditional BiLSTM-CRF model and consider the polysemy of a word in combination with the context. The  $P$  value, R value, and F1 score have all been improved. It can be seen that BERT has strong semantic analysis capabilities.

In order to solve the problems of ignoring context information, low model efficiency, and susceptibility to word segmentation in electronic CVD medical entity recognition processing. We propose a BERT-BiGRU-CRF neural network model to identify named entities in electronic medical records of cerebrovascular diseases. Specifically, the BERT layer first converts the electronic medical record text into a low-dimensional vector, then inputs the vector into the BiGRU layer to capture contextual features, and finally uses a CRF to capture dependency between adjacent tags. The entity extraction model proposed in this paper has achieved good recognition results.

## 2. BERT-BiGRU-CRF Model Construction

In the NER field, the use of deep neural network models for entity recognition has become the mainstream. This article uses BiGRU-CRF as a benchmark to extract named entities in the field of cerebrovascular diseases. The reason why the BERT pre-training language model is chosen is that the text vector is used as the input of the model, and the granularity of Chinese division is divided into character-level and word-level. Existing research shows that character-level pretraining schemes show better results [37, 39], while the BERT pre-training language model is a character-level pretraining program. That is, each word in the text is converted into a vector by querying the word vector table, as the model input; the model output is the vector representation combined with the context.

The overall structure of the BERT-BiGRU-CRF model is shown in Figure 1. The model is mainly divided into 3 layers. The first layer is the BERT layer. Through the BERT pre-training language model, each word in the sentence is converted into a low-dimensional vector form. The second layer is the BiGRU layer, which aims to automatically extract semantic and temporal features from the context. The third layer is the CRF layer, which aims to solve the dependency between the output tags to obtain the global optimal annotation sequence of the text.

In this study, the named entity recognition model was used to identify the medical named entity in the electronic medical record of cerebrovascular disease. The specific steps are as follows:

- (1) EMR data preprocessing, that is, processing the original electronic medical record text data set and express the electronic medical record text set as  $H = \{h_1, h_2, \dots, h_n\}$ , where the  $i$ -th electronic medical record text is expressed as  $h_i = \langle w_{i1}, w_{i2}, \dots, w_{in} \rangle$ . Predefined entity category  $C = \{c_1, c_2, \dots, c_m\}$  is divided and annotated according to character-level, and the characters and predefined categories are separated by spaces when annotated.
- (2) Construct the electronic medical record text training data set.
- (3) Model training, that is, training the BERT-BiGRU-CRF named entity recognition model. Take the electronic medical record test text collection  $D_{\text{test}} = \{d_1, d_2, \dots, d_N\}$  as input, and take the entity and its corresponding category pair as output:  $\{\langle m_1, c_1 \rangle, \langle m_2, c_2 \rangle, \dots, \langle m_p, c_p \rangle\}$ , where the entity  $mi = \langle h_i, b_i, e_i \rangle$  represents the entity that appears in the document;  $h_i, b_i$ , and  $e_i$ , respectively, represent the start and end positions of  $m_i$  in  $h_i$ , and no overlap between entities is required, ie  $e_i < b_i + 1$ .  $C_{mi}$  represents the predefined category of the entity  $m_i$ , then calculates the  $F1$  score according to the precision rate and the recall rate, and uses the  $F1$  score as the model comprehensive evaluation index.

**2.1. BERT Pre-training Language Model.** Bidirectional Encoder Representation from Transformers (BERT) [40] is an unsupervised and deep bidirectional language representation model for pre-training. In order to accurately represent the context-related semantic information in the EMR, it is necessary to call the interface of the model to obtain the embedded representation of each word in the electronic medical record. BERT uses the deep two-way transformer encoder as the main structure of the model. Transformer introduces the self-attention mechanism and also draws on the residual mechanism of the convolutional neural network, so the training speed of the model is fast and the expression ability is strong. And also abandoning the RNN loop structure, the overall structure of the BERT model is shown in Figure 2.

En is the coded representation of the word, Trm is the transformer structure, and Tn is the word vector of the target word after training. The operating principle of the model is to use the transformer structure to construct a multi-layer bidirectional Encoder network, which can read the entire text sequence at one time, so that each layer can integrate the contextual information. The input of the BERT model adopts the embedding addition method. By adding three vectors, Token Embeddings, Segment Embeddings, and Position Embeddings, the purpose of pre-training and predicting the next sentence is achieved. In Chinese electronic medical record text processing, the semantics of characters or words in different positions have different semantics. Transformer indicates that the information embedded in the sequence of the tag sequence is its relative position or absolute position information, as shown in the following formulae:

$$PE(P_{\text{pos}}, 2i+1) = \cos\left(\frac{P_{\text{pos}}}{1000^{(2i/d_{\text{model}})}}\right), \quad (1)$$

$$PE(P_{\text{pos}}, 2i) = \sin\left(\frac{P_{\text{pos}}}{1000^{(2i/d_{\text{model}})}}\right), \quad (2)$$

where  $P_{\text{pos}}$  is the position of the word in the text,  $i$  represents the dimension, and  $d_{\text{model}}$  is the dimension of the encoded vector. The odd position is encoded using the cosine function. Even positions are coded using a sine function.

In order to better capture word-level and sentence-level information, the BERT pre-training language model is jointly trained by two tasks: Masked Language Model and Next Sentence Prediction. The Masked LM model [36] is similar to cloze filling. 15% of the words in the random mask corpus are marked with the "MASK" form, and then the BERT model is used to correctly predict the masked words. The strategy adopted in the training is that for 15% of the words, only 80% of the words are actually replaced with [mask], 10% of the words will be randomly replaced with other words, and the remaining 10% are unchanged. The Next SP model is to train the model to understand the relationship between sentences, that is, to judge whether the next sentence is the next sentence of the previous sentence. The specific method is to randomly select 50% correct sentence pairs from the text corpus, and 50% randomly select sentence pairs to judge the correctness of the sentence pairs. The Masked LM word processing and Next SP sentence processing are jointly trained to ensure that the information is represented by the vector of each word, so the model is comprehensive and semantically accurate. It fully depicts the characteristics of the character-level, word-level, sentence-level and even the relationship between sentences and increases the generalization ability of the BERT model.

