

Research Article

A Hybrid Model Based on Deep Convolutional Network for Medical Named Entity Recognition

Tingzhong Wang ¹, Yongxin Zhang,¹ Yifan Zhang,¹ Hao Lu,¹ Bo Yu,¹ Shoubo Peng,² Youzhong Ma ¹ and Deguang Li ¹

¹School of Information Technology, Luoyang Normal University, Luoyang 471934, China

²Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

Correspondence should be addressed to Tingzhong Wang; wangtingzhong@lynu.edu.cn

Received 8 February 2023; Revised 23 April 2023; Accepted 10 May 2023; Published 23 May 2023

Academic Editor: Ping-Feng Pai

Copyright © 2023 Tingzhong Wang 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.

The typical pretrained model's feature extraction capabilities are insufficient for medical named entity identification, and it is challenging to express word polysemy, resulting in a low recognition accuracy for electronic medical records. In order to solve this problem, this paper proposes a new model that combines the BERT pretraining model and the BiLSTM-CRF model. First, word embedding with semantic information is obtained by pretraining the corpus input to the BERT model. Then, the BiLSTM module is utilized to extract further features from the encoded outputs of BERT in order to account for context information and improve the accuracy of semantic coding. Then, CRF is used to modify the results of BiLSTM to screen out the annotation sequence with the largest score. Finally, extensive experimental results show that the performance of the proposed model is effectively improved compared with other models.

1. Introduction

Medical named entity recognition is the foundational technology of natural language processing used in the medical field, and there are two primary implementation approaches: rule-based approach and model-based approach depending on the requirement of extracting named entities. Rule-based methods primarily employ regular expressions and dictionaries, which are simple to update and provide high certainty, while the disadvantages of the methods are also obvious, as regular expressions need to write a large number of rules, and dictionaries need to collect a large number of synonyms; thus, regular expressions usually are used to express well-defined entities, and dictionaries are used to express specialized words. The classic model-based methods are the combination of BiLSTM and CRF, which have the advantage of strong generalization ability but the disadvantage of requiring a large number of labeled samples. However, with the increasing demand for medical treatment and the large population of our country,

a sizable quantity of electronic medical records was created. And deep learning is rapidly advancing in bioinformatics [1–3], disease recognition, medical image analysis, clinical decision-making, and medication discovery at the same time, which has been used to create a variety of models for data feature learning, information mining, state simulation and recognition, evaluation, prediction, and other functions through multidimensional quantification of recorded information such as text, image, and field information, and the practice of deep learning in the biomedical field has achieved better results than traditional algorithms. Thus, the application of deep learning in automatic disease coding, multidata source integration analysis, public health, and other aspects is worthy of further exploration.

As an emerging discipline and rapidly developing branch in the field of machine learning, deep learning has been widely used in many fields, such as image recognition automatic speech recognition, automatic machine translation, natural language processing, and so on. To learn the internal characteristics of data, deep learning employs an artificial

neural network (ANN) composed of neurons with multiple layers of nonlinear processing. In fact, the concept of neural networks was proposed as early as the 1940s, but the modern ANN did not emerge until the 1980s. However, due to over fitting and gradient problem, the ANN is replaced by machine learning algorithms such as support vector machine and random forest. The recent advancement of deep learning has repopularized the ANN; the main reason is that as computing hardware improves (the appearance of high-performance CPU and GPU), the network framework can have more hidden layers. Furthermore, the deep learning algorithm has been improved in recent years, including the Dropout method which is used to solve overfitting, the activation function ReLU to solve the gradient attenuation problem, and the convolutional and pooling process into the network framework.

Deep learning methods based on neural networks have made significant progress in the field of natural language in recent years; natural language processing (NLP) focuses on the interaction between computer and human language, especially text mining and classification are widely used in application scenarios such as topic detection, document classification, and scene understanding. More and more researchers are interested in using neural network models to handle classification, recognition, and prediction problems, in which models learn and extract higher-level features from raw data, and deep learning has also produced promising results in the research of named entity recognition (NER), the key basic task of NLP. With the rapid expansion of biomedical texts, a large amount of biomedical knowledge exists primarily in the form of unstructured free-text in various forms, and the information extraction has become a critical basis for biomedical research; thus, the use of named entity recognition in the biomedical field has a significant application value, while traditional methods could not achieve efficient recognition performance due to the large size of biomedical data and problems of out-of-vocabulary. As a result, researchers began to apply the named entity method based on deep learning to the field of biomedicine. For natural language documents in the medical field, such as medicine, clinical cases, hospital records, and inspection reports, these texts contain a large number of medical professional knowledge and medical terminology, combining entity recognition technology with medical professional field and using machine to read medical texts could significantly improve the efficiency and quality of clinical research and serve downstream subtasks.

Unstructured free-texts in medicine are typically made up of natural language sentences or sentence sets, and entity extraction is the process of extracting medical entities from these unstructured free-texts, such as diseases and symptoms. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are currently the most popular deep learning models in the field of named entity recognition. In this field, the RNN could process sequence data efficiently and plays an important role in the development of NLP. The methods based on deep learning can reduce the dependence of feature engineering [4, 5]. The combination of LSTM and CRF eliminate the need for most feature

engineering, which surpasses the previous traditional methods and reduces the out-of-vocabulary problem at the same time. A typical NER model, such as the RNN-CRF model, is mainly composed of the embedding layer (mainly including word vector, character vector, and some additional features), bidirectional RNN layer, TANH hidden layer, and the final CRF layer, and experimental results show that RNN-CRF has achieved good results, which has reached and even exceeded the CRF model with rich features and has become the most mainstream model in the current learning-based NER method. In terms of feature extraction, the RNN-CRF model inherits the advantages of deep learning methods, which could achieve good results by using word vectors and character vectors without using feature engineering. Furthermore, the model could be improved further if there are high-quality dictionary features.

