

Order number	Author	Title	Publication year	Dataset size	Data type	Data number and source of TCM	Disease	System	Methods description (system/model/soft)	Method type	Supervised type	Core content of extracted TCM information	NLP tasks related to information extraction	Evaluation and Best Performance	Comparative Evaluation	Competitive Evaluation	Objective	Relevant outcomes	Challenges
4	Cao et al. 2014	A semi-supervised learning method for mining of traditional Chinese prescriptions and drugs	2014	CA	Ancient literature	518 medical records from "Classified Medical Records of Contemporary Prescriptions" and "Classified Medical Records of Contemporary Physicians' Prescriptions"	Not specified	None	Bootstrapping Approach to Recognize Names of Traditional Chinese Prescriptions and Drugs using Semantic Knowledge	Machine learning	Semi-supervised	Prescriptions, drugs	Entity extraction	F-Value of prescription and drug are 74.9% and 90.8% respectively	✓	Compared with human-computer interactions during recognition process	Recognition names of Traditional Chinese prescriptions and drugs in an annotated ancient medical features	In this paper, a method named Bootstrap, which uses simple context features instead of lexical and syntactic knowledge, was proposed. In the method, iteration process was simplified and regularities in regularity through by pre-processing of context were avoided. Furthermore, human-computer interaction is added to Bootstrap for generating a new framework, which while the semantic, DRG was improved	Due to the diversity of ancient medical literatures, low coverage rate of current semantic lexicons and the ambiguity of the lexicon words, low recall rate and low accuracy are realized by using only lexicons to recognize TCM terms.
10	Wang Y. et al. 2014	Segmented methods for symptoms name recognition in free-text clinical records of traditional Chinese medicine	2014	JA	Clinical records	11632 TCM clinical record collected by TCM doctors during their routine diagnostic work from April 2006 to June 2008	Not specified	None	MMH / MEMM / CRF	Machine learning	Supervised	Symptom	Entity extraction	The best F-measure achieved by CRF reaches 95.28 (Precision: 94.77% and Recall: 95.49%)	✓	Comparison among MMH, MEMM, CRF	Detect symptom name from free-text clinical record in machine learning methods	The domain-specific adaptation of sequence labeling requires a appropriate and effective, and the CRF outperforms MMH and MEMM for the CRF task	The methods that are verified on the well-measured data cannot be directly applied to knowledge discovery in the free-text clinical records. In this study, the authors proposed a method for TCM clinical records in time-consuming, and labor-intensive. Hence there is an urgent need for the development of an effective method to automatically detect clinical records
12	Jiang Q. et al. 2014	Free-text mining of TCM medical records	2014	CA	Clinical records	more than 600 free-text of TCM medical records	Not specified	None	CRF / MMH / MEMM	Machine learning	Supervised	Symptoms, signs, TCM diagnosis, Chinese medicines (drug) prescriptions, TCM syndrome type, etc.	Entity extraction	CRF: Precision=90.57%, Recall=85.05%, F-score=88.10%; MMH=84.29%, Recall=79.92%, F-score=82.86%	✓	Comparisons among CRF, MMH and MEMM	Aim to give an extracted method with text mining based on conditional random fields	CRF is more appropriate and precise than MMH, MEMM and MEMM for clinical terms recognition in TCM medical records	How to extract medical terms from TCM medical records based on conditional random fields.