**2.2. BiGRU Layer.** Gated Recurrent Unit (GRU) [34] gated recurrent unit structure is a variant of long and short-term memory neural network (LSTM). The LSTM structure includes forget gates, input gates and output gates. In traditional recurrent neural network (RNN) training, gradient disappearance or explosion problems often occur. LSTM

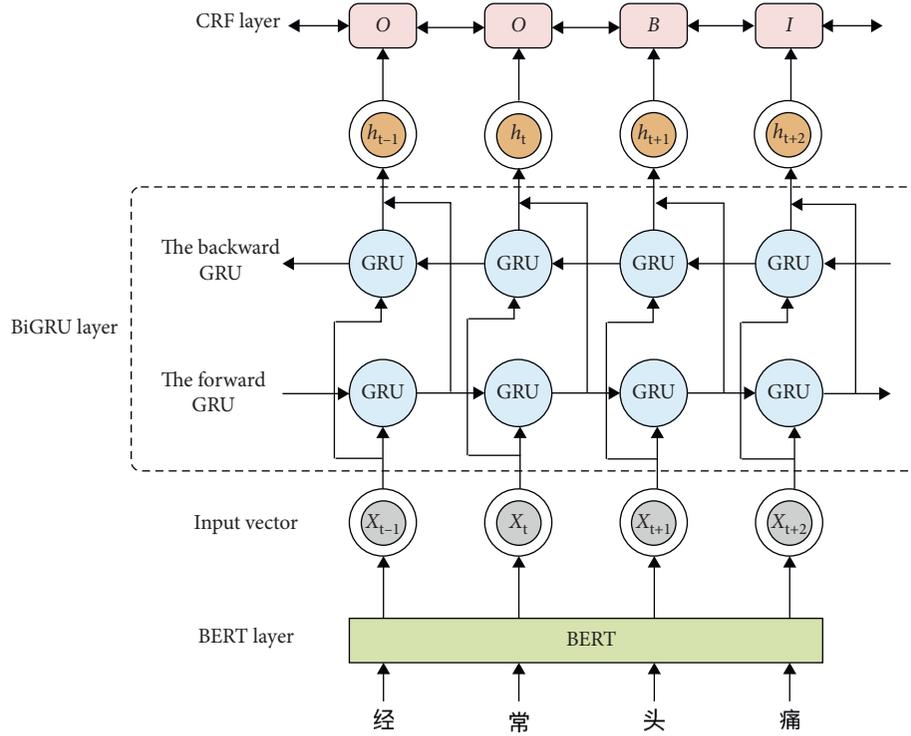


FIGURE 1: BERT-BiGRU-CRF model structure.

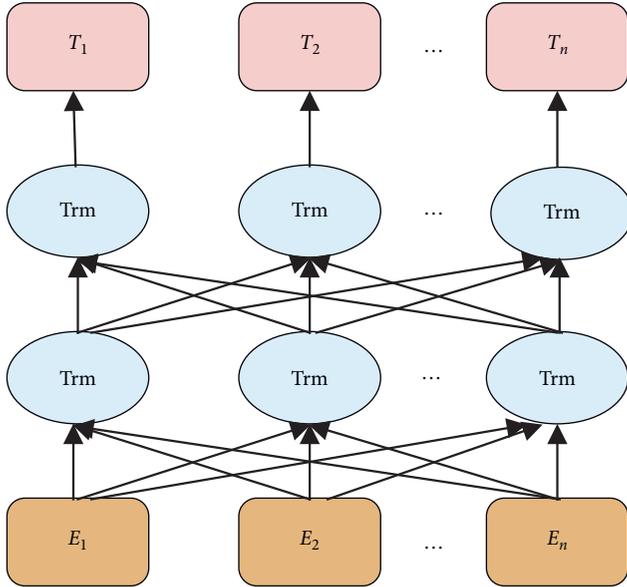


FIGURE 2: BERT network model.

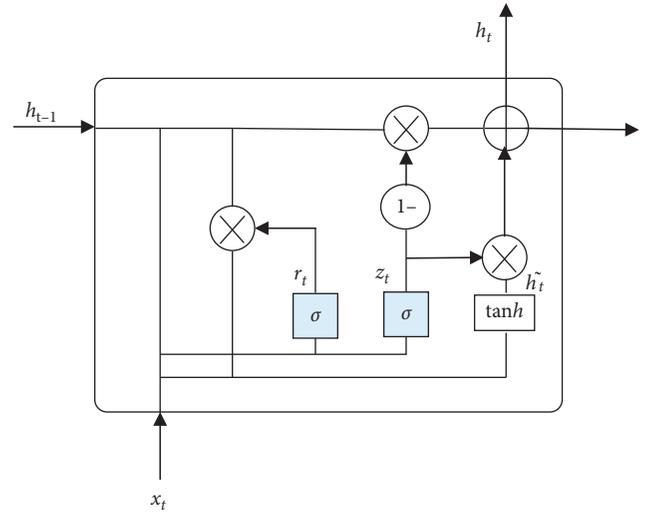


FIGURE 3: GRU structural unit.

only solves the problem of gradient disappearance to a certain extent, and the calculation is time-consuming. The GRU structure includes an update gate and a reset gate, and the GRU combines the forget gate and the input gate in the LSTM into an update gate. Therefore, GRU not only has the advantages of LSTM, but also simplifies its network structure. In the task of entity recognition of cerebrovascular disease electronic medical record, GRU can extract features effectively. Its network structure is shown as in Figure 3.