At present, ANN-CRF, CNN-CRF, or RNN-CRF models that combine the neural network and CRF has become the mainstream model of NER. The CNN and RNN have their respective advantages, while RNN-CRF is more widely used because of the natural sequence structure of the RNN. The NER method based on the neural network structure inherits the advantages of deep learning methods, which could reach the level of current mainstream technology without a large number of artificial features which only need word vector and character vector, and adding high-quality dictionary features could further improve the effect. With the development of computing power and the proposal of word embedding, the neural network could effectively deal with many NLP tasks. The processing of this kind of method for sequence annotation tasks (such as CWS, POS, and NER) is similar: map the token from discrete one-hot representation to low dimensional space to become dense embedding, then input the embedding sequence of the sentence into the RNN, and use the neural network to automatically extract features, and softmax function to predict the label of each token. In this paper, we adopt the BERT model as our backbone network to strengthen the understanding of features, and then enhance the accuracy of prediction results through the feature extraction of BiLSTM and the decoding ability of CRF. Thus, a hybrid framework based on BERT-BiLSTM-CRF with good entity recognition function is proposed, and extensive experimental results show that the performance of the proposed model is effectively improved compared with other models.

1.1. Related Work. The main purpose of NER is to identify and classify specific entity from medical records, such as side symptoms, drugs, and treatment; this task is usually solved as a sequence label problem, in which entity boundary and category label are jointly predicted [6, 7]. Researchers have explored a variety of methods to excavate medical records because of the enormous amount of information in medical texts. In recent years, machine learning methods and deep learning methods have been widely used to solve NER problems, and a major difference between the two methods is that machine learning methods rely on artificial feature engineering, while deep learning methods are end-to-end,

both methods adopt a two-stage approach to solve natural language processing (NLP) tasks. The first stage is to train the language model on a large corpus, and the second stage is to apply the pretrained language model to downstream tasks. Because of their success in various NLP tasks, leveraging unsupervised pretraining becomes very useful, especially when task-specific annotations are difficult to express [8]. Zhu et al. [9] proposed an online medical prediagnosis framework with high efficiency and privacy protection by using nonlinear kernel SVM, and an efficient privacy protection classification scheme is proposed by using lightweight multiparty random masking and polynomial aggregation technology.

Wang et al. [10] focused on the memory mechanism, proposed a gradient-based learning algorithm to deal with the gradual reduction of classification tasks as the gradient disappears, and combined the RNN model with the auxiliary model to overcome memory conflicts. Many named entity recognition models are extended based on statistical methods, including the method of hidden Markov models (HMM), and NER has also made some major breakthroughs in its development process. Among them, the word vector proposed by Bengio [11] and the new method of combining convolutional neural networks (CNNs) and CRF proposed by Collobert [12] were applied to NER research, which obtained a better entity recognition results. He and Wang [13] compared character-level based and word-level based in word segmentation, and found that the recognition effect of character based is better than word based for NER. Collobert [12] proposed the combination of the sentence method and window method for named entity recognition, which is one of the earliest representative works using neural networks for NER. Xu et al. [14] combined word segmentation and named entity recognition for training in order to fuse word segmentation information. Dos Santos and Zadrozny [15] proposed a CharWNN architecture to process characters by using the CNN on the basis of POS and achieved good results. Huang et al. [16] employed BLSTM and CRF to the benchmark annotation dataset in NLP, which reduced the dependence on word embedding and improved the accuracy rate.

Lample [17] and Ma and Hovy [18] proposed an end-to-end method combining BiLSTM, CNN, and CRF to solve NER problems. Yang et al. [19] proposed a training model based on word vector in 2016; the model used the GRU-CRF (gated recurrent unit conditional random field) method and achieved good results in multiple text annotation tasks. Chiu and Nichols [20] proposed a hybrid bidirectional LSTM and CNN neural network model, used CNN training to obtain character-level vectors and combined with word vectors, and then used the combination of the two as the input of the BLSTM deep neural network model, which achieved good results on CoNLL-2003 and Onto Notes5.0 with $F1$ values of 91.62% and 86.28%. Long et al. [21] made good use of the CNN to get the presentation character of words and label the words by using BiLSTM and CRF to improve the performance of named entity recognition on SIGHAN 2006 Bakeoff-3. Li et al. [22] used the bidirectional long short-term memory (LSTM) neural network model based on

the conditional random field in the study of biological texts to identify irregular entities. Chen et al. [23] proposed an annotation system for medical named entity recognition based on active learning, which extract the entity of treatments, problems, and laboratory-related experiments from medical records; the annotation system iteratively builds a named entity recognition model based on the annotated sentences and selects the next sentence for annotation. Wang et al. [24] combined the advantages of both BiLSTM and CRF to present a weakly supervised BiLSTM-CRF method to solve the problem of clinical performance audits in emergency medical services (EMS). The experimental results show that this approach may further improve the efficiency of clinical audits.

NER is usually regarded as a sequence labeling problem, and usually a large-scale corpus is used to train the annotation model, so as to label each position of the sentence, commonly used models in NER tasks include the generative model HMM and discriminative model CRF. Now, CRF is a mainstream model of NER problems, its objective function not only considers the input state feature function, but also includes the label transition feature function, and SGD could be used to learn model parameters during training. At the same time, when the model is known, the optimal sequence that maximizes the objective function is obtained from the input sequence, which is a dynamic programming problem and could be decoded by the Viterbi algorithm to get the optimal label sequence. The advantage of CRF is that it can take advantage of wealth of internal and contextual features in the process of annotating a location. With the continuous progress of technology, the research on named entities has changed from traditional machine learning to deep learning, in which the deep learning framework models are mainly based on the convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory network (LSTM), and so on. Zeng [25] proposed a piecewise convolutional neural network (PCNN) based on the CNN. Vaswani [26] used the attention mechanism to better capture contextual information. Devlin [27] proposed the bidirectional encoder pretraining model from transformers (BERT) based on the previous algorithm experience, and the model has achieved good results in Chinese data source scenarios and multiple NLP fields due to its' high flexibility. Jiang et al. [28] combined BERT with LSTM-CRF and achieved better results than other models. With the continuous development of named entities, the drug relocation technology combining named entities and medicine has emerged. Luo et al. [29, 30] proposed a drug relocation algorithm based on comprehensive similarity and random walk; they first proposed a comprehensive similarity calculation method for drug and disease similarity by combining the drug and disease feature information with the known drug disease relationship. Chiang and Butte [31] proposed a view of drug repositioning from the perspective of the disease, which considers two diseases to be similar when they can be treated by multiple of the same drugs. If a drug has a therapeutic effect on only one of the diseases, it is considered that the drug also has a potential therapeutic

relationship for the other disease and could be used as a candidate drug for the treatment of the disease. Zhang et al. [32] proposed a series of studies on drug relocation algorithms based on collaborative filtering; the predicted value of the relationship between drugs and diseases could be calculated by collaborative filtering from the perspective of multisource data.

