

RESEARCH

Open Access



# Chinese medical entity recognition based on the dual-branch TENER model

Hui Peng<sup>1</sup>, Zhichang Zhang<sup>1\*</sup>, Dan Liu<sup>1</sup> and Xiaohui Qin<sup>1</sup>

## Abstract

**Background** Named Entity Recognition (NER) is a long-standing fundamental problem in various research fields of Natural Language Processing (NLP) and has been practiced in many application scenarios. However, the application results of NER methods in Chinese electronic medical records (EMRs) are not satisfactory, mainly due to the following two problems: (1) Existing methods do not take into account the impact of medical terminology on model recognition performance, resulting in poor model performance. (2) Existing methods do not fully utilize the Chinese language features contained in EMR, resulting in poor model robustness. Therefore, it is imminent to solve these two problems regarding the performance of the NER model for EMRs.

**Methods** In this paper, a TENER-based radical feature and entity augmentation model for NER in Chinese EMRs is proposed. The TENER model is first used in the pre-training stage to extract deep semantic information from each layer of the feature extractor. In the decoder part, the recognition of medical entity boundary and entity category are divided into two branch tasks.

**Results** We compare the overall performance of the proposed model with existing models on different datasets using the computed F1 score evaluation metric. The experimental results show that our model achieves the best F1 score of 82.67%, 74.37%, 70.16% on the CCKS2019, ERTCMM, and CEMR data sets. Meanwhile, in the CMeEE challenge, our model surpassed the top-3 with the F1 score of 68.39%.

**Conclusions** Our proposed model is the first to divide the NER task into a two-branch tasks, entity boundary and types recognition. Firstly, the medical entity dictionary information is integrated into TENER to obtain the feature information of professional terms in Chinese EMRs. Secondly, the features of Chinese radicals in Chinese EMRs extracted by CNN are added to the entity category recognition task. Finally, the effectiveness of the model is validated on four datasets and competitive results are achieved.

**Keywords** Electronic medical records, Named entity recognition, TENER, Char-Entity-Transformer, Dual-branch

## Background

With the rapid development of medical informatization, medical institutions have produced a large number of EMRs texts, which contain critical information

about patients in the diagnosis and treatment process, such as chief complaint, diagnosis results, treatment process and drug use. The EMRs contain extremely rich knowledge of clinical experience, which is closely related to the patient's health status. Effective use of this clinical information can greatly assist physicians in improving the accuracy of their diagnosis. As shown in Table 1, the text information in EMRs is in the form of unstructured data. Among it, the words in bold represent medical entities. Considering this situation, most of the existing methods is to transform unstructured

\*Correspondence:

Zhichang Zhang  
zzc@nwnu.edu.cn

<sup>1</sup> College of Computer Science and Engineering, Northwest Normal University, 967 Anning East Road, 730070 Lanzhou, China



**Table 1** A labeled example of Chinese EMRs entity recognition

Sample medical record	Entity type
<b>Outpatient cerebral infarction and subcortical arterioscl-erotic encephalopathy</b> were admitted to our department.	disease
A small amount of <b>phlegm sounds</b> can be heard on bilateral lung auscultation.	symptom
<b>Head CT</b> showed high-density foci in the brain parenchyma.	check
<b>Nasal endoscopic</b> double sieve, double maxillary sinus.	check
Multiple Myeloma Light Chain Stage <b>IIIA</b>	disease diagnosis clas- sification

data into structured data. NER is usually used to complete the transformation of unstructured data into structured data, which is the main carrier of relevant medical knowledge in EMRs. NER in electronic medical records is mainly to identify the boundaries of medical entities and determine their categories, which is usually regarded as a task of sequence labeling.

BERT [1], a pre-trained model based on multilayer Transformer, has achieved great success on several NLP tasks. At present, pre-trained models have received extensive attention from academia and industry as a new paradigm for processing. However, the absolute position encoding, Mask word and FP32 methods used in the training process also affect the performance of BERT in the Chinese medical entity recognition task. Subsequently, TENER (Transformer Encoder for Named Entity Recognition) [2] added position and direction information and introduced relative position encoding based on BERT, which improved the performance of the model in Chinese NER tasks. Therefore, TENER is used as the encoder our proposed model. Considering the particularity of medical data and the important features of medical vocabulary, we integrate the medical entity dictionary information into TENER through the Char-Entity-Transformer method. Through the fusion of word and entity features, the self-attention performance is enhanced to solve the problem of poor pattern recognition performance caused by many ambiguous terms in the medical field.

At the same time, we noticed that some specific types of entity words in the Chinese EMRs data set often have different characteristics from general entity words, especially in the radical kinds of particular entities. Many words that make up the entity words of disease often have the radicals. For example, these words, “scar”, “epilepsy”, “pain”, all have the radical “sickness”. Similarly, many words that make up the entity words of body parts often have the radical “moon”, such as “muscle”, “bone”, “kidney”. Radical information also has specific reference value in predicting labels, especially for medical entities composed of multiple components,

such as diseases in the “body part + symptoms” (mouth ulcers) format, which will play a key role. However, the medical NER model has not fully utilized this information. We add radical information as a basic feature to our model to improve the model’s ability to recognize Chinese EMRs.

Compared with traditional text recognition, Chinese EMRs text has many differences. It has the characteristics of concise language and strong structure. In view of these characteristics, this paper proposes a divide-and-conquer solution, which is to identify the boundaries and types of medical entities in Chinese EMRs. Conditional random field (CRF) is used after the TENER encoding layer to identify the boundaries of medical entities. And A-Softmax [3] is used to identify the entity categories. Our contributions can be summarized as follows:

- The proposed model integrates the medical entity dictionary information into TENER for the first time. It solves the problem of poor recognition of medical entities caused by the vague expression of professional terms in EMRs.
- We propose the idea of decomposing the NER task into two-branch tasks of entity boundary and type recognition to solve. In the task of entity category recognition, Chinese radical features are introduced to promote the prediction of entity categories.
- The best F1 score achieved by the experiment was 82.67%, 74.37%, 70.16% on the CCKS2019, ERTCMM, and CEMR data sets. And the proposed model achieved the F1 score of 68.39% in the CMEE challenge, surpassing the top-3.

## Related work

### Chinese NER

Early NER methods include rule-based methods and dictionary-based methods [4, 5]. With the rise of deep learning, it gradually applied to NER tasks. The method based on deep learning can automatically learn the feature information of the data, so as to obtain the feature

representation of the word and complete the entity label prediction. The deep learning based methods are usually divided into dynamic structure and adaptive embedding structure.

Lattice LSTM [6] introduces a compatible LSTM [7] that brings lexical information to Chinese NER tasks. In the LR-CNN [8] model, CNN is used for stacked encoding and rethink mechanism to solve the problem of vocabulary conflict. However, Lattice LSTM and LR-CNN have slow inference speed and cannot capture long-distance dependence. The CGN [9] is developed from the collaborative graph network (GAN). LGN [10] aggregates local and global information to build a graph network, making full use of lexical feature information. FLAT [11] introduces vocabulary information by designing a position vector, using a Transformer to capture long-distance dependence and improve inference efficiency. The above methods belong to dynamic structure methods, but these methods have poor generalization performance.

The adaptive embedding structure is independent of the model and has strong portability. WC-LSTM [12] dynamically encodes the input of Lattice LSTM through four encoding strategies. Multi-digraph [13] introduces entity dictionaries to better model the information interaction between characters and dictionaries through a graph structure. Simple-Lexicon [14] introduces lexical information through the Soft-lexicon method. The model structure is simple and the inference speed is fast.