In the GRU structure, the update gate is  $z$  and the reset gate is  $r$ . The update gate  $z_t$  is to calculate how much electronic medical record information of the previous hidden layer state  $h_{t-1}$  needs to be transmitted to the current hidden state  $h_t$ . If  $z_t$  takes the value  $[0, 1]$ , it needs to be transmitted when it is close to 1, and the information needs to be ignored when the value is close to 0. The reset gate  $r_t$  calculation formula is similar to the update gate principle, but the weight matrix is different. The calculation of  $z_t$  and  $r_t$  is shown in formulae (3) and (4). First, the electronic medical record data  $x_t$  input at time  $t$ , the state  $h_{t-1}$  of the hidden layer at the previous time, and the corresponding weights

are, respectively, multiplied and added to the  $\sigma$  function. After the calculation of  $z_t$  and  $r_t$  is completed, the content that needs to be memorized at time  $t$  can be calculated. Secondly, use the reset gate to determine the hidden state of the electronic medical record at  $t - 1$ . The information that needs to be ignored at time  $t$ . Then, input  $r_t$ ,  $h_{t-1}$ ,  $x_t$ , and use tanh function to calculate the candidate hidden strong state. Finally,  $h_t$  transfers the cerebrovascular disease electronic medical record information retained in the current unit to the next unit; that is, at time  $t$ , the product of  $z_t$  and  $\tilde{h}$  represents the cerebrovascular disease information that the hidden unit  $h_t$  needs to retain. The product of  $(1 - z_t)$  and  $h_{t-1}$  indicates how much information needs to be forgotten. The calculation is shown in formulae (5) and (6) for details:

$$z_t = \sigma(w_z * [h_{t-1}, x_t]), \quad (3)$$

$$r_t = \sigma(w_r * [h_{t-1}, x_t]), \quad (4)$$

$$\tilde{h} = \tanh(w_h * r_t \mu_t h_{t-1}, x_t), \quad (5)$$

$$h_t = (1 - z_t)\mu h_t - 1 + z_t\mu\tilde{h}_t, \quad (6)$$

where  $x_t$  is the input of the electronic medical record of cerebrovascular disease at time  $t$  and  $h_{t-1}$  is the state of the hidden layer at the previous time;  $h_t$  is the hidden state at time  $t$ ;  $w$  is the weight matrix;  $w_z$  is the update gate weight matrix and  $w_r$  is the reset gate weight matrix;  $\sigma$  is the sigmoid nonlinear transformation function and  $\tanh$  is the activation function;  $\tilde{h}$  is the hidden state of candidate.

From the operating principle of the GRU unit, it can discard some useless information, and the structure of the model is simple, which reduces the computational complexity. However, the simple GRU cannot fully utilize the context information of the electronic medical record. Therefore, this paper designs the backward GRU to learn the backward semantics, and the GRU neural network forwards and backwards to extract the key features of the named entity in the electronic medical record of cerebrovascular disease, namely, the BiGRU model. The specific structure is shown in Figure 4. Based on the GRU principle, forward GRU is to obtain the above semantic feature ( $h_t$ ), and backward GRU is to obtain the following semantic features  $h_t$ , and finally, the above and the following semantic features are combined to get  $h_t$ . Refer to formulae (7) and (8) for details:

$$\vec{h}_t = G\vec{R}U(x_t), \quad (7)$$

$$\overleftarrow{h}_t = \overleftarrow{G}R\overleftarrow{U}(x_t), \quad (8)$$

$$h_t = \langle \vec{h}_t, \overleftarrow{h}_t \rangle, \quad (9)$$

where  $\vec{h}_t$  is the hidden layer state, the purpose is to obtain the above information from the GRU;  $\overleftarrow{h}_t$  is the hidden layer state, the purpose is to obtain the following information from the GRU;  $G\vec{R}U(x_t)$  means that it is represented by features from front to back;  $\overleftarrow{G}R\overleftarrow{U}(x_t)$  means that it is

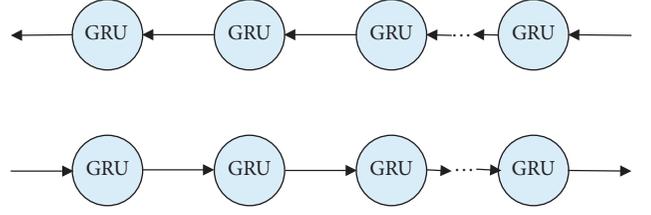


FIGURE 4: BiGRU structure.

represented by the back-to-front feature; and the final hidden layer state of  $h_t$  is the feature of the electronic medical record report.

**2.3. CRF Layer.** The NER problem can be regarded as a sequence labeling problem. The BiGRU layer outputs the hidden state context feature vector  $h$ , denoted as  $h = (h_1, h_2, \dots, h_t)$ . This vector only considers the context information in the electronic medical record and does not consider the inter-label dependencies. Therefore, this paper adds a CRF layer to label the global optimal sequence and converts the hidden state sequence  $h = (h_1, h_2, \dots, h_t)$  into the optimal label sequence  $y = (y_1, y_2, \dots, y_t)$ . CRF calculation principle [34]: firstly, for the specified electronic medical record input sequence  $x = (x_1, x_2, \dots, x_t)$ , it calculates the score of each location, shown in formula (10). Secondly, calculate the probability of normalized sequence  $y$  through the Softmax function, shown in formula (11). Finally, the label sequence with the highest score is calculated using the Viterbi algorithm, shown in formula (12):

$$\text{score}(h, y) = \sum_{t=1}^T A_{y_{t-1}, y_t} + \sum_{t=1}^T W_{y_t}^T h_t, \quad (10)$$

$$P\left(\frac{y}{h}\right) = \frac{e^{\text{score}(h, y)}}{\sum_{y' \in Y(h)} e^{\text{score}(h, y')}} \quad (11)$$

$$y^* = \arg \max_{y' \in Y(h)} (h, y'), \quad (12)$$

where  $A$  is the transfer score matrix between tags;  $\text{score}(h, y)$  is the position score;  $W_{y_t}^T$  is the parameter vector;  $p(y/h)$  normalized probability function;  $Y(h)$  represents all possible tag sequences; and formula (10) is to calculate the score ( $h, y$ ) of each position in the input sequence from the output feature matrix of the BiGRU layer and the CRF transition matrix.