This paper investigates the problem of drug and medical record identification primarily through text research of medical drug information, which is then combined with other fields of technology analysis in order to extract drug information. Yan [33] proposed the transformer encoder for the NER (TENER) model and designed the attention mechanism with orientation and relative position information to improve the problem that a transformer could not capture orientation information and relative position information; the $F1$ value of this model reaches 92.74% on MSRA Chinese corpus and 88.43% on Onto Notes5.0. Khan et al. [34] proposed a multitask transformer model for biomedical named entity recognition, the problem of training a slot tagger with multiple datasets containing different slot types was regarded as a multitask learning, and the context information of input representation was captured by the encoder of the transformer model, then generate the shared context embedding vector. Finally, a specific task representation for each task/dataset was created, which increases the efficiency and effectiveness in terms of time and memory. Huang et al. [35] proposed a novel two-stage method, which first uses a feature-based binary classifier to identify positive instances, and then uses a classifier based on long-term and short-term memory (LSTM) to classify positive instances into specific categories; the $F1$ -score value of the experimental result is 69.0%. Qin and Zeng [36] proposed a deep learning model combining the BiLSTM network and CRF to extract entities in electronic medical records; the highest $F1$ -score value in i2b2 and VA public data set has reached 85.37%. Akbik et al. [37] dynamically constructed the “memory” of context embedding to store the word embedding generated by each word, and applied a pool operation to extract the global representation of each word. In this way, the embedding of words is related not only to the current sentence, but also to the previous text in the document. The method effectively solves the problem of embedding rare characters in unspecified context, and t reached the highest $F1$ -score value of 93.18% and German reached 88.27% on conll2003 English data set.

In summary, although the traditional named entity model algorithm has yielded some results, there is still much effort for improvement. In view of the insufficient feature extraction ability of the traditional pretraining model and the difficulty in expressing the polysemy of words in the traditional medical named entity recognition, this paper proposes a model for medical record named entity recognition, which combines the BERT pretraining model and the BiLSTM-CRF model. Firstly, the sentences to be recognized are input into the BERT model for pretraining to obtain the word embedding with semantic information, and then it is

used as the input layer of BiLSTM module for semantic coding. The BERT model could significantly reduce the amount of training and increase training speed and accuracy by using the Chinese training model files that Google has made available. In addition, word vector training in Chinese is easier than word vector training in English, the semantic coding is more accurate than the traditional word vector model, and it could generate dynamic word vector in accordance with the context. Finally, the BiLSTM model's output is modified by the addition of CRF, which ultimately produces the annotation sequence with the highest score. Experimental results show that the proposed model effectively increases the precision of biomedical named entity recognition.

1.2. Proposed Model. This paper presents a hybrid model that consists of three modules: BERT, BiLSTM, and CRF. Generally, the model first carries out pretraining through BERT to obtain word embedding with semantic information, then takes it as the input layer of BiLSTM module for semantic coding, and finally fed the results to the CRF layer to calculate the optimal tag sequence. The overall structure is shown in Figure 1.

BERT is the framework to train deep bidirectional transformers for language understanding from Google [38], its acronym is the “bidirectional encoder representations from transformers.” It is a self-coding language model (autoencoder LM) as a whole, and designs two tasks to pretrain the model. The first task is to use the mask information to train the language model. Generally speaking, when inputting a sentence, randomly select some words to be predicted, and then replace them with a special symbol, and let the model learn these words according to the given label. The second task adds a sentence level continuity prediction task on the basis of the two-way language model, that is, to predict whether the two paragraphs of text input into BERT are continuous text. The introduction of this task can better enable the model to learn the relationship between continuous text segments. Compared with the original RNN and LSTM, BERT can execute concurrently, extract the relational features of words in sentences at the same time, and extract the relational features at multiple different levels, so as to reflect the sentence semantics more comprehensively. Compared with Word2vec, it can obtain the word meaning according to the sentence context, so as to avoid ambiguity. At the same time, the disadvantages are also obvious. There are too many model parameters and the model are too large. It is easy to overfit when training with a small amount of data. The main input of the BERT model is the original word vector of each word (or token) in the text. This vector can be initialized randomly or pretrained by algorithms such as Word2vec as the initial value; the output is the vector representation of each word in the text after integrating the full-text semantic information. In BERT, the pretraining model provided by Google has mastered a lot of natural language semantic and grammatical knowledge through unsupervised