### TENER

For the NER task, Since Transformer utilizes absolute position encoding, it lacks the ability of direction awareness and does not sufficiently extract contextual feature information. To improve the performance of Transformer on NER tasks, it is critical to improve the position and direction perception characteristics of the Transformer. Inspired by the idea of the relative distance [15] and Transformer-xl [16], a new relative position encoding was used in TENER. It can enhance the model's ability to sense the direction of contextual feature information. The calculation formula of relative position coding is shown in formula (1).

$$R_t, R_{-t} = \begin{bmatrix} \sin c_0 t \\ \cos c_0 t \\ \vdots \\ \sin(c \frac{d}{2} t) \\ \cos(c \frac{d}{2} t) \\ \vdots \\ \sin(c \frac{d}{2} t) \\ \cos(c \frac{d}{2} t) \end{bmatrix}, \begin{bmatrix} -\sin c_0 t \\ \cos c_0 t \\ \vdots \\ -\sin(c \frac{d}{2} t) \\ \cos(c \frac{d}{2} t) \\ \vdots \\ -\sin(c \frac{d}{2} t) \\ \cos(c \frac{d}{2} t) \end{bmatrix} \quad (1)$$

It can be seen from formula (1) that the relative position coding is directional, that is, the relative distance coding

in different directions is inconsistent. It is also used in the computation of attention score, as shown in Eq. (2).

$$A_{t,j}^{rel} = Q_t K_j^T + Q_t R_{t,j}^T + u K_j^T + v R_{t,j}^T \quad (2)$$

where  $Q_t K_j^T$  represents the attention score between any two tokens,  $Q_t R_{t,j}^T$  represents the offset of the  $t$ -th token in determining the relative distance,  $u K_j^T$  represents the offset of the  $j$ -th token, and  $v R_{t,j}^T$  represents the offset in determining the relative distance and direction.

### Methods

The architecture of our proposed model is shown in Fig. 1. In the encoder part, the TENER pre-trained model backbone network is used. The medical entity dictionary information is integrated into TENER to obtain the feature information of professional terms in Chinese electronic medical records. Our model is a two-branch model based on TENER. One of the tasks is the recognition of entity boundary and the other is the recognition of entity types. The boundary information of the entity is obtained through CRE, and the entity category is classified through A-Softmax. The features of Chinese radicals in Chinese EMRs extracted by CNN are added to the entity category recognition task.

### Encoding layer

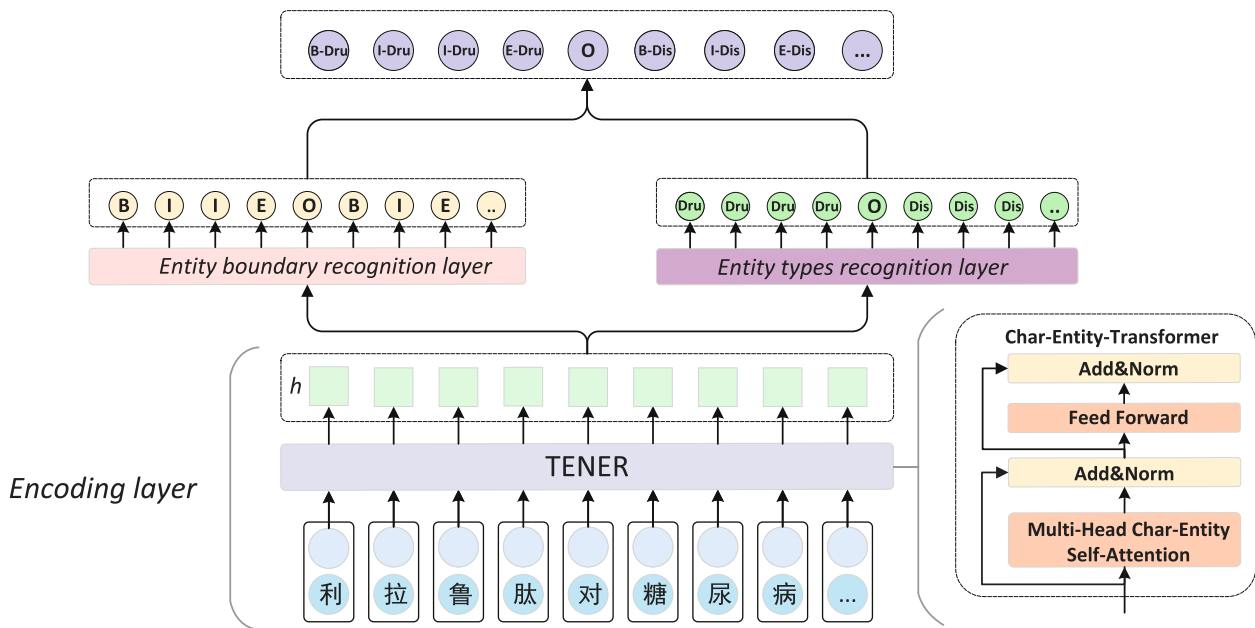
Given the character sequence  $c = \{c_1, c_2, \dots, c_T\}$  and the entity dictionary  $\varepsilon_{ent}$  extracted from the training text, we use the maximum entity matching algorithm to obtain the corresponding entity tag sequence  $e = \{e_1, e_2, \dots, e_T\}$ . In particular, we label each character with the index of the longest entity in the  $\varepsilon_{ent}$  that includes the character, and label characters with no entity matches with 0.

Figure 2 shows the Char-Entity-Transformer structure. Following TENER, given the character sequence  $c = \{c_1, c_2, \dots, c_T\}$ , the representation of the  $t$ -th ( $t \in \{1, 2, \dots, T\}$ ) character in the input layer is the sum of character and position embeddings, represented as:

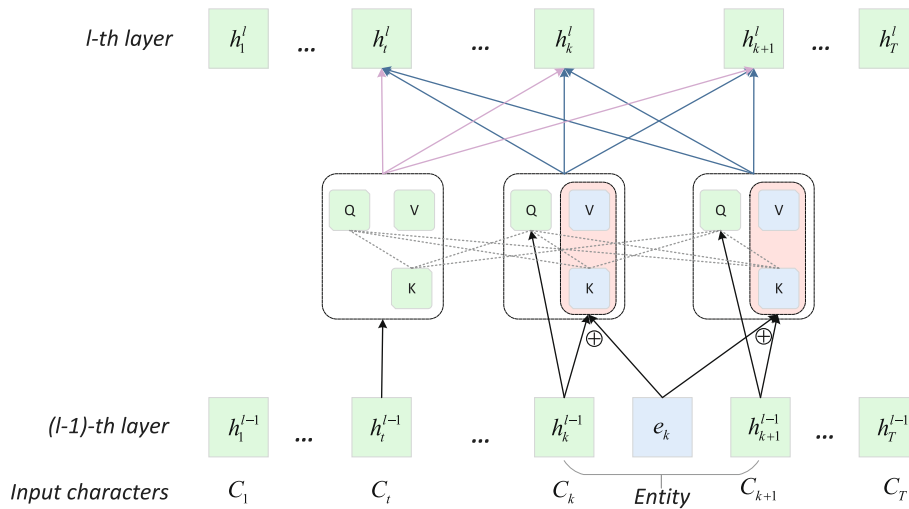
$$h_t^1 = E_c[c_t] + E_p[t] \quad (3)$$

where  $E_c$  and  $E_p$  represent character encoding table and position encoding table respectively.