**2.4. Training Process.** The process of deep network model training is a process of repeatedly adjusting parameters so that loss reaches a minimum. However, due to the strong learning ability of deep network models, the problem of model generalization is prone to occur. For example, the problem of model under-fitting and over-fitting leads to poor adaptability of the model to new sample data. Therefore, regularization methods can generate many models with small parameter values. In other words, such

models have strong anti-interference ability and can adapt to different datasets and different “extreme conditions”. It can increase the generalization capabilities of the model during the network training process. The method to solve the problem in this paper is the L2 regularization method, which can avoid the over-fitting problem, that is, adding regularization calculation to the cost function, shown in the following formula:

$$\text{loss} = E_{\text{in}} + \lambda \sum_j w_j^2, \quad (13)$$

where  $E_{\text{in}}$  is the training sample error that does not include the regularization term;  $\lambda$  is the adjustable parameter of regularization; and  $w_i$  represents the weight parameter.

### 3. Experimental Design

**3.1. Data Preparation.** The experimental data in this article was obtained from the Ai’ai medical electronic medical records website. The electronic medical record data is a total of 1,300 electronic medical records related to cerebrovascular diseases, which are composed of general patient information, chief complaint, medical history, physical examination, and diagnosis. In addition, this article sorts out the types of entities in the published papers related to named entities in electronic medical records, as shown in Table 1. According to the frequency of occurrence of entities in published literature, electronic medical record entities for cerebrovascular diseases are divided into five entities: disease, symptom, body part, examination, and treatment, which are also proposed by CCKS.

**3.2. Data Preprocessing.** The electronic medical record information is preprocessed, that is, line breaks and invalid characters, etc., are removed, and 36400 sentences are finally obtained. The testing dataset and the training dataset are divided into 2:8. The labeling system used in this article is BIO labeling. The five types of entities are disease, symptoms, body parts, examination, and treatment. Therefore, there are 11 labels, namely, O, Disease-B, Disease-I, Body-B, Body-I, Symptom-B, Symptom-I, Examination-B, Examination-I, Treatment-B, and Treatment-I. We conduct named entity labeling with doctors. Among the 300 medical records, we designated two annotators to annotate them at the same time, used Cohen’s kappa to calculate the consistency of the annotations, and obtained a kappa value of 0.8. The labels to be predicted are shown in Table 2.

**3.3. Experimental Settings.** The experimental model in this article is built using tensorflow deep learning framework and Python programming language. The parameter update method is to update the parameters of the BiGRU-CRF model, and BERT is a fixed parameter. Table 3 lists the hyperparameter values of the experimental model in this article. These values have been modified according to relevant literature [34–38, 41] and have not been adjusted for the cerebrovascular electronic medical record data set in this article. The model parameter optimization in this

TABLE 1: Sorting of entities.

Entity category	Literature
Symptom	[28, 31, 32, 37, 39, 40, 42]
Treatment	[28, 31, 32, 37, 39, 40, 42]
Body part	[31, 32, 34, 42]
Examination	[28, 31, 32, 40]
Disease	[28, 31, 37, 40, 42]
Surgery	[34, 40]
Anatomy	[31, 32]
Independent symptoms	[31, 39]
Drug	[31, 39]

TABLE 2: BIO annotation example.

Text	Label	Meaning
头	Body-B	The beginning of the body part entity
部	Body-I	The middle of the body part entity
疼	Symptom-B	The beginning of the symptom entity
痛	Symptom-I	The middle of the symptom entity
3	O	Non-entity
小	O	Non-entity
时	O	Non-entity

TABLE 3: Parameter settings.

Parameter	Value
Dropout	0.5
Embedding	50
GRU-dim	200
Regularized intensity	1’10-8
Batch_size	64
Lr (learning rate)	0.015
LSTM_dim	200
Dr (decay rate)	0.05

paper adopts the stochastic gradient descent method (SGD), the initial learning rate is 0.015, the update of the learning rate adopts the step decay method, and the decay rate is 0.05. The model has achieved good experimental results in the training set and test set.

**3.4. Evaluation.** This article uses the most commonly used evaluation index in the field of named entity recognition: precision rate ( $P$ ), recall rate ( $R$ ), and  $F1$  score ( $F1$ ). That is,  $P$  is the recognition rate of correctly recognized named entities,  $R$  is the rate of correctly recognized named entities in the test set, and  $F1$  is the harmonic average of  $P$  and  $R$ , which is the comprehensive evaluation index of the model. Among them, the higher the  $P$  and  $R$  values, the higher the accuracy and the recall rate, but in fact, the two are contradictory in some cases. Therefore, the  $F1$  score is often used to evaluate the overall performance of the model. The calculation formula is:

$$P = \frac{T_{\text{num\_entities}}}{S_{\text{num\_entities}}}, \quad (14)$$

$$R = \frac{T_{\text{num\_entities}}}{C_{\text{num\_entities}}}, \quad (15)$$

$$F1 = \frac{2P * R}{(P + R)}, \quad (16)$$

where  $T_{\text{num\_entities}}$  is the number of correct entities identified;  $S_{\text{num\_entities}}$  is the total number of all entities identified; and  $C_{\text{num\_entities}}$  is the number of entities in the test set.

## 4. Experimental Results

*4.1. Comparative Experiment Analysis.* In order to prove the entity recognition effect based on the BERT-BiGRU-CRF model, this article uses BiLSTM-CRF as the baseline model, and the selected comparison models are BiLSTM-CRF, BERT-BiLSTM-CRF, and BERT-BiGRU-CRF. Model introduction:

- (1) BiGRU-CRF model: this model inputs word vectors into the model for training.
- (2) BiLSTM-CRF model: this model is a classic model in the NER field. It uses trained word vectors and then uses the BiLSTM-CRF model to extract entities.
- (3) BERT-BiLSTM-CRF model: this model is based on the Google BERT model. Many scholars have embedded BERT in the BiLSTM-CRF model and achieved better recognition results in NER research.

*4.2. Model Performance Comparison.* The comparison model proposed in this paper first uses electronic medical record data for training and then uses a test set for testing. The specific comparison results are shown in Table 4.