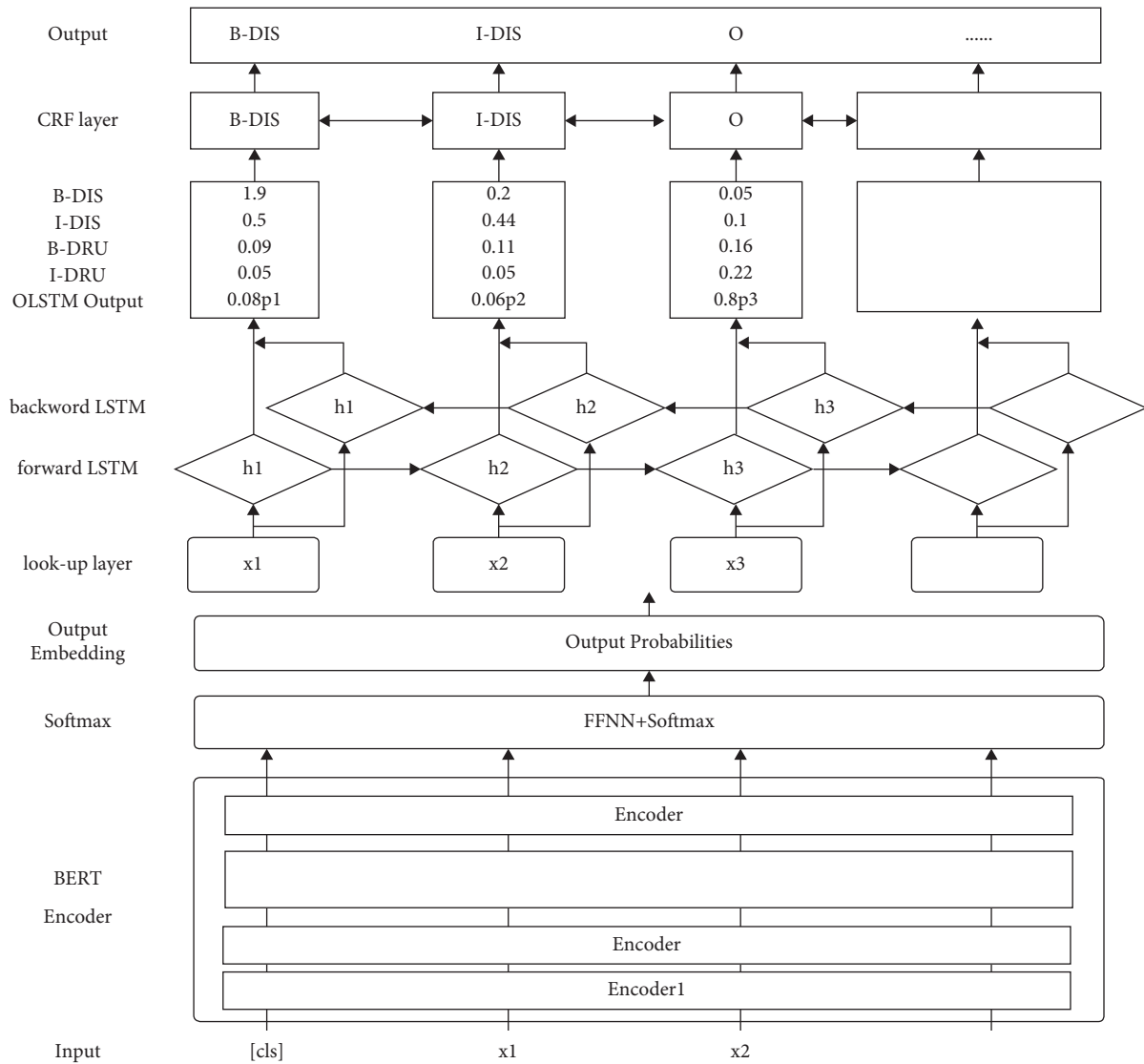


FIGURE 1: Overview architecture of the proposed hybrid model.

learning, and the effect is better when doing downstream natural language processing tasks. The BERT model uses the encoder with 12 layers or 24 layers bidirectional transformer as the feature extractor, its detailed structure is shown in Figure 2.

In each encoder, there is the superposition of multiple BERT layers, and multilayer BERT layers constitute the BERT encoder of a transformer, its detailed structure is shown in Figure 3.

Self-attention is the most important basic structural unit in the whole transformer framework. The whole calculation process of a self-attention revolves around a formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (1)$$

K , Q , and V vectors are obtained by multiplying token embedding and three different weight matrices during training. Figure 4 shows the whole self-attention calculation process with a sentence of thinking machines.

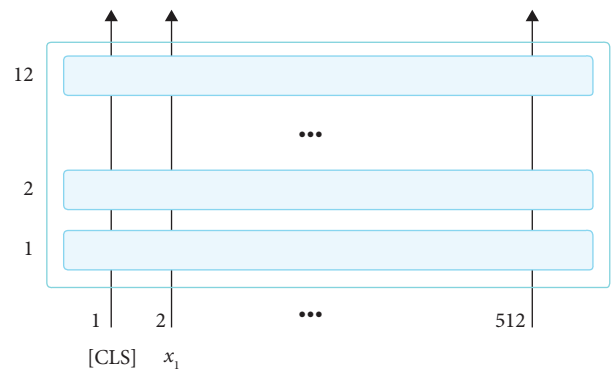


FIGURE 2: Internal structure of the BERT encoder framework.

The input of the BERT model, whether a single sentence or multiple sentences, will convert the token in the sentence into embedding and then transfer it to the model. It is composed of the addition of token embedding, segment