Given the  $(l - 1)$ -th layer character hidden sequence  $\{h_1^{l-1}, h_2^{l-1}, \dots, h_T^{l-1}\}$ , the calculation process of the  $l$ -th layer query matrix  $Q^l = \{q_t^l\}_{t=1}^T \in \mathbb{R}^{T \times H_c}$  is the same as the query matrix in the traditional self-attention. But for *key* matrix  $K^l = \{k_t^l\}_{t=1}^T \in \mathbb{R}^{T \times H_c}$ , *value* matrix  $V^l = \{v_t^l\}_{t=1}^T \in \mathbb{R}^{T \times H_c}$ , we calculate the combination of hidden characters and their corresponding entity coding as follows:



**Fig. 1** The structure of the dual-branch TENER model. As shown in the Fig. 1, “利” is the entity’s beginning and “肽” is the end of the entity. Among it, “利拉鲁肽” is labeled as the Dru(Drug), and “糖尿病” is labeled as the Dis(Diseases)



**Fig. 2** The structure of Char-Entity-Transformer

$$\begin{aligned}
 q_t^l &= h_t^{l-1} W_{h,k}^t \\
 k_t^l &= \begin{cases} h_t^{l-1} W_{h,k}^l & \text{if } e_t = 0, \\ \frac{1}{2}(h_t^{l-1} W_{h,k}^l + E_{ent}^T[e_t] W_{e,k}^l) & \text{else;} \end{cases} \\
 v_t^l &= \begin{cases} h_t^{l-1} W_{h,v}^l & \text{if } e_t = 0, \\ \frac{1}{2}(h_t^{l-1} W_{h,v}^l + E_{ent}^T[e_t] W_{e,v}^l) & \text{else;} \end{cases}
 \end{aligned} \tag{4}$$

where the  $W_{h,q}^t, W_{h,k}^l, W_{h,v}^l \in \mathbb{R}^{H_c \times H_c}$  are the trainable parameters of the  $l$ -th layer, and the  $W_{e,k}^l, W_{e,v}^l \in \mathbb{R}^{H_e \times H_c}$ ,

are trainable parameters for the corresponding entities, and  $E_{ent}$  is the entity encoding table.

As shown in formula (4), if a character does not have a corresponding entity, the representation is equal to the baseline self-attention. To illustrate how the character and its corresponding entity are jointly coded, a set of entity codes  $\{E_{ent}[e_1], E_{ent}[e_2], \dots, E_{ent}[e_T]\}$  is used, where  $e \in \mathbb{R}^{T \times H_e}$ . Among it, the attention score  $S_i^l$  of the  $i$ -th character character is as follows:

$$\begin{aligned}
 S_i^l &= \text{softmax}\left(\frac{q_i^l K^{lT}}{\sqrt{d_k}}\right) \\
 &= \text{softmax}\left(\frac{q_i^l (h_{l-1} W_{h,k}^l + e W_{e,k}^l)^T}{2\sqrt{d_k}}\right) \\
 &= \left(\frac{\sqrt{s_i^c s_i^e}}{\sum_j \sqrt{s_j^c s_j^e}}\right)^T_{t=1}
 \end{aligned} \tag{5}$$

s.t.

$$\begin{aligned}
 s_i^c &= \exp\left(\frac{q_i^l (h_{l-1}^T W_{h,k}^l)^T}{\sqrt{d_k}}\right) \\
 s_i^e &= \exp\left(\frac{q_i^l (e^T W_{e,k}^l)^T}{\sqrt{d_k}}\right)
 \end{aligned}$$

where the attention score  $s_i^c$  for each character is the same as the traditional self-attention calculation, and the char-to-entity attention score  $s_i^e$  represents the similarity between the character and the corresponding entity.

### Entity boundary recognition layer

Generally, in a tag sequence, there are interdependent and mutually constrained relationships between tags. For example, the I (inside) tag should be after the B (begin) tag or the I tag. A set of entity tags should have the same entity type as possible. CRF can describe the dependencies between tags through the transition matrix, and obtain the globally optimal tag sequence.

The optimization goal of the CRF layer is to increase the proportion of the score of the true label sequence to the total. Let  $[y]_1^T$  be the true label sequence, and  $[j]_1^T$  takes all possible label sequences. The formula for the proportion of the true label sequence to the total is as follows:

$$S([x]_1^T, [i]_1^T) = \sum_{t=1}^T (A_{[i]_{t-1}, [i]_t} + [M([x]_1^T)]_{[i]_t, t}) \tag{6}$$

$$p([y]_1^T || [x]_1^T) = \frac{e^{S([x]_1^T, [y]_1^T)}}{\sum e^{S([x]_1^T, [j]_1^T)}} \tag{7}$$

$$Loss_C = -S([x]_1^T, [y]_1^T) + \log\left(\sum_j e^{S([x]_1^T, [j]_1^T)}\right) \tag{8}$$

The loss function is estimated by the negative log-likelihood of the optimization objective, and the calculation formula is as formula (8).

### Entity types recognition layer

#### Radical features

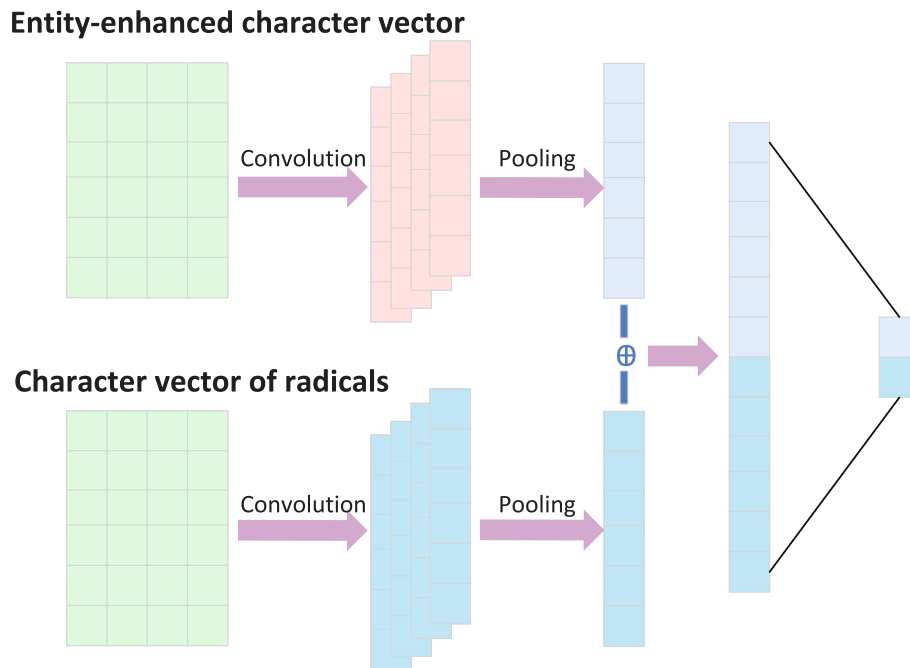
Traditional NER models are aimed at generic datasets. The features extracted by the underlying network are limited to the semantic-level features of the context, and lack

attention to domain-specific datasets, such as the unique features of Chinese EMRs. Just like in English, you can guess the meaning and nature of the word based on the root and affix of the word. Chinese radicals and strokes also contain a lot of word meaning information, and the flexible use of this information can improve the effect of the model. The difference between Chinese electronic medical record data and general data set lies in the particularity of its domain, so that most of the characters constituting entity words in this type of data are limited and have special characteristics. For example, the Chinese five elements “Gold, Wooden, Water, Fire, Dust” are also often used as the characteristic radicals of various entity words in the medical field. Gold corresponds to the common trace elements in the body and the names of some medicines, such as “× × ×Na”, “× × × Ca”. Wooden corresponds to “× × ×surgery”, “physical examination× × ×”, “× × ×thrombosis”, “vertebra”, and the names of some Chinese herbal medicines. Water corresponds to various body fluids (plasma, tissue fluid, lymph fluid) and some symptoms, such as “wet”, “slip”, “spill”, “dissolve”, “burst”. Fire is the end character for various inflammation, as well as some symptom entity characters, such as “scald”, and “rot”. Dust corresponds to “droop”, and body parts modification words “wall”, and so on. In addition, there are others radicals such as “heart”, “car”, “gas” and “mouth”, which also have important reference value for identifying medical entity words.