From the comparison results of Table 4 and Figure 5, we can see that in terms of comprehensive evaluation indicators. In terms of precision rate, recall rate, and  $F1$  score, the BERT-BiGRU-CRF model proposed in this article has increased by 2.9%, 5.0%, and 3.95%, respectively, compared with the BiGRU-CRF model. The difference between the two models is the embedding of BERT. It shows that BERT embedding can improve the recognition effect of entities. Compared with the BiLSTM-CRF model, the increase was 3.14%, 4.40%, and 4.34%, respectively. Compared with the BERT-BiLSTM-CRF model, the increase was 1.25%, 0.77%, and 1.01%, respectively. Therefore, all  $P$ ,  $R$ , and  $F1$  score are improved compared to the baseline model, indicating that the BERT-BiGRU-CRF model is more applicable to electronic medical record recognition in the CVD field. This is mainly due to the stronger ability of embedding BERT to extract features, which enables word vectors to fuse context information. On the other hand, the BiGRU-CRF model can input bidirectional information before and after the sequence, which can effectively avoid entity ambiguity.

Figure 6, in terms of entity types, horizontally compares the recognition effects of various entities under different models and compare BiLSTM-CRF, BERT-BiLSTM-CRF, and BiGRU-CRF. In terms of disease entities, they were increased by 9.87%, 2.73%, and 9.63%, respectively; on the symptom entity, they were increased by 1.62%,  $-0.33\%$ , and 3.29%; on the body part entity, they were increased by 2.85%, 0.45%, and 3.10%. On the examination entity, they were increased by 0.76%,  $-0.25\%$ , and 0.49%. In terms of the treatment entities, they were increased by 3.8%, 2.46%, and 3.29%, respectively. The overall recognition effect of different entities is compared longitudinally, the effect of checking entity recognition is higher than the comparison model, and the  $F1$  score reaches 90%. However, the recognition effect of entities in the treatment category is relatively poor because the entities are relatively long and cannot clearly identify the boundaries of each entity. In short, the recognition effect of the BERT-BiGRU-CRF model proposed in this paper is higher than that of the control group.

*4.3. Model Training Time.* Model training is the process of parameter update. This article analyzes the relationship between the four models in the first 10 rounds of Epoch and  $F1$ . It can be seen from Figure 7 that the  $F1$  score of the neural network model without BERT is continuously rising from a lower level, while the  $F1$  score of the neural network model with BERT can be maintained at a higher level, and it takes iterations to reach the optimal  $F1$  score, fewer times. In addition, as a whole, the  $F1$  score of the BERT-BiGRU-CRF model is the highest. From the comparison of training time, Table 5 lists the time required for each model iteration. The training time of the BERT-BiGRU-CRF model for one round is 37 seconds shorter than that of the BERT-BiLSTM-CRF model. This is due to the simple structure of the BiGRU-CRF model and the higher efficiency of the model in calculation. In addition, comparing BiGRU-CRF and BERT-BiGRU-CRF models, it is worth noting that the BERT with full word mask is added to the neural network model, which improves the overall training efficiency of the model. The overall training efficiency is improved.

In summary, the BERT-BiGRU-CRF entity recognition model proposed in this paper has a better recognition effect than the control group. This model can make full use of context information, further avoid ambiguity, and effectively avoid repetition between entities, and the granularity of word segmentation in this article is small, which can improve the accuracy of entity recognition.

*4.4. Entity Recognition Result.* This paper uses the BERT-BiGRU-CRF named entity recognition model to identify 9393 entities (without deduplication). Among them, electronic medical records have the most descriptions of body parts, followed by symptoms and examination entities, while treatment and disease types are less. The specific results are shown in Figure 8.

TABLE 4: Comparison results of each model.

Model	Evaluation index (%)	Entity type					Comprehensive value
		Disease	Symptoms	Body parts	Examination	Treatment	
BiLSTM-CRF	<i>P</i>	84.01	89.09	85.79	91.16	83.28	86.67
	<i>R</i>	79.50	89.83	86.24	92.04	85.06	86.53
	<i>F1</i>	81.69	89.46	86.01	91.60	84.16	86.60
BERT-BiLSTM-CRF	<i>P</i>	88.64	90.38	87.82	91.98	84.12	88.59
	<i>R</i>	89.02	92.46	89.01	93.24	87.09	90.16
	<i>F1</i>	88.83	91.41	88.41	92.61	85.58	89.37
BiGRU-CRF	<i>P</i>	83.93	88.20	86.33	91.90	84.29	86.93
	<i>R</i>	80.02	87.38	85.20	91.84	85.22	85.93
	<i>F1</i>	81.93	87.79	85.76	91.87	84.75	86.43
BERT-BiGRU-CRF	<i>P</i>	90.65	90.26	88.64	92.33	87.29	89.83
	<i>R</i>	92.48	91.92	89.08	92.39	88.81	90.94
	<i>F1</i>	91.56	91.08	88.86	92.36	88.04	90.38

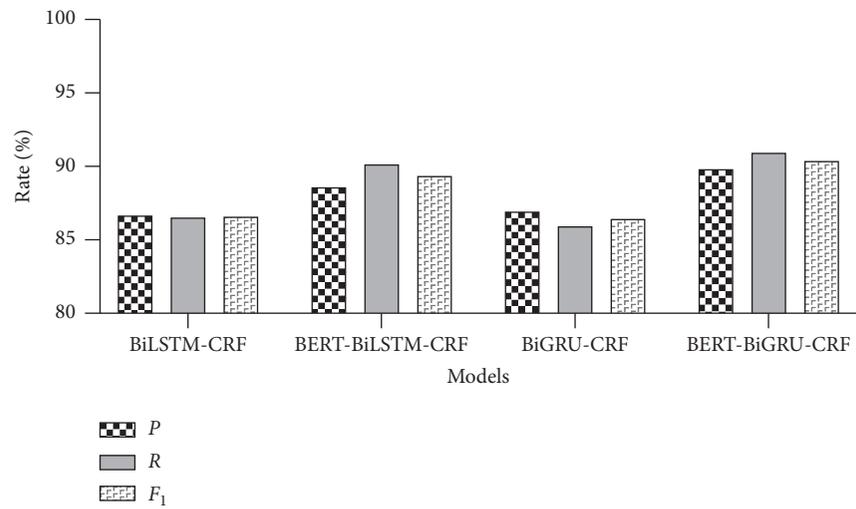


FIGURE 5: Comparison of various indicators of different models.