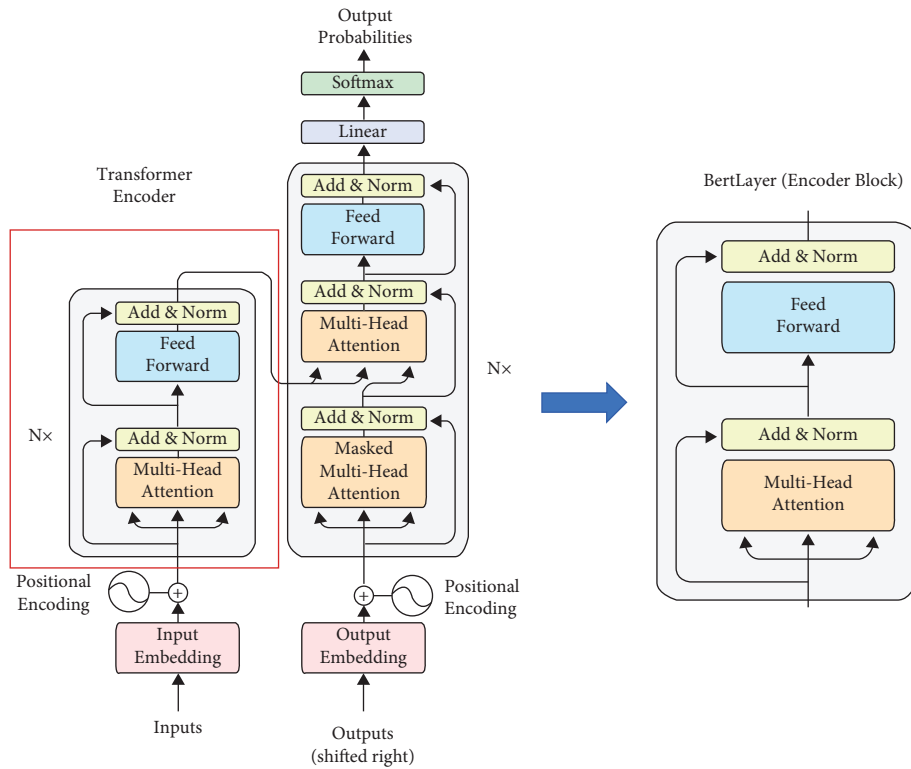


FIGURE 3: Internal structure of a transformer.

embedding, and position embedding, as shown in Figure 5. Among them,

- (1) Token embedding is a word vector that embeds tokens into a dimensional space. BERT selects 768 and 1,024 dimensions, respectively, with the change of the number of structural layers. In addition, BERT also makes special processing on the input of token. The first token is a special character of CLS, which can be used for downstream tasks. There will be a special symbol SEP in the middle of the sentence and at the end of the sentence.
- (2) Segment embedding is used to distinguish two sentences because pretraining not only does LM, but also does the classification task with two sentences as input.
- (3) Unlike the transformer in the previous chapter, position embedding is not a trigonometric function, but a vector learned from training records. To illustrate the problem more clearly, Figure 5 displays the abovementioned process.

In the traditional LSTM model, the model can obtain the previous information, but cannot obtain the later information, and the context information has very important reference significance in the NER task. Especially in the drug name recognition in this paper, the drug name may contain the names or abbreviations of other drugs or diseases. Therefore, we must refer to the following information: LSTM (long short-term memory) is a variant of the recurrent neural network (RNN), it solves the problem that the gradient of the RNN disappears and cannot learn the

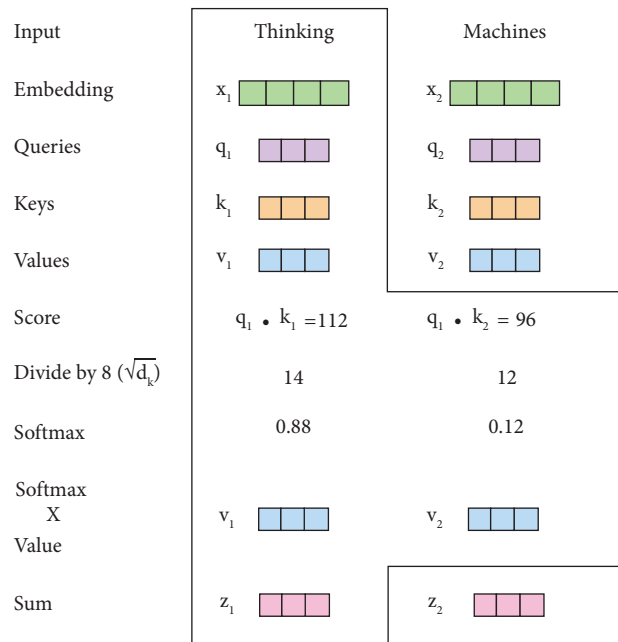


FIGURE 4: Detailed process of converting word vectors from records.

characteristics of long-term dependence through cell gate state and gate control in the hidden layer. Its structure is shown in Figure 6.

The core of the LSTM mainly includes the following structures: forgetting gate, input gate, output gate, and memory cell. The joint function of input gate and forgetting

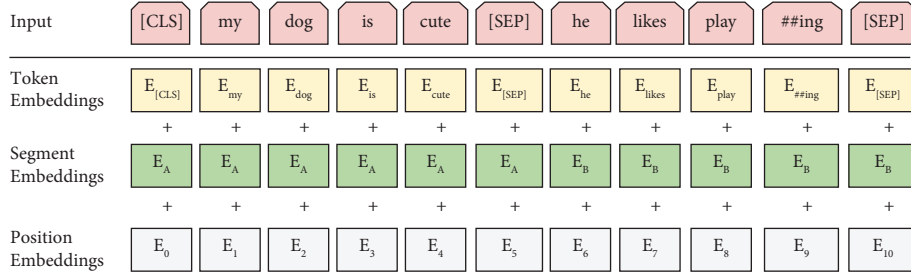


FIGURE 5: Actual operational process of a sentence.

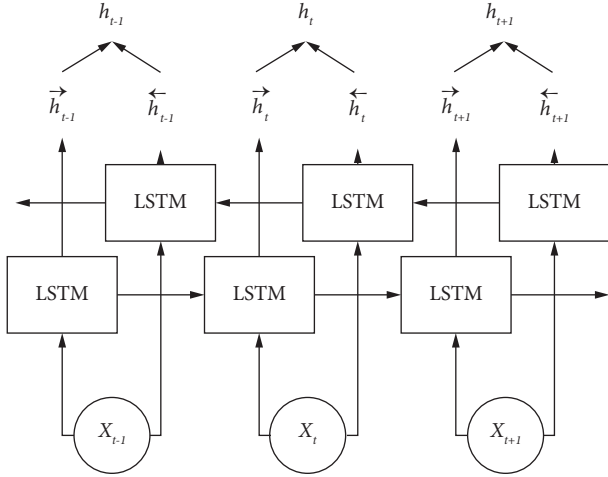


FIGURE 6: Internal structure of the LSTM module.

gate is to abandon useless information and transfer useful information to the next moment. For the output of the whole structure, it is mainly obtained by multiplying the output of memory cell and the output of output gate. Its structure is expressed as follows:

$$\begin{aligned}
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\
 \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \\
 C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \\
 O_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\
 h_t &= O_t \cdot \tanh(C_t).
 \end{aligned} \tag{2}$$

Here, σ is the activation function; W is the weight matrix; b is the offset vector; C_t is the current memory status; i_t , f_t , and O_t mean the input gate, forgetting gate, and output gate, respectively; and h_t is the final output. A bidirectional LSTM is constructed to obtain context information by using the point that LSTM can obtain the abovementioned information. Figure 7 shows its internal structure.