As shown in Fig. 3, a character-level convolutional neural network (CNN) is designed to capture the fine-grained radical features. It consists of two parts: Entity-enhanced character vector and Character vector of radicals. The CNN structure is the same, and different convolution kernels (2, 3, 4) are selected to extract the key information of the context. The input of the Entity-enhanced character vector part comes from The output of TENER (768 dimensions), the input of Character vector of radicals comes from the Word2Vec (768 dimensions) of the Chinese dictionary. It takes the vector of each character and its radical in the medical data as input, and outputs the sentence vector containing the character and radical information. The obtained character vector is concatenated with the radical vector, and then the concatenated vector is input into the fully connected layer to obtain the result of entity classification. In this way, more and more fine-grained character-level features can be captured.

#### A-Softmax

The Softmax classification algorithm is the most commonly used classification algorithm. The traditional Softmax algorithm uses the same format when learning samples of the same and different classes. This results in poor intra-class and inter-class discrimination of learned features. However, the A-Softmax algorithm increases the difficulty of



**Fig. 3** The structure of the radical feature extraction module

learning when learning similar samples, which makes the features more distinguishable. This makes predictions for the model difficult and leads to poor classification results. Therefore, we adopt A-Softmax as the classification algorithm for our proposed model. This has a positive impact on the classification of entity types in EMRs. The loss of the entity types recognition layer is as follows:

$$Loss_F = \frac{1}{N} \sum_i -\log\left(\frac{e^{\|x_i\|\varphi(\theta_{y_i,i})}}{e^{\|x_i\|\varphi(\theta_{y_i,i})} + \sum_{j \neq y_i} e^{\|x_i\|\cos(\theta_{j,i})}}\right) \quad (9)$$

Traditional NER tasks in EMRs have only one input and one output. For the two different recognition tasks of entity boundary and type, we use the network structure shown in Fig. 1 to process the two tasks. Finally, the loss of entity boundary and types recognition layer are trained together as the global loss. The specific formula is as follows:

$$Loss = \alpha Loss_B + \beta Loss_F \quad (10)$$

where the subscripts B and F respectively represent the entity boundary recognition layer and the entity types recognition layer.

### The semi-supervised module

As shown in Fig. 4, the train process of our proposed model mainly includes two steps. The first step is to train the

model(M) with a small amount of labeled data. The second step is to use the trained model(M) to predict unlabeled samples. Then the unlabeled samples with high confidence are selected and added to the labeled data iteratively. These two steps are repeated until the preset stopping conditions are reached, and then an ideal model M' can be obtained. Algorithm 1 is the process of a semi-supervised training method. Parameters of Semi-supervised module shown in Table 2. The loss function calculation process of the semi-supervised training method is shown in Eq. (11).

---

**Input:**  $L; U; M; \theta; H; I_{Itermax}$ , where  $M$  is the Model used in the iteration process,  $\theta$  is the number of unlabeled samples selected in the next iteration, and  $I_{Itermax}$  is the maximum number of iterations, and  $H$  is the selection metric  $S(U_{time}, \theta, M, H)$  is the selection function.  
**Output:** output:Model M after training.

- 1:  $t_{time} = 0, L_{time} = L, U_{time} = U$ , where  $t_{time}$  is the number of iterations,  $L_{time}$  is the labeled sample set at the  $t_{time}$  iteration,  $U_{time}$  is the unlabeled sample set at the  $t_{time}$  iteration.
- 2: **while**  $U \neq \emptyset$  or  $t_{time} < I_{Itermax}$  **do**
- 3:   Train M on  $L_{time}$
- 4:    $S_{time} = S(U_{time}, \theta, M, H)$ ,  $S_{time}$  is the selected unlabeled sample set
- 5:    $U_{time} = U_{time} \cup S_{time}; L_{time} = L_{time} \cup S_{time}$
- 6:    $t_{time} = t_{time} + 1$
- 7:  $L = L_{time}$  Training model M
- 8: Output the trained model M'

---

**Algorithm 1** The semi-supervised training process

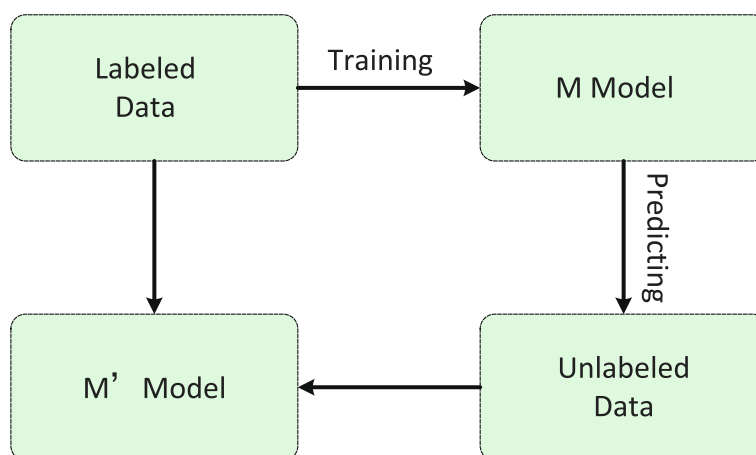


Fig. 4 The structure of semi-supervised module

Table 2 Parameters of semi-supervised module

Parameters	Significations
$L = (x_i, y_i)$	A label set
$x_i$	The labeled sample, $x_i \in R^d$
$R^d$	The d-dimensional real number set
$y_i$	The true label, $y_i \in \{w_1, w_2, \dots, w_s\}$
$s$	The number of different labels
$U$	Unlabeled set
$U = \{x_1, x_2, \dots, x_m\}$	$m$ is the number of unlabeled samples, $m > n$

$$Loss = (1 - \theta) \times Loss_{LD} + \theta \times Loss_{ULD} \tag{11}$$

where the subscripts *LD* and *ULD* respectively represent labeled data and unlabeled data.

### Experiment

We evaluated the proposed model on four datasets, including three publicly available NER datasets (CCKS 2019, ERTCMM, CMeEE) and our own data set (CEMR), which is also one of the contributions of this paper. However, since CEMR contains private data of patients (name, age, home address, etc.), it cannot be made public. The statistics of the four datasets are shown in Table 3.

Table 3 Statistics of the four datasets

Data Set	Training	Validation	Test	Category
CCKS2019	800	200	379	6
ERTCMM	1200	300	497	13
CMeEE [18]	15000	5000	3000	9
CEMR	3000	500	500	5

### Dataset

The total annotated entities in the dataset CCKS2019 are divided on the six types as: Diseases and Diagnosis, Imaging Examinations, Laboratory Tests, Surgery, Drugs, and Anatomical Parts.

The data of ERTCMM (Entity Recognition of Traditional Chinese Medicine’s Manual) comes from the instructions of traditional Chinese medicine, which contains 13 types: Drug, Drug ingredient, Disease, Symptom, Syndrome, Disease group, Food, Food group, Person group, Drug group, Drug dosage, Drug taste, Drug efficacy.

The CMeEE data set is jointly provided by several institutions in China including Peking-University, Zhengzhou University, Pengcheng Laboratory and Harbin Institute of Technology (Shenzhen). It includes 9 types of entity: Disease, Clinical manifestations, Drugs, Medical equipment, Medical procedures, Body, Medical test items, Microorganisms, Departments. The total number of words in the annotated data reaches 2.2 million, which contains 938 files. The average number of words per file is 2355. The data set includes 504 common pediatric diseases, 7085 body parts, 12907 clinical manifestations, and 4354 medical procedures. Slightly different from traditional NER, there is a nested relationship between entities. Nested entities are a common phenomenon in a medical text, so the model processing is more complicated than commonly used NER models.