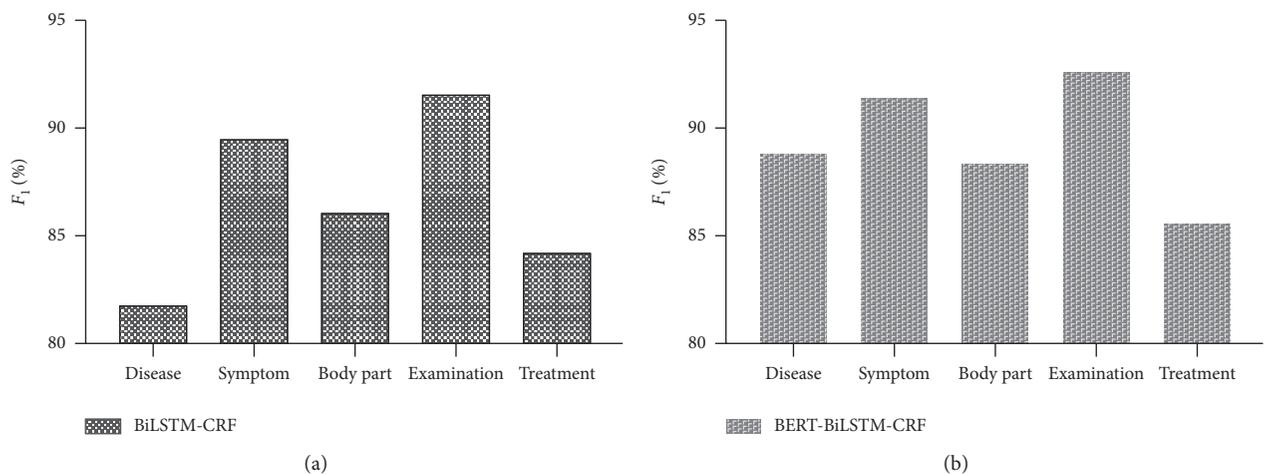


FIGURE 6: Continued.

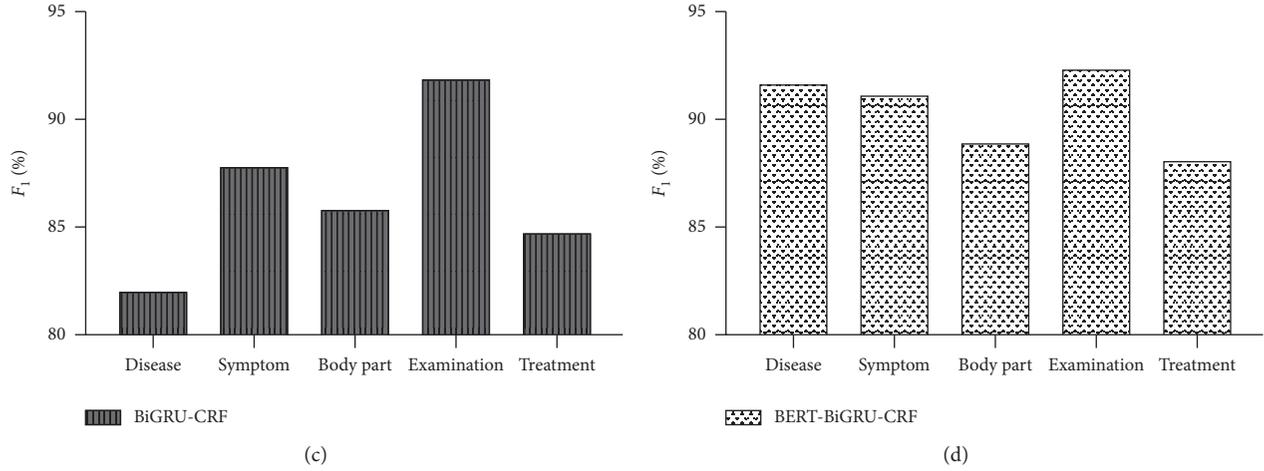
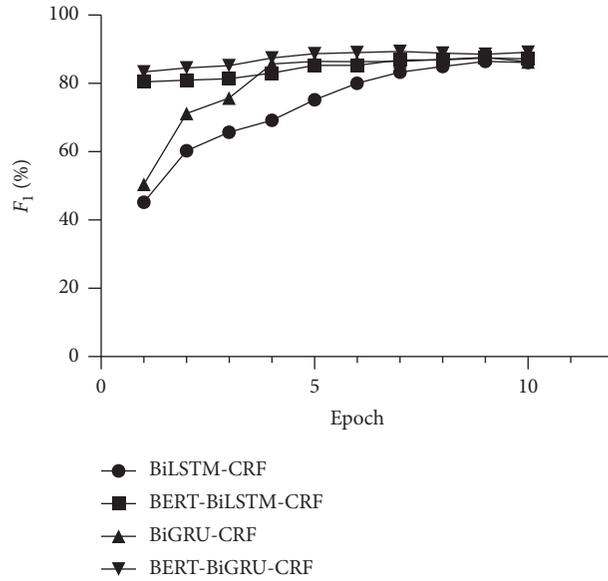
FIGURE 6: Comparison of entity  $F_1$ .FIGURE 7: Epoch and  $F_1$  relationship.

TABLE 5: Time of one round of model training.