BiLSTM is the recurrent neural network that is composed of forward LSTM and backward LSTM. The forward LSTM can obtain the information of the current text vector and the front LSTM hidden layer, and the backward LSTM can obtain the information of the current text vector and the

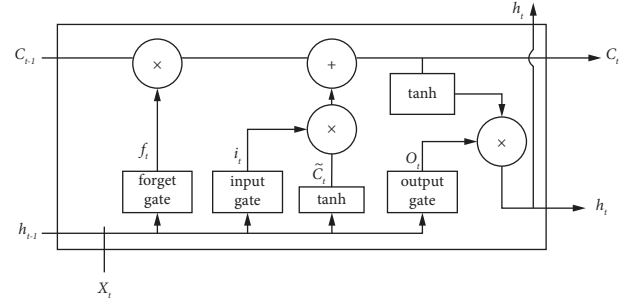


FIGURE 7: Internal structure of BiLSTM.

rear LSTM hidden layer. Take a sentence as a unit, then a record may consist of n words:

$$X = (x_1, x_2, \dots, x_n). \tag{3}$$

The forward and backward LSTM hidden layer at each time is output to the same output unit. Finally, the sum of the word vector outputs from the forward LSTM and backward LSTM, respectively, mean the vector h_t and vector h_i ; they are directly combined into the final result $H = (h_t \text{ and } h_i)$ with the following format:

$$H = (h_1, h_2, \dots, h_n) \in R^n. \tag{4}$$

To obtain the automatically extracted features, the linear layer is used after dropout to map the hidden state vector to K dimension (K is the number of labels in the annotation dataset). In this paper, $K = 5$ (B-DIS, I-DIS, B-DRU, I-DRU, and O). In addition, for the record with n words, each word has a prediction value. Therefore, the final predicted result P is with the following format:

$$P = (p_1, p_2, \dots, p_n) \in R^{n \times k}. \tag{5}$$

In the RNN based BiLSTM, the process of labeling each token is an independent classification, and the tags predicted above cannot be directly used (the abovementioned information can only be transmitted by the hidden state), resulting in the predicted tag sequence may be illegal. For example, the tag B-DIS cannot be followed by I-DIS, but softmax will not use this information.

To make sure the prediction results can cover each label, the CRF layer is used to map each group elements into a one-shot vector. For a record work x , its scores can be calculated by the following formula:

$$\text{score}(x, y) = \sum_{i=1}^n P_{i,y_i} + \sum_{i=1}^n A_{y_{i-1},y_i}. \quad (6)$$

Here, y is the label sequence with shape $Y = (y_1, y_2, \dots, y_n)$ and A is the matrix with shape $(K+2) \times (K+2)$. A_{ij} represents the transfer score from the i th label to the j th label. It can be seen that the score of the whole sequence is equal to the score sum of each position, which is determined by the transfer matrix A of CRF output and predicted result P by LSTM, and then the normalized probability can be obtained by softmax:

$$P(y|x) = \frac{\exp(\text{score}(x, y))}{\sum_{y'} \exp(\text{score}(x, y'))}. \quad (7)$$

At last, we calculate the final prediction by the following expression:

$$y^* = \arg \max_{y'} \text{score}(x, y'). \quad (8)$$

2. Experiments

2.1. Dataset. In this paper, we use BIO annotation method to construct a library with rich semantics. The detailed information about the dataset is shown in Table 1, where L means the number of record words, SN means the number of corresponding sentences, and AL indicates the average length for one sentence.

There are five markers to be predicted, including “B-DIS,” “I-DIS,” “B-DRU,” “I-DRU,” and “O,” in which B is the beginning of the named entity, I is the middle part of the named entity, and O is the component of the unnamed entity. In addition, the evaluation metrics precision (P), recall (R), and $F1$ -score ($F1$) are used in this paper. Table 2 shows their corresponding formulas.

2.2. Implementation Details. The operating systems of the experimental computer platform are Windows 10 and Ubuntu 16.04, the processor is Intel Core Xeon W-2200, the graphics card is NVIDIA Geforce RTX 3080Ti, and the memory is 64 GB. The implementation language is Python 3.6, and the deep learning framework is PyTorch 0.4.0 framework. The optimizer selects the random gradient descent method, the momentum is 0.9, the learning rate is set to 0.00001, the attenuation rate is 0.001, the number of batch size is 5, and the optimizer iterates 100 times. More information is shown in Tables 3 and 4.

3. Results and Analysis

In order to verify the effectiveness of the proposed model, we compare its performance with those models including BERT, BiLSTM-CRF, BERT-CRF, and BERT-BiLSTM-CRF. Table 5 shows the corresponding results, which clearly means only when both precision and recall are high, $F1$ -score can reflect the performance of a model. The precision value of BERT-CRF is the highest, but its recall is lower than

TABLE 1: The information of the constructed dataset.

Classification	L	SN	AL
Train	565,636	8,551	66.1
Validation	55,239	822	67.2
Test	108,701	1,629	66.7

TABLE 2: Evaluation metric and its formula.

Evaluation metric	Description
Precision	$TP/TP + FP$
Recall	$TP/TP + FN$
$F1$ -score ($F1$)	$2 \times TP/2 \times TP + FP + FN$

TABLE 3: Detailed information of computer platform.

Device name	Resource allocation
Operating system	Windows 10
CPU	Xeon W-2200
GPU	RTX 3080 Ti 12 G
Memory size	32 G
Python	3.6
Tensorflow component	1.14

TABLE 4: Detailed parameters of the trained model.