13	Jiang Q. et al. 2014	Multi-combined features text mining of TCM clinical cases	2014	CA	Clinical records	about 600 TCM medical cases obtained from traditional Chinese medicine journal of Hospital of Guangdong Province	Not specified	None	CRF	Machine learning	Supervised	Symptoms or signs, TCM diagnosis, TCM syndrome type, Chinese medicines (drug), and names of TCM prescriptions	Entity extraction	With all features: P=89.14%, R=82.91%, F=86.39%	✓	Compared with different combined labeling features	In order to improve accuracy of recognitions, CRF model is used for terms recognition in TCM medical cases with multi-labeling features	How to extract the data and terms automatic, intelligently, and clearly into a database	
14	Wang J. et al. 2014	Relation extraction from Traditional Chinese Medicine journal publications	2014	CA	Journal publications	31 pieces of TCM journal publication	Not specified	None	Decision Tree / rules	Machine learning, rule-based method	Supervised	Effect relation, conditional effect relation	Relation extraction	Rule-based approach: F-score=48%, feature-based approach: F-score=41%	✓	Comparison between the rule-based method and the feature-based approach	Extract relations from traditional Chinese medicine journal publications	The rule-based approach avoids the rule-based approach which may be due to the small size of the dataset and the feature-based approach which may be due to the small size of the dataset. The feature-based approach and compare the relation extraction and classification tasks with the more Chinese-specific NLP problems, such as word segmentation, word co-occurrence analysis, etc.	How to extract the data and terms automatic, intelligently, and clearly into a database
15	Wan H. et al. 2014	Extracting relations from traditional Chinese medicine journal publications	2014	JA	Journal publications	301610 abstracts in published articles on TCM from a Chinese publication database	Not specified	None	Heterogeneous factor graph model (HFGM)	Machine learning	Semi-supervised	Herb-symptoms relation, herb-disease relation, formula-symptoms relation, formula-disease relation, syndrome-symptoms relation	Relation extraction	Precision=90.39%, Recall=86.09%, F1=88.58%	✓	Comparison with a traditional TSM classifier	Collectively and globally extract multiple types of relations from the entire corpus of Chinese medicine journal publications	The challenge of relation extraction from TCM data is the complexity of the TCM system. Multiple types of relations exist in TCM data, such as herb-symptoms, herb-disease, formula-symptoms, formula-disease, syndrome-symptoms, syndrome-disease, etc.	How to extract the data and terms automatic, intelligently, and clearly into a database
16	Liang J. et al. 2017	A Novel Approach towards Medical Entity Recognition in Chinese	2017	JA	Clinical records	872 extraction records from Second Affiliated Hospital Zhejiang University School of Medicine	Not specified	None	CRF	Machine learning	Supervised	TCM drug name	Entity extraction	Recognition of traditional Chinese medicine drug names: precision=94.2%, recall=92.81%, F-measure=93.5%	✓	Compared with the baseline system which used the maximum matching algorithm and professional drug dictionaries (dictionary-based method)	Proposed a new cascaded-type approach towards medical entity recognition in Chinese	The proposed approach avoids the side effect of the baseline system which used the maximum matching algorithm and professional drug dictionaries (dictionary-based method) and improves the recognition performance of drug named entities in the linguistic (Chinese) domain	Much less work has been done on medical entity recognition than that written in Chinese, or in the setting of differentiation of Chinese drug names between traditional Chinese medicine and Western medicine
19	Zhang H. et al. 2019	TCMR: A TCM Entity Recognition Framework	2019	CA	Web data	7500 TCM related entity data from BaiduBaidu and HupuBaidu	Not specified	TCM entity filter (TCMEF)	Short Text TCM Classifier (STTC)	Deep learning	Supervised	Patient name	Entity extraction	F1 score of TCMER=92.75%	✓	Compared with general word-based short text classification algorithms and Latent Dirichlet Allocation based model	Propose a three-phase filtering framework for TCM entities	The three-phase filtering framework for TCM entities uses the stroke features of the Chinese characters in the TCM text to filter the noise and improve the recognition performance of TCM entities. The proposed approach avoids the side effect of the baseline system which used the maximum matching algorithm and professional drug dictionaries (dictionary-based method) and improves the recognition performance of drug named entities in the linguistic (Chinese) domain	The entities in text style and character shape in TCM files are often shorter than 20, many of them are even shorter than 10 characters. It is hard for general short text classification algorithms to detect enough features for these files