Model	Training time (s)
BiLSTM-CRF	472
BERT-BiLSTM-CRF	356
BiGRU-CRF	429
BERT-BiGRU-CRF	319

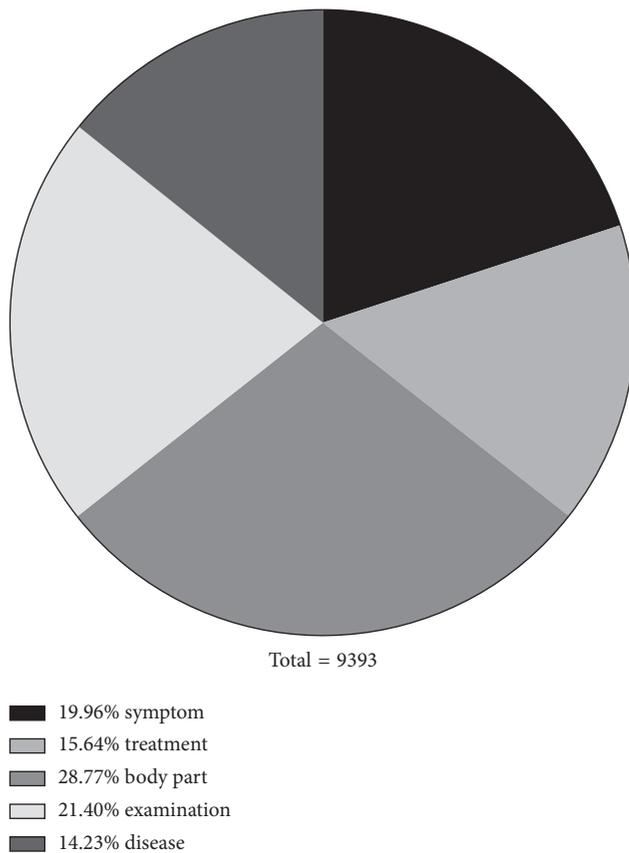


FIGURE 8: Distribution of various entities.

## 5. Conclusions

Aiming at the text data of electronic medical records of cerebrovascular diseases, this paper proposes a BERT-BiGRU-CRF entity recognition model to identify five key entities in the field of cerebrovascular diseases, which are “disease, symptoms, body parts, examination, and treatment.” The model obtains the word vector combined with context information through the BERT layer and then obtains the optimal annotation sequence through the BiGRU-CRF neural network model. It not only guarantees a simple network structure and fast training speed but also can solve the problem of ambiguity in combination with context information. Next, on the one hand, we will study the construction of high-quality dictionaries. On the other hand, we will extract the relationship between different entities based on NER to construct a knowledge map in the field of cerebrovascular diseases, which is conducive to the further potential information of electronic medical records in the field of CVD.

## Data Availability

The Chinese electronic medical record data used to support the findings of this study have been deposited in the Ai’ai medical repository.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the Beijing Municipal Commission of Science and Technology Project (Z131100005613017), the model and demonstration application of the collaborative prevention and treatment of major diseases in Beijing’s medical reform.

## References

- [1] H.-J. Jang and K.-O. Cho, “Applications of deep learning for the analysis of medical data,” *Archives of Pharmacal Research*, vol. 42, no. 6, pp. 492–504, 2019.
- [2] J. R. Ubbens and I. Stavness, “Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks,” *Frontiers in Plant Science*, vol. 8, no. 1190, p. 2245, 2018.
- [3] L. Shen, Q. Li, W. Wang et al., “Treatment patterns and direct medical costs of metastatic colorectal cancer patients: a retrospective study of electronic medical records from Urban China,” *Journal of Medical Economics*, vol. 23, no. 5, pp. 456–463, 2020.
- [4] Z. Li, T. Liu, L. Ding, Z. Liu, X. Li, and Y. Wang, “Research progress of machine learning in the diagnosis and treatment of cerebrovascular diseases,” *Chinese Journal of Stroke*, vol. 15, no. 3, pp. 283–289, 2020.
- [5] S. Wu, B. Wu, and M. Li, “Stroke in China: advances and challenges in epidemiology, prevention, and management,” *Lancet Neurology*, vol. 18, no. 4, pp. 394–405, 2019.
- [6] J. M. Wardlaw, C. Smith, and M. Dichgans, “Small vessel disease: mechanisms and clinical implications,” *The Lancet Neurology*, vol. 18, no. 7, pp. 684–696, 2019.
- [7] W. J. Powers, A. A. Rabinstein, T. Ackerson et al., “Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American heart association/American stroke association,” *Stroke*, vol. 50, no. 12, pp. e344–e418, 2019.
- [8] Z. Li, A. B. Singhal, and Y. Wang, “Stroke physician training in China,” *Stroke*, vol. 48, no. 12, pp. e338–e340, 2017.
- [9] E. Smit, G. Saposnik, G. Biessels et al., “Prevention of stroke in patients with silent cerebrovascular disease: a scientific statement for healthcare professionals from the American heart association/American stroke association,” *Stroke*, vol. 48, no. 2, pp. E44–E71, 2017.
- [10] J. Virtanen, M. Varpela, F. Biancari, J. Jalkanen, and H. Hakovirta, “Association between anatomical distribution of symptomatic peripheral artery disease and cerebrovascular disease,” *Vascular*, vol. 28, no. 3, pp. 295–300, 2020.
- [11] J. Virtanen, M. Varpela, and F. Biancari, “Association between anatomical distribution of symptomatic peripheral artery disease and cerebrovascular disease,” *Vascular*, vol. 28, no. 32, pp. 95–300, 2020.
- [12] A. Baluja, R. Moisés, A. Cordero et al., “Prediction of major adverse cardiac, cerebrovascular events in patients with diabetes after acute coronary syndrome,” *Diabetes & Vascular Disease Research*, vol. 17, no. 1, pp. 67–69, 2019.
- [13] C. Romana, T. Eva, D. Edin, J. Raka, and T. Milan, “Brain image segmentation based on firefly algorithm combined with K-means clustering,” *Studies in Informatics and Control*, vol. 28, no. 2, pp. 167–176, 2019.
- [14] B. Ambale-Venkatesh, X. Yang, C. Wu et al., “Cardiovascular event prediction by machine learning the multi-ethnic study