In order to facilitate the study of medical entity recognition tasks and related topics on Chinese medical record texts, according to the I2B2/VA English medical record text labeling guidelines [17], 4000 real Chinese medical record texts were manually annotated. Finally, it is used as the medical entity recognition dataset (CEMR). All

electronic medical records come from the top-3 hospitals in Gansu Province, China. It contains 5 types of entity: Symptoms and characteristics, Examinations, Abnormal examination results, Diseases, and Treatments.

**Experimental parameter setup**

Each experiment of the model in this paper is repeated three times, and the average result of the ranking is the final result. To be fair, we set the same hyper-parameters (namely, hidden layer size, number of layers, number of attention heads) in the pre-trained model. In addition, unless otherwise specified, for training, all models are trained using the same optimizer Adam.

Adversarial training is used in the model training process, fast gradient method (FGM) and projected gradient descent method (PGD) are used to introduce noise and adjust parameters to alleviate the problem of poor robustness of the model. Finally, the generalization ability of the model is improved. At the same time, mixed-precision training is used to improve the problem of reduced computational efficiency due to adversarial training. FP16 is used for storage and multiplication in the memory to speed up training, and FP32 is used for accumulation to avoid rounding errors. And FP32 is expanded by  $2^k$  times before back-propagation loss to prevent the loss from overflowing. During training, the model weights of the last few epochs are weighted averaged to get a smoother and better performing model. To prevent the model from over-fitting, we use 5-fold cross-validation. The differential learning rate is used in the training process. The learning rate of the TENER layer uses 0.00002, and the other layers use 0.02. The parameters of the model are shown in Table 4.

**Table 4** Hyperparameters

Parameter Name	Value
Clipping gradient	2
The number of TENER encoder layer	24
Dropout	0.3
Hidden size	768
Weight decay	0.01
The number of hidden layer	12
Gradient accumulation steps	2
The number of attention head	12
Hidden activation function	GELU
Epoch	200
Max sequence len	128
Batch size	16

**Results**

We summarized the overall performance by computing the F1 score. The results are illustrated in Table 5. We conduct experiments on four datasets using different pre-trained models and downlink encoders. To fairly evaluate the performance of all methods, precision (P), recall (R), and F1 score (F1) are used as evaluation metrics to evaluate the recognition ability of the models.

Table 5 shows the results of the four datasets on the current mainstream pre-trained models. Tables 6 and 10 are the comparisons with the existing state-of-the-art methods and the results of the challenge on the CCKS2019 dataset. Tables 7 and 11 are the comparison results with the existing state-of-the-art methods and the challenge results on the ERTCMM dataset. Tables 8 and 12 are the comparison with the existing methods and challenge results on the CMeEE dataset. Table 9 shows the comparison results between our proposed method and the existing state-of-the-art methods on the CEMR dataset.

In Table 5, although the precision (85.70%) of our model is not the highest, lower than ELECTRA-Rtransformer-CRF (86.18%). This is due to ELECTRA borrowing ideas from GANs and pre-trained reinforcement learning. However, our proposed model achieves the highest F1 score on all four datasets, which demonstrates the superiority of our model over models. In Table 6, the recall of ELMo-TENER-CRF (87.61%) is higher than that of our model (87.18%), which is because there is a small amount of polysemy in the CCKS2019 dataset. Nonetheless, we achieved the highest F1 score. In Table 10, F1 score of the top-3 in “surgery” is 1.35% higher than our model. Our model is 0.79% higher than top-1 (85.62%) overall, and we achieve the highest F1 score on the other three datasets. And our model achieves the best results on the CCKS2019 dataset. In both Tables 7 and 11, our model achieves the highest F1 score. This shows that our model has a good ability to identify multiple medical entity categories (The number of ERTCMM medical entity category is 13, model contains rich external medical entities). The top-1 in the challenge used adversarial training, multi-head selection, pointer network, model fusion and other methods. In Table 8, the F1 score of the human-labeled result is 67.0%, and our model outperforms the human-labeled by 1.39%. In Table 12, we surpassed the top-3 in the challenge (68.135%) by 0.264% and achieved the top-3. This proves that our model can capture long-distance dependence. (The average sentence length in CMeEE dataset is 2355 words). In our annotated dataset CEMR, our proposed model still outperforms other mainstream models. Overall, our model achieves decent experimental results on four different datasets, which demonstrates the superiority of our model.



**Table 5** Overall results of data augmentation, where Rt stands for R-Transformer, TE stands for TENER

Model	CCKS2019			ERTCMM			CMeEE			CEMR		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BERT-Rt-CRF	82.42	83.23	82.82	61.42	72.44	66.48	63.45	65.27	64.34	63.04	62.09	62.56
TENER-LSTM-CRF	84.99	87.12	86.02	65.65	74.19	69.66	64.07	67.29	65.64	66.54	64.92	65.72
TENER-Rt-CRF	83.14	81.71	81.20	67.35	75.33	71.12	64.53	63.08	63.79	67.22	66.82	67.03
RoBERTa-TE-CRF	84.11	<b>87.24</b>	85.65	67.25	73.22	70.11	62.74	60.11	61.39	66.23	67.06	66.60
RoBERTa-Rt-CRF	85.61	86.24	85.92	67.68	73.45	70.45	62.52	58.02	60.18	67.38	68.65	68.01
ELECTRA-Rt-CRF	<b>86.18</b>	86.23	86.20	68.19	76.44	72.08	64.71	63.32	64.00	69.25	66.03	67.60
<b>Ours</b>	85.70	87.18	<b>86.41</b>	<b>68.34</b>	<b>81.56</b>	<b>74.37</b>	<b>67.45</b>	<b>69.37</b>	<b>68.39</b>	<b>69.32</b>	<b>71.04</b>	<b>70.16</b>

**Table 6** Experimental results at CCKS2019, where (baseline) is the model comparison baseline

Model	P(%)	R(%)	F1 score(%)
BiLSTM-CRF(baseline)	81.49	80.52	81.00
IDCNN-CRF(baseline)	80.56	81.47	81.01
BERT-CRF(baseline)	81.82	79.33	80.56
Ra-RC(Wu et al.[19])	83.31	82.44	82.87
ELMo-ET-CRF (Wan et al.[20])	83.65	<b>87.61</b>	85.59
(ELMo)-lattice-LSTM-CRF(Li et al.[21])	84.69	85.35	85.02
<b>Ours</b>	<b>85.70</b>	87.18	<b>86.41</b>

**Table 7** Experimental results at ERTCMM. where (baseline) is the model comparison baseline, (LSTM) means that the method is based on LSTM

Model	P(%)	R(%)	F1 score(%)
RoBERT-CRF(baseline)	65.79	73.97	69.64
RoBERT-BiLSTM-CRF(baseline)	67.56	73.36	70.34
Lattice LSTM	60.52	76.15	67.44
Soft-Lexicon(LSTM)	62.27	75.52	68.26
BERT-Soft-Lexicon(LSTM)	67.01	76.66	71.51
<b>Ours</b>	<b>68.34</b>	<b>81.56</b>	<b>74.37</b>

## Discussion

### Ablation study

To verify the contributions of the three main modules (entity dictionary, dual branch, radical feature), we conduct ablation experiments on four datasets. From the results showed in Table 13, it can be seen that three main modules have improved the performance of this model to varying degrees, where “-” means to delete the component, and “→” means to replace the

**Table 8** Experimental results at CMeEE, the laboratory results are derived from (Zhang et al. [18])

Model	F1 score (%)
BERT-base	62.1
RoBERTa-large[22]	62.1
ALBERT-xxlarge[23]	61.8
ZEN[24]	61.0
MacBERT-base[25]	60.7
MacBERT-large[25]	62.4
PCL-MedBERT[26]	60.6
Human[18]	67.0
<b>Ours</b>	<b>68.39</b>

**Table 9** Experimental results at CEMR

Model	P(%)	R(%)	F1 score(%)
Lattice-LSTM[6]	66.82	70.14	68.44
LR-CNN[8]	65.29	63.84	64.56
CGN[9]	65.92	70.01	67.90
LGN[9]	69.83	66.48	68.11
FLAT[11]	69.26	70.92	70.08
<b>Ours</b>	<b>69.32</b>	<b>71.04</b>	<b>70.16</b>

component. “TENER→BERT” means to replace the coding layer of TENER with the coding layer of BERT.