21	Chen Y. et al. 2018	Knowledge Management in Traditional Chinese Medicine	2018	JA	Web data	China Food Composition and three Chinese food composition and three Chinese food composition	Not specified	None	CRF for the entity extraction, SVM, naive Bayes (NB), LSTM, and KNN for relation extraction	Machine learning, deep learning	Supervised	Entities: food material, dish, nutritional element, symptom, and crowd relation; "good for" or "bad for" relationship between "food material" and "symptom"; and "same" or "different" relationships between two "food material" entities	Entity extraction, relation extraction	concept extraction and relationship recognition were all above 80%	✓	Compared among SVM, NB, LSTM, KNN	To help people retrieve and learn healthy diet knowledge more efficiently and comprehensively, the paper designs a knowledge graph to integrate healthy diet information in the literature and provides a semantic retrieval system	The data regarding healthy diet available in the Internet is increasing rapidly and is distributed on multiple sources. It is time-consuming for users to learn about healthy diet on the Internet. They need to search data on multiple platforms, choose and integrate information, and then understand what they have learned	
22	Chen Y. et al. 2019	A general approach for recognizing deep learning-based medical entity recognition in traditional Chinese medicine	2019	JA	Journal publications	10810 abstracts from papers published in 114 traditional Chinese medicine journals	Not specified	None	Combine BERT with a one-dimensional convolutional neural network (LSTM) to extract the entity information from the pre-trained	Deep learning	Supervised	Herb-symptoms relation, herb-disease relation, formula-symptoms relation, formula-disease relation, syndrome-symptoms relation	Relation extraction	The 15-CHN fine-tuning approach: Precision=95.18%, Recall=90.41%, F1-measure=92.82%	✓	Compared with SVM, iterative SVM, HFGM, L1-CHN general embeddings, L1-CHN TCM embeddings, BERT with pre-trained model	Propose a pre-trained model and a fine-tuning technique to improve three approaches related to entity relation: consuming human labeling	The 15-CHN fine-tuning approach achieves the best performance on four of the nine relation types	Most of the current deep learning approaches for medical relation extraction require large-scale training data to pre-train embeddings of the words
23	Gong F. et al. 2019	On building a diabetes knowledge base for mining the health data	2019	JA	Web data	7700 researches and 90041 cases from eight prevalent chronic diseases on health care in 2019	Diabetes	None	Insurance matching algorithm	Dictionary-based method	Supervised	Traditional Chinese medicine	Entity extraction	The accuracy of overall knowledge base is 98.1%	✓	Compared with insurance matching with pre-processing	Constructing a high-quality knowledge base of a specific disease to support the disease detection tasks (like EHR-based detection)	Introduced an approach to constructing diabetes knowledge base	To support the medical professionals' diagnosis and analysis against a mass of labor cost to collect experimental data, which is also very poor. To reduce the cost and to ensure the data quality, there is a growing need of automatic relation extraction from large-scale health data
24	Jin L. et al. 2019	Named Entity Recognition in Traditional Chinese Medicine	2019	CA	Text book	75 modern TCM clinical cases and 14 ancient TCM clinical cases from two authoritative clinical case books and literatures in QCMET	Not specified	None	TCMRG-LSTM-CRF	Deep learning	Supervised	Medicine, dose, prescription, process, disease, symptom, syndrome, meridian, property/flavor and function	Entity extraction	F1 score=89.41%	✓	Compared with CRF, BiLSTM-CRF, BiLSTM-CRF Pretrained	These methods cannot effectively solve the problem of low recognition rate of new words, which is common in TCM field	Propose TCMRG-LSTM-CRF model that offers knowledge integration to strength the learning ability and recognize new words. The model integrates knowledge integration and knowledge graph construction and considers relations from context word. The experiment results show that the proposed model can effectively solve the problem of low recognition rate of new words, which is common in TCM field	How to effectively solve the problem of low recognition rate of new words, which is common in TCM field
26	Song B. et al. 2020	Incorporating Lexicon for Named Entity Recognition of Traditional Chinese Medicine Books	2020	CA	Ancient literature	376 TCM books from the Han Dynasty to the Qing Dynasty	Not specified	None	BiLSTM-CRF	Deep learning	Supervised	Symptom	Entity extraction	Precision=76.80%, Recall=72.11%, F1=74.28%	✓	Compared with MMH, CRF	In order to improve the performance of NER for TCM books	Lexicon information and large-scale of unlabeled corpus data are not fully exploited in the previous methods	