Parameter variables	Values
Transformer layers	12
Hidden layers	768
Batch_size	8
Clip	0.5
Dropout_rate	0.5
Learning_rate	1e-5
Max_seq_length	202
Train_epochs	10
Warmup_proportion	0.1
lstm_size	128

our model. Therefore, the proposed model achieves the best performance among them.

According to the abovementioned test results (in Table 5, Figures 8(a) and 8(b)), we can find that the effect of using the BERT model or BiLSTM-CRF model is not as good as that of the BERT-BiLSTM-CRF model proposed in this paper. This is because in BiLSTM-CRF, although this model can use context information for training and context label information for constraints, according to this model, only corpus can be used, and the results of BERT pretraining are not used, so it needs to be retrained. Therefore, the prediction results are not obvious in small corpus. The BERT model is used alone because the pre-training model is introduced, which can better represent the polysemy of words, but it lacks the ability to fuse context information, so it can only carry out representation migration, resulting in unsatisfactory results. The BERT-BiLSTM-CRF used in this paper first migrates the representation of the pretraining model to the medical field through BERT. After BiLSTM is trained, it effectively integrates the context information and defines the

TABLE 5: The performance of all models tested on the constructed dataset.

Model	Precision (%)	Recall (%)	F1-score (%)	FLOPs	Time (ms)
BiLSTM-CRF	96.00	96.10	96.05	65.45G	3.27
BERT	96.32	95.81	96.06	528.93G	52.73
BERT-CRF	96.75	96.12	96.43	541.21G	53.26
BERT-BiLSTM-CRF	96.60	96.33	96.47	561.62G	54.89

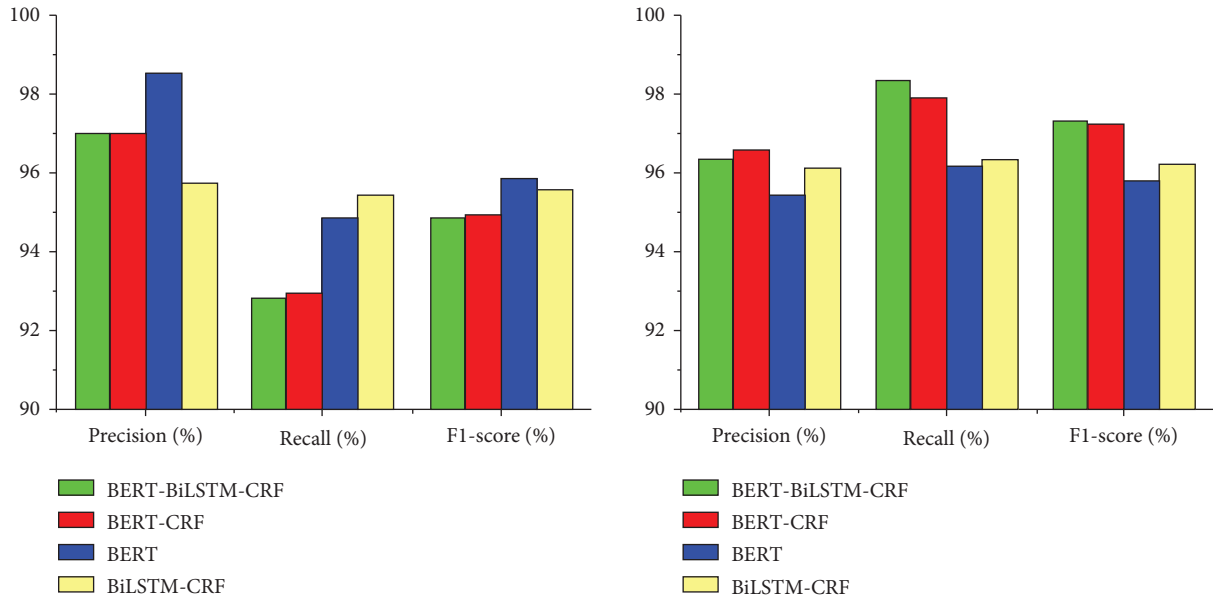


FIGURE 8: The performance of the drug and disease test results.

semantics of words in sentences. At the same time, CRF restricts the label prediction to ensure that the output of the prediction is legal and greatly improves the $F1$ -score value. It can be seen that the introduction of BERT based on the traditional named entity recognition model can achieve better results and improve the recognition accuracy in the medical field than the single model based on the neural network.

4. Conclusion

Aiming at the polysemy of words that cannot be represented based on traditional word vector representation in Chinese named entity recognition, in this study, we introduce a drug selection system with intelligent recommendation mechanism based on NER (named entity recognition) to provide online medical service diagnosis support. For the task of Chinese entity recognition, this paper obtains the contextualized word vector through the BERT language pre-processing model, which naturally supports text classification tasks and does not need to modify the model when doing text classification tasks. Then, considering the meaning relationship and adjacent temporal information between words embedding, we build the BERT-BiLSTM-CRF model for entity recognition. The experiment results show that the proposed model can achieve better recognition results than the traditional neural network models.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The paper was presented in “2022 IEEE 24th Int Conf on High-Performance Computing & Communications, 8th International Conference on Data Science and Systems, 20th International Conference on Smart City, and 8th International Conference on Dependability in Sensor Cloud and Big Data Systems and Application (HPCC/DSS/SmartCity/DependSys) as shown in [39].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Science and Technology Innovation Team of Henan University (22IRTSTHN016), special project of the Key Research and Development Plan Of Henan Province (221111111700), and the teaching reform research and practice project of Higher Education in Henan Province in 2021 (2021SJGLX502).