In particular, the addition of our entity dictionary has a specific effect on the model. The entity dictionary collects many medical entities and improves the model’s ability to recognize medical terminology. Since TENER’s whole word masking and relative position coding make the overall performance higher than BERT, the change of character-level encoder has little impact on the result. After merging the dual branches of the model into a single branch, it can be found that the experimental results of the four datasets all drop

**Table 10** The results of the CCKS2019 dataset in the challenge [27]

Model	Disease and diagnosis	An examination	Test	Surgery	Drug	Anatomy	Overall
Top-1	84.29	86.29	76.94	83.33	96.02	86.18	85.62
Top-2	-	-	-	-	-	-	85.59
Top-3	82.81	88.01	75.65	<b>86.79</b>	94.49	85.99	85.16
Our	<b>85.12</b>	<b>86.31</b>	<b>77.01</b>	85.42	<b>96.13</b>	<b>86.37</b>	<b>86.41</b>

**Table 11** The results of the ERTCMM dataset in the challenge [28]

Model	P(%)	R(%)	F1 score(%)
Top-1	67.71	78.95	72.90
Top-2	65.89	81.23	72.76
Top-3	66.60	79.74	72.58
Ours	<b>68.34</b>	<b>81.56</b>	<b>74.37</b>

**Table 12** The results of the CMeEE dataset in the challenge [29]

Model	F1 score (%)
Top-1	<b>69.383</b>
Top-2	69.105
Top-3	68.135
Ours	68.399

**Table 13** The ablation experiment results on four datasets - F1 score

Model	CCKS2019 F1 score (%)	ERTCMM F1 score (%)	CMeEE F1 score (%)	CEMR F1 score (%)
Ours	<b>86.41</b>	<b>74.37</b>	<b>68.39</b>	<b>70.16</b>
- Entity dictionary	82.13	70.26	63.55	67.29
TENER→BERT	80.35	64.93	61.04	63.05
Single	85.37	68.32	64.96	68.21
- Radicals feature	83.75	71.02	64.47	66.19
TENER→Glove 100d	84.38	66.32	63.96	67.36
A-Softmax→Softmax	86.02	69.57	65.75	69.23
- Semi-supervised	86.33	73.41	65.83	70.03

significantly. This proves that the dual-branch module can effectively identify the boundaries and categories of entities. From the experimental results, it can be found that the experimental results of removing the radical feature module are reduced in all four datasets.

This proves that the radical feature can promote the performance of the model to recognize Chinese electronic medical records. The encoding dimension of Glove-100 is much smaller than that of TENER. If the encoded dimension is changed to a larger value, the training speed of the model will decrease significantly. At the same time, we also verified A-Softmax can maintain the compactness within the class and the distance between the classes in the medical entity, and the F1 score on the CEMR data set has increased by 0.93. The experimental results show that these three modules can help the model learn Chinese EMRs information. And the semi-supervised model solve the problem of a small amount of data in the clinical field and the particularity of clinical data.

The experimental results show that the three main modules (entity dictionary, dual branch, radical feature) in this model have played an important role. The addition of entity dictionary enables the model to learn more domain knowledge and improve its understanding of context. Dual The branch can better improve the ability of entity recognition and entity classification, and the radial feature module improves the model's ability to perceive Chinese, which is an important inspiration for the identification of Chinese medical electronic medical records.

Figure 5 is a heat map of the attention weight matrix. The weight matrix used in the heatmap is the attention weights of the 12th multi-head self-attention in the last layer in the TENER module. In the heatmap, the darker the color, the higher the correlation between the two words. Because the weight on the diagonal represents the relationship between each word in the sentence and itself, so the weight is generally larger. Among it, (a) is no the heat map of the weight matrix obtained after the network model of the entity dictionary is coded. (b) is the heat map of the weight matrix obtained after the network model of the entity dictionary is coded. Among them, "phlegm sound" is used as an entity. In (a), the color between each word of the entity is lighter, and the degree of entity association is weak, which leads to the low performance of the model entity recognition. The difference is that in (b) because of The addition of the entity

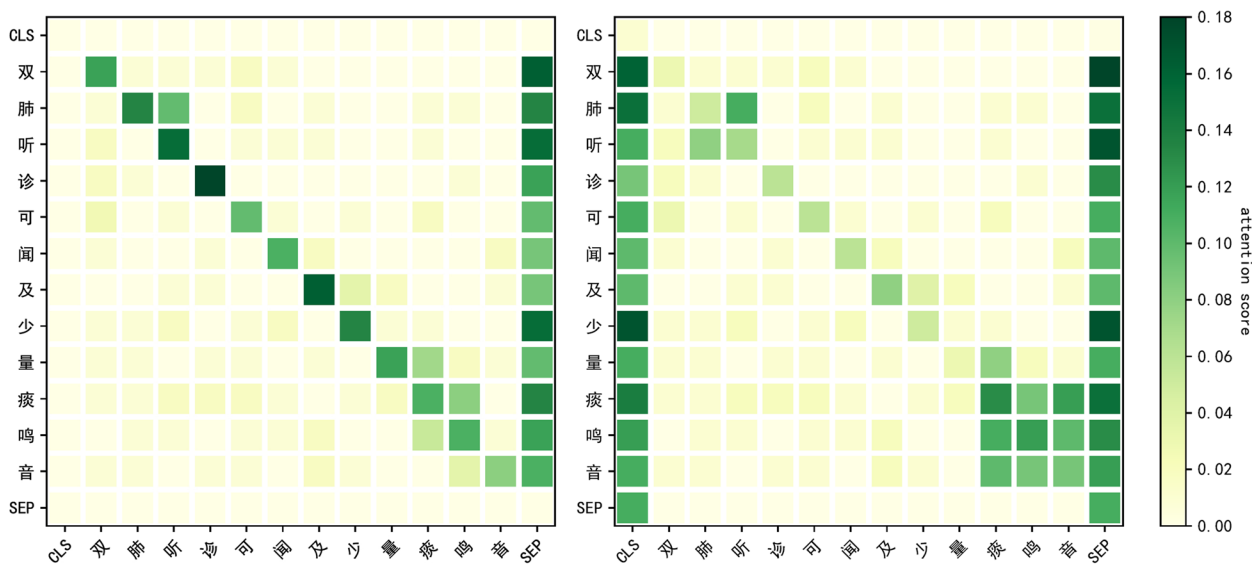


Fig. 5 The heat map of attention weight matrix

dictionary makes the association between entities close and improves the recognition performance of model entities. When the heat maps are different, it can be clearly found that the proposed model can understand the entity relationship in the sentence at a deeper level. And the entity information enhances the context relationship. This is because the attention mechanism contains entity dictionary information. The pairing of the heat map can be intuitively explained: fusing entity information in the self-attention mechanism can improve the ability of the

self-attention mechanism to capture sentence semantics, thereby improving the overall performance of the model.