- of atherosclerosis," *Circulation Research*, vol. 121, no. 9, p. 1092, 2017.
- [15] I. Korvigo, M. Holmatov, A. Zaikovskii, and M. Skoblov, "Putting hands to rest: efficient deep cnn-rnn architecture for chemical named entity recognition with no hand-crafted rules," *Journal of Cheminformatics*, vol. 10, no. 28, 2018.
- [16] H. Afify, K. Mohammed, and A. Hassaniien, "Multi-images recognition of breast cancer histopathological via probabilistic neural network approach," *Journal of System and Management Sciences*, vol. 1, no. 2, pp. 53–68, 2020.
- [17] A. Ekbal and S. Saha, "Simultaneous feature and parameter selection using multi-objective optimization: application to named entity recognition," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 4, pp. 597–611, 2020.
- [18] Y. Meng and L. Long, "A deep learning Approach for a source code detection model using self-attention," *Complexity*, vol. 2020, Article ID 5027198, 15 pages, 2020.
- [19] L. Gligic, A. Kormilitzin, P. Goldberg, and A. Nevado-Holgado, "Named entity recognition in electronic health records using transfer learning bootstrapped neural networks," *Neural Networks*, vol. 121, pp. 132–139, 2020.
- [20] C. Dumitrescu and I. Dumitrache, "Combining deep learning technologies with multi-level gabor features for facial recognition in biometric automated systems," *Studies in Informatics and Control*, vol. 28, no. 2, pp. 221–230, 2019.
- [21] J. Chen, Z. Chang, and J. Xu, "Recognition and classification of biomedical named entities based on HMM," *Computer Age*, no. 10, pp. 40–42, 2006.
- [22] M. Sui and L. Cui, "CRF model combining multiple characteristics for chemical substance-disease named entity recognition," *Modern Library and Information Technology*, no. 10, pp. 91–97, 2016.
- [23] X. Sun, C. Sun, and F. Ren, "Biomedical named entity recognition based on deep conditional random fields," *Pattern Recognition and Artificial Intelligence*, vol. 29, no. 11, pp. 997–1008, 2016.
- [24] A. Cocos, A. Fiks, and A. Masino, "Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts," *Journal of the American Medical Informatics Association*, vol. 24, no. 4, pp. 813–821, 2017.
- [25] F. Zhang and M. Wang, "Medical named entity recognition based on deep learning," *Computing Technology and Automation*, vol. 36, no. 1, pp. 123–127, 2017.
- [26] L. Qin, N. Yu, and D. Zhao, "Applying the convolutional neural network deep learning technology to behavioural recognition in intelligent video," *Tehnicki Vjesnik-Technical Gazette*, vol. 25, no. 2, pp. 528–535, 2018.
- [27] U. Zuperl and F. Cus, "A cyber-physical system for smart fixture monitoring via clamping simulation," *International Journal of Simulation Modelling*, vol. 18, no. 1, pp. 112–124, 2019.
- [28] J. M. Giorgi and G. D. Bader, "Towards reliable named entity recognition in the biomedical domain," *Bioinformatics*, vol. 36, no. 1, pp. 280–286, 2020.
- [29] I. Unanue, E. Borzeshi, and M. Piccardi, "Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition," *Journal of Biomedical Informatics*, vol. 76, pp. 102–109, 2017.
- [30] L. Weston, V. Tshitoyan, J. Dagdelen et al., "Named entity recognition and normalization applied to large-scale information extraction from the materials science literature," *Journal of Chemical Information and Modeling*, vol. 59, no. 9, pp. 3692–3702, 2019.
- [31] Z. H. Kilimci and S. Akyokus, "Deep learning- and word embedding-based heterogeneous classifier ensembles for text classification," *Complexity*, vol. 2018, Article ID 7130146, 10 pages, 2018.
- [32] S. Chowdhury, X. Dong, and L. Qian, "A multitask bi-directional RNN model for named entity recognition on Chinese electronic medical records," *BMC Bioinformatics*, vol. 19, no. 17, pp. 75–84, 2018.
- [33] M. Al-Smadi, S. Al-Zboon, Y. Jararweh, and P. Juola, "Transfer learning for Arabic named entity recognition with deep neural networks," *Ieee Access*, vol. 8, no. 31, pp. 37–45, 2020.
- [34] M. Gridach and H. Haddad, "Arabic named entity recognition: a bidirectional GRU-CRF approach," in *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science*, vol. 10761, pp. 264–275, , no. 1, Springer, Berlin, Germany, 2018.
- [35] Y. Kim and T. Lee, "Korean clinical entity recognition from diagnosis text using bert," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 242, 2020.
- [36] J. Lee, W. Yoon, S. Kim et al., "Biobert: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics (Oxford, England)*, vol. 36, no. 4, pp. 1234–1240, 2020.
- [37] P. Yang and W. Dong, "Chinese named entity recognition method based on BERT embedding," *Computer Engineering*, vol. 46, no. 4, pp. 40–45+52, 2020.
- [38] Y. Cho and Y. Lee, "Biomedical named entity recognition using deep neural networks with contextual information," *BMC Bioinformatics*, vol. 20, no. 1, p. 735, 2019.
- [39] X. Chen, C. Ouyang, Y. Liu, and Y. Bu, "Improving the named entity recognition of Chinese electronic medical records by combining domain dictionary and rules," *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, p. 2687, 2020.
- [40] F. Li, Y. Jin, W. Liu, B. Rawat, P. Cai, and H. Yu, "Fine-tuning bidirectional encoder representations from transformers (Bert)-based models on large-scale electronic health record notes: an empirical study," *JMIR Medical Informatics*, vol. 7, no. 3, Article ID e14830, 2019.
- [41] Y. Feng, H. Yu, and G. Sun, "Named entity recognition method based on BLSTM[J]," *Computer Science*, vol. 45, no. 2, pp. 261–268, 2018.
- [42] J. Wu, Y. Cheng, and H. Hao, "Research on Chinese professional term extraction based on BERT embedded BiLSTM-CRF model," *Journal of Information*, vol. 39, no. 4, pp. 409–418, 2020.