References

- [1] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018.
- [2] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," in *Proceedings of the NAACL-HLT*, pp. 103–112, Los Angeles, CA, USA, June, 2015.
- [3] X. Zhou, L. Yue, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020.
- [4] M. Gridach, "Character-level neural network for biomedical named entity recognition," *Journal of Biomedical Informatics*, vol. 70, pp. 85–91, 2017.
- [5] Q. Zhu, X. Li, A. Conesa, and C. Pereira, "GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text," *Bioinformatics*, vol. 34, no. 9, pp. 1547–1554, 2018.
- [6] J. Lei, B. Tang, X. Lu, K. Gao, M. Jiang, and H. Xu, "A comprehensive study of named entity recognition in Chinese clinical text," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 808–814, 2014.
- [7] C. Dong, "Character-based LSTM-CRF with radical-level features for Chinese named entity recognition," *Natural Language Understanding and Intelligent Applications*, pp. 239–250, Springer, Berlin, Germany, 2016.
- [8] 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, Article ID 103422, 2020.
- [9] H. Zhu, X. Liu, R. Lu, and H. Li, "Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM," *IEEE journal of biomedical and health informatics*, vol. 21, no. 3, pp. 838–850, 2017.
- [10] J. Wang, L. Zhang, Q. Guo, and Z. Yi, "Recurrent neural networks with auxiliary memory units," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1652–1661, 2018.
- [11] Y. Bengio, "A neural probabilistic language model," *Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.
- [12] R. Collobert, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, pp. 2493–2537, 2011.
- [13] J. He and H. Wang, "Chinese named entity recognition and word segmentation based on character," in *Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing*, Hyderabad, India, January, 2008.
- [14] Y. Xu, Y. Wang, T. Liu et al., "Joint segmentation and named entity recognition using dual decomposition in Chinese discharge summaries," *Journal of the American Medical Informatics Association*, vol. 21, pp. e84–e92, 2013.
- [15] C. Dos Santos and B. Zadrozny, "Learning character-level representations for part-of-speech tagging," in *Proceedings of the International Conference on Machine Learning*, PMLR, Pittsburgh, PA, USA, July, 2014.
- [16] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, <https://arxiv.org/abs/1508.01991>.
- [17] G. Lample, "Neural architectures for named entity recognition," 2016, <https://arxiv.org/abs/1603.01360>.
- [18] X. Ma and E. Hovy, "End-to-end sequence labeling via bidirectional lstm-cnns-crf," 2016, <https://arxiv.org/abs/1603.01354>.
- [19] Z. Yang, R. Salakhutdinov, and W. Cohen, "Multi-task cross-lingual sequence tagging from scratch," 2016, <https://arxiv.org/abs/1603.06270>.
- [20] J. P. C. Chiu and E. Nichols, "Named entity recognition with bidirectional LSTM-CNNs," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 357–370, 2016.
- [21] S. Long, "A method of Chinese named entity recognition based on CNN-BiLSTM-CRF model," in *Proceedings of the International Conference of Pioneering Computer Scientists, Engineers and Educators*, Springer, Chengdu, China, August, 2018.
- [22] F. Li, M. Zhang, B. Tian, B. Chen, G. Fu, and D. Ji, "Recognizing irregular entities in biomedical text via deep neural networks," *Pattern Recognition Letters*, vol. 105, pp. 105–113, 2018.
- [23] Y. Chen, T. A. Lask, Q. Mei et al., "An active learning-enabled annotation system for clinical named entity recognition," *BMC Medical Informatics and Decision Making*, vol. 17, no. S2, pp. 82–44, 2017.
- [24] H. Wang, W. L. K. Yeung, Q. X. Ng et al., "A weakly-supervised named entity recognition machine learning approach for emergency medical services clinical audit," *International Journal of Environmental Research and Public Health*, vol. 18, no. 15, p. 7776, 2021.
- [25] D. Zeng, "Distant supervision for relation extraction via piecewise convolutional neural networks," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September, 2015.
- [26] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, 2017.
- [27] J. Devlin, "Bert: pre-training of deep bidirectional transformers for language understanding," 2018, <https://arxiv.org/abs/1810.04805>.
- [28] S. Jiang, "A BERT-BiLSTM-CRF model for Chinese electronic medical records named entity recognition," in *Proceedings of the 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, IEEE, Xiangtan, China, October, 2019.
- [29] H. Luo, J. Wang, M. Li et al., "Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.
- [30] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, and J. Wang, "Computational drug repositioning using low-rank matrix approximation and randomized algorithms," *Bioinformatics*, vol. 34, no. 11, pp. 1904–1912, 2018.
- [31] A. P. Chiang and A. J. Butte, "Systematic evaluation of drug–disease relationships to identify leads for novel drug uses," *Clinical Pharmacology and Therapeutics (St. Louis)*, vol. 86, no. 5, pp. 507–510, 2009.
- [32] J. Zhang, C. Li, Y. Lin, Y. Shao, and S. Li, "Computational drug repositioning using collaborative filtering via multi-source fusion," *Expert Systems with Applications*, vol. 84, pp. 281–289, 2017.
- [33] H. Yan, "TENER: adapting transformer encoder for named entity recognition," 2019, <https://arxiv.org/abs/1911.04474>.
- [34] M. R. Khan, M. Ziyadi, and A. H. Mohamed, "Mt-bioner: multi-task learning for biomedical named entity recognition using deep bidirectional transformers," 2020, <https://arxiv.org/abs/2001.08904>.

- [35] D. Huang, Z. Jiang, L. Zou, and L. Li, "Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks," *Information Sciences*, vol. 415-416, pp. 100–109, 2017.
- [36] Y. Qin and Y. Zeng, "Research of clinical named entity recognition based on bi-lstm-crf," *Journal of Shanghai Jiaotong University*, vol. 23, no. 3, pp. 392–397, 2018.
- [37] A. Akbik, T. Bergmann, and V. Roland, "Pooled contextualized embeddings for named entity recognition," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, Minneapolis, Minnesota, June, 2019.
- [38] W. Fan and R. Yang, "Hot public appeal extraction and visual analysis combined BERT and spatio-temporal location," in *Proceedings of the Spatial Data and Intelligence: Third International Conference*, Wuhan, China, August, 2022.
- [39] D. Li, Y. Zhang, Y. Zhang et al., "A hybrid model based on deep convolutional neural network for medical named entity recognition," in *Proceedings of the 2022 IEEE 24th Int Conference on High Performance Computing & Communications*, pp. 2353–2357, IEEE, Chengdu, China, 2022, December.