Figure 6 shows the F1 score of the four data sets. Among them, (a) is the F1 score change of the model with the supervised training method, (b) is the F1 score change of the model with the semi-supervised training method.

We can clearly find that after adding the semi-supervised training method to the model, in the CCKS2019 dataset, the effect achieved by 80% of the data in (a) is similar to the effect achieved by 50% of the data in (b).

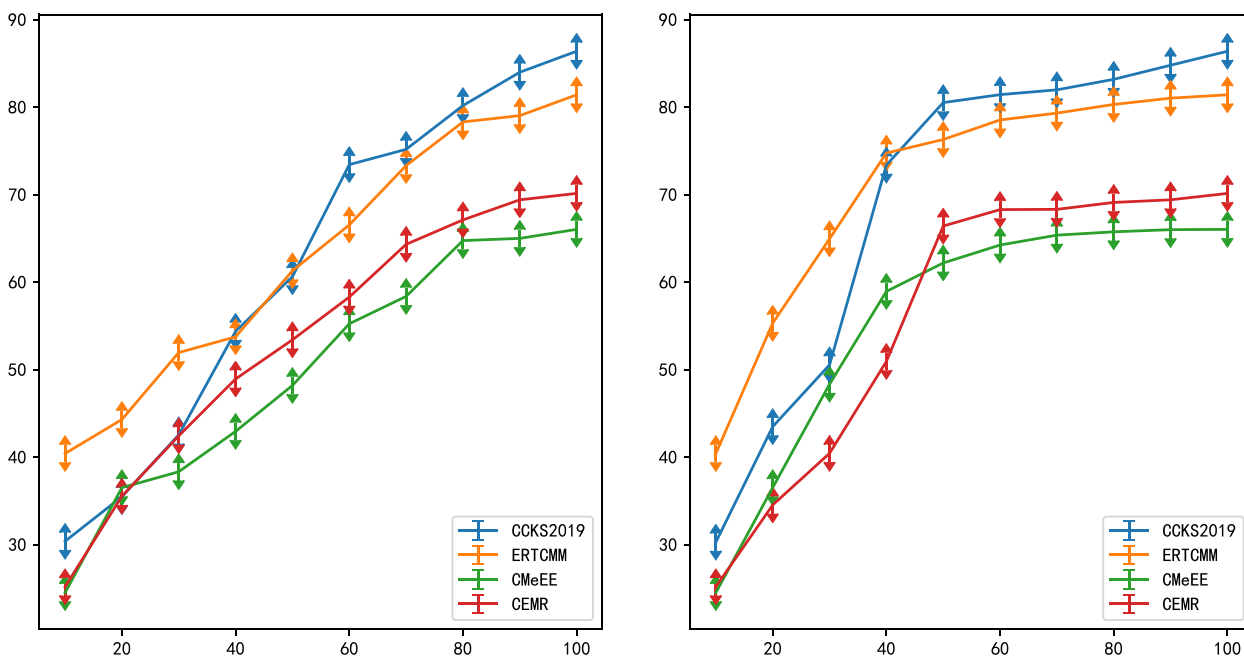


Fig. 6 The performance of the model before and after adding the semi-supervised training method

The same performance is also observed in the other three datasets. Therefore, the use of semi-supervised models can effectively reduce the dependence on the amount of data. Experiments show that the model in this paper has a high recognition ability for medical clinical data when the training data is scarce. It can alleviate model performance degradation due to lack of medical data.

**Case study**

We compare the recognition results of the proposed model with three other existing models (BiLSTM-CRF, BERT-CRF, TENER-Rtransformer-CRF) on a case. Table 14 shows a case study on the ERTCMM dataset. Red indicates that the boundary is recognized incorrectly, green indicates that the boundary is recognized correctly, blue indicates that the category is recognized incorrectly, and black indicates that the category is recognized correctly. Among them, “removal of blood stasis” is a “symptom”, “blood stasis” is a “syndrome”, and “hyperplasia of mammary glands” is a “disease grouping”.

As can be seen from the Table 14, our model can correctly identify the categories of all three entities, while the other three existing models can only correctly identify a part of the entities. BiLSTM-CRF and BERT-CRF cannot accurately identify entity boundaries. Although TENER-Rtransformer-CRF can accurately identify entity boundaries, it cannot accurately identify entity categories. TENER-Rtransformer-CRF has lower entity recognition performance for similar categories.

**Conclusions**

We propose a novel, dual-branch TENER model for entity recognition of Chinese EMRs. Results on the four datasets show that it is highly effective to explicitly integrate the Chinese medical entity dictionary into TENER pre-training. Multiple ablation experiments demonstrate that the dual-branch model formed by adding the entity boundary recognition module is effective for the Chinese medical entity recognition task. And the addition of Chinese radicals makes the model pay more attention to the extraction of unique features of Chinese EMRs. Our method outperforms the state-of-the-art methods for Chinese NER. Worth emphasizing is the limitation of this study is Chinese medical data, and the categories of data entities are distributed evenly. At the same time, there is no Chinese electronic medical record data of nested entities. The existence of polysemy problems (Ambiguity, Lack of context, Out-of-vocabulary words, Overfitting, Data imbalance) has affected the performance of NER, and solving polysemy problems is crucial to improving the accuracy of the NER model. In future work, we can use Context-aware embeddings, Rule-based systems, Named Entity Linking, Domain-specific dictionaries, Ensemble models and other methods to reduce the impact of polysemy on NER. In the future, we aim to increase the processing speed of the model and reduce parameters. Simultaneously, the generalization ability of the model is improved, so that the model has a good performance in multilingual data sets.

**Table 14** Examples from the ERTCMM test set

Model	Soften firmness and dissipate knots, <b>remove blood stasi</b> and relieve pain. In traditional Chinese medicine, heat toxin hurts yin and <b>blood stasis</b> . For <b>breast hyperplasia disease</b> caused by qi stagnation and blood stasis.	symptom symptome disease grouping
BiLSTM	Soften firmness and dissipate knots, <b>remove blood stasi and relieve pain</b> . In traditional Chinese medicine, heat toxin hurts yin and <b>blood stasis</b> . For <b>breast hyperplasia</b> disease caused by qi stagnation and blood stasis.	drug efficacy symptom disease
BERT-CRF	Soften firmness and dissipate knots, <b>remove blood stasi</b> and relieve pain. In traditional Chinese medicine, <b>heat toxin hurts yin and blood stasis</b> . For <b>breast hyperplasia</b> disease caused by qi stagnation and blood stasis.	symptom symptom disease
TENER-Rtransformer-CRF	Soften firmness and dissipate knots, <b>remove blood stasi</b> and relieve pain. In traditional Chinese medicine, heat toxin hurts yin and <b>blood stasis</b> . For <b>breast hyperplasia disease</b> caused by qi stagnation and blood stasis.	symptom symptom disease grouping
Ours	Soften firmness and dissipate knots, <b>remove blood stasi</b> and relieve pain. In traditional Chinese medicine, heat toxin hurts yin and <b>blood stasis</b> . For <b>breast hyperplasia disease</b> caused by qi stagnation and blood stasis.	symptom symptome disease grouping

## Abbreviations

BERT	Bidirectional Encoder Representations from Transformers
BiLSTM	Bidirectional long short-term memory
CCKS	China Conference on Knowledge Graph and Semantic Computing
CEMR	Chinese electronic medical record
CGN	Collaborative Graph Network
CMeEE	Chinese Medical Entity Extraction dataset
CNN	Convolutional neural network
CRF	Conditional random field
ELMo	Embeddings from Language Models
EMRs	Electronic medical records
F1	F1 Score
FGM	Fast gradient method
FLAT	Flat-Lattice Transformer
GAN	Generative Adversarial Network
GloVe	Global Vectors
IDCNN	Iterated Dilated convolutional neural network
K	Key
LD	Labeled data
LGN	Lexicon-Based Graph Neural Network
LR-CNN	CNN-Based Chinese NER with Lexicon Rethinking
LSTM	Long short-term memory
NER	Named Entity Recognition
NLP	Natural Language Processing
P	Precision
Q	Query
R	Recall
RoBERT	Robustly Optimized BERT
TENER	Transformer Encoder for Named Entity Recognition
ULD	Unlabeled data
V	Value
WC-LSTM	Word-Character long short-term memory
ZEN	Chinese (Z) text encoder Enhanced by N-gram representations

## Acknowledgements

We thank the anonymous reviewers for their insightful comments.

## Authors' contributions

Zhang Z and Peng H led the method application, experiment conduction and the result analysis. Liu D and Qin X participated in the data extraction and preprocessing. All authors provided theoretical guidance and the revision of this paper.

## Funding

This work is supported by the National Natural Science Foundation of China (No. 62163033), the Talent Innovation and Entrepreneurship Project of Lanzhou, China (No. 2021-RC-49), the Natural Science Foundation of Gansu Province, China (No.22JR5RA145, No.21JR7RA781, No.21JR7RA116), the Major Research Project Incubation Program of Northwest Normal University, China (No. NWNLU-LKZD2021-06). The funding body had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and materials

The datasets used and analyzed during the current study are available from the first author upon reasonable requests.

CCKS2019: <http://openkg.cn/dataset/yidu-s4k>,  
ERTCMM: <https://tianchi.aliyun.com/dataset/86819>,  
CMeEE: <https://tianchi.aliyun.com/dataset/95414>.

## Declarations

### Ethics approval and consent to participate

Our study adhered to the Declaration of Helsinki. The data constituting the CEMR dataset have been licensed by a tertiary hospital in Lanzhou, Gansu, China, and patients have agreed to use their electronic medical record data. Participants gave informed consent to participate in the study. The CEMR dataset was approved by the Medical Ethics Committee of the Second People's Hospital of Gansu Province.

## Consent for publication

Not applicable.

## Competing interests

The authors declare no competing interests.

Received: 30 June 2022 Accepted: 18 July 2023

Published online: 24 July 2023

## References

- Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.
- Yan H, Deng B, Li X, Qiu X. TENER: Adapting Transformer Encoder for Named Entity Recognition. arXiv preprint arXiv:1911.04474. 2019.
- Liu W, Wen Y, Yu Z, Li M, Raj B, Song L. SpheroFace: Deep hypersphere embedding for face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. p. 212–220.
- Friedman C, Alderson PO, Austin JHM, Cimino JJ, Johnson SB. A General Natural-language Text Processor for Clinical Radiology. *J Am Med Inform Assoc*. 1994;1(2):161–74.
- Gaizauskas R, Demetriou G, Humphreys K. Term recognition and classification in biological science journal articles. In: In Proc. of the Computational Terminology for Medical and Biological Applications Workshop of the 2nd International Conference on NLP. Citeseer; 2000.
- Zhang Y, Yang J. Chinese NER Using Lattice LSTM. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics; 2018. p. 1554–1564.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–80.
- Gui T, Ma R, Zhang Q, Zhao L, Jiang YG, Huang X. CNN-Based Chinese NER with Lexicon Rethinking. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence; 2019. p. 4982–4988.
- Sui D, Chen Y, Liu K, Zhao J, Liu S. Leverage lexical knowledge for Chinese named entity recognition via collaborative graph network. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), China: Association for Computational Linguistics; 2019. p. 3830–3840.
- Gui T, Zou Y, Zhang Q, Peng M, Fu J, Wei Z, Huang X J. A lexicon-based graph neural network for Chinese NER. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). 2019:1040-1050.
- Li X, Yan H, Qiu X, Huang X. FLAT: Chinese NER Using Flat-Lattice Transformer. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics; 2020. p. 6836–6842.
- Liu W, Xu T, Xu Q, Song J, Zu Y. An Encoding Strategy Based Word-Character LSTM for Chinese NER. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 2379–2389.
- Ding R, Xie P, Zhang X, Lu W, Li L, Si L. A Neural Multi-digraph Model for Chinese NER with Gazetteers. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 1462–1467.
- Ma R, Peng M, Zhang Q, Wei Z, Huang X. Simplify the Usage of Lexicon in Chinese NER. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics; 2020. p. 5951–5960.
- Shaw P, Uszkoreit J, Vaswani A. Self-Attention with Relative Position Representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics:

- Human Language Technologies, Volume 2 (Short Papers). New Orleans, Louisiana: Association for Computational Linguistics; 2018. p. 464–468.
16. Dai Z, Yang Z, Yang Y, Carbonell J, Le Q, Salakhutdinov R. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 2978–2988.
  17. Stubbs A, Kotfila C, Özlem Uzuner. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. *J Biomed Inform.* 2015;58:S11–9. Supplement: Proceedings of the 2014 i2b2/UTHealth Shared-Tasks and Workshop on Challenges in Natural Language Processing for Clinical Data.
  18. Zhang N, Chen M, Bi Z, Liang X, Li L, Shang X, et al. CBLUE: A Chinese Biomedical Language Understanding Evaluation Benchmark. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland: Association for Computational Linguistics; 2022. p. 7888–7915.
  19. Wu Y, Huang J, Xu C, Zheng H, Zhang L, Wan J. Research on Named Entity Recognition of Electronic Medical Records Based on RoBERTa and Radical-Level Feature. *Wireless Communications and Mobile Computing.* 2021;2021:1–10.
  20. Wan Q, Liu J, Wei L, Ji B. A self-attention based neural architecture for Chinese medical named entity recognition. *Math Biosci Eng.* 2020;17(4):3498–511.
  21. Li Y, Wang X, Hui L, Zou L, Li H, Xu L, et al. Chinese Clinical Named Entity Recognition in Electronic Medical Records: Development of a Lattice Long Short-Term Memory Model With Contextualized Character Representations. *JMIR Med Inform.* 2020;8(9):19848.
  22. Liu Z, Lin W, Shi Y, Zhao J. A Robustly Optimized BERT Pre-training Approach with Post-training. In: China National Conference on Chinese Computational Linguistics. Springer; 2021. p. 471–484.
  23. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942.* 2019.
  24. Diao S, Bai J, Song Y, Zhang T, Wang Y. ZEN: Pre-training Chinese Text Encoder Enhanced by N-gram Representations. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Online: Association for Computational Linguistics; 2020. p. 4729–4740.
  25. Cui Y, Che W, Liu T, Qin B, Wang S, Hu G. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Online: Association for Computational Linguistics; 2020. p. 657–668.
  26. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Iii HD, et al. Datasheets for datasets. *Commun ACM.* 2021;64(12):86–92.
  27. YiduCloud. [https://www.biendata.xyz/competition/ccks\\_2019\\_1/final-leaderboard/](https://www.biendata.xyz/competition/ccks_2019_1/final-leaderboard/). Accessed 29 June 2022.
  28. Tianchi Big Data Competition. <https://tianchi.aliyun.com/competition/entrance/531824/rankingList/1>. Accessed 29 June 2022.
  29. Chinese Biomedical Language Understanding Evaluation. <https://tianchi.aliyun.com/dataset/dataDetail?spm=5176.12282016.0.0.29486d92TsAmDP&datald=95414#4>. Accessed 29 June 2022.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