29	Feng J. et al. 2020	Research on Named Entity Recognition of Traditional Chinese Medicine	2020	CA	Clinical records	131 electronic medical records from the Hubei Provincial Hospital of Traditional Chinese Medicine	Not specified	None	BiLSTM-CRF, BiLSTM-CRF, BiLSTM-CRF	Deep learning	Supervised	Symptom	Entity extraction	The BERT model: Precision=89.94%, Recall=82.74%, F1=86.10%	✓	Compared with BiLSTM-CRF and BiLSTM-CRF	In order to explore the effective method of named entity recognition of TCM ER	The BERT model: Precision=89.94%, Recall=82.74%, F1=86.10%	(1) The naming entity of traditional Chinese medicine has complex characteristics and uncertain length. In addition, there are sub-entities in the entity. Therefore, the task of named entity recognition in traditional Chinese medicine is more complex and more difficult than that in general fields. (2) Character-based model can avoid word segmentation errors, but it can also lose the word information contained in the word sequence in the process
31	Lu L. et al. 2020	A semi-supervised approach for extracting TCM clinical terms from free-text records	2020	JA	Text book	Formulae of Chinese Medicine (2000 words)	Not specified	None	BiLSTM-CRF	Deep learning	Semi-supervised	Traditional Chinese medicine, symptoms, patterns, diseases and formulae	Entity extraction	F1 value=78.78%, Precision=77.58%, Recall=78.7%	✓	Compared with the supervised method	Propose a semi-supervised approach for extracting TCM clinical terms based on character vector	The method reduces the cost of manual labeling and improves the experiment results on the help of the free-text work	(1) Ancient Chinese language is inherently used in the TCM corpus which brings difficulties to TCM NER research. (2) Manual labeling of supervised methods are costly and labor-intensive
32	Zhang M. et al. 2020	Traditional Chinese Medicine Knowledge Service based on Semi-Supervised BERT-BiLSTM-CRF Model	2020	CA	Ancient literature	A total of 32614 characters from "Treatise on Febrile Diseases"	Not specified	None	BERT-BiLSTM-CRF	Deep learning	Semi-supervised	Symptoms, syndromes, treatment, Chinese medicine, prescriptions, pulse, tongue, efficacy	Entity extraction	Accuracy=81.24%, Recall=80.94%, F1 value=81.04%	✓	Compared with CRF, BiLSTM-CRF	This paper proposes a semi-supervised method to extract entities from TCM texts	The experiment shows that TCM entity recognition accuracy of the embedded Semi-BERT-BiLSTM-CRF model reaches 81.24%, which effectively improves the TCM entity recognition accuracy and reduces the manual labeling work. The results of this research can be applied to scenarios such as auxiliary diagnosis of TCM and expert system after subsequent improvement and refinement	The existing methods are not accurate enough to solve TCM named entity recognition and require a lot of manual labeling data
34	Qin C. et al. 2020	Named Entity Recognition of TCM Text Based on BiLSTM Model	2020	CA	Ancient literature	A total of 32614 characters from "Treatise on Febrile Diseases"	Not specified	None	BERT-BiLSTM-CRF	Deep learning	Supervised	Symptoms, disease, names, text, prescription names, and drug names	Entity extraction	The best recognition effect of drug, with an F-value as high as 88.79%	✓	Compared with BiLSTM-CRF, LSTM-CRF	Aiming at the problem of fuzzy entity recognition and less labeled data in the field of traditional Chinese medicine, a named entity recognition model based on BiLSTM-CRF is constructed and tested on the cross-validation data set	By constructing the word vector through the fast model and then using it as a BiLSTM-CRF, both the accuracy rate, the recall rate, and the F1 value are improved	The problem of fuzzy entity recognition and less labeled data in the field of TCM
35	Wan H. et al. 2020	A temporal semantic network system for traditional Chinese medicine based on temporal knowledge graphs	2020	CA	Clinical records	More than 100,000 cases abstracts from a Chinese publication database	Not specified	None	Multi-view graph model to extract relation (MVGR)	Deep learning	Supervised	Herb-symptoms relation, herb-disease relation, formula-symptoms relation, formula-disease relation, syndrome-symptoms relation	Relation extraction	Precision=80.38%, Recall=81.42%, F1 score=84.07%	✓	Compared with SVM-C, PTM, Adabo, BiLSTM	Extract relations from Chinese clinical records	Very limited related information system has been constructed on clinical records written in Chinese, especially TCM clinical records	
37	Tang C. et al. 2020	Implementing a temporal semantic network system for traditional Chinese medicine based on temporal knowledge graphs	2020	CA	Clinical records	Classic Chinese Medicine Clinical Case Library	Not specified	None	Temporal facts	Rule-based method	Supervised	Temporal facts	Relation extraction	Average Precision=0.825	✓	Compared with off-the-shelf entity linking	Implement a temporal semantic network system	Propose a new model labeled for TCM based on the temporal knowledge graph in this paper	Current TCM-based retrieval systems cannot effectively deal with the temporal relations of search sentences, which leads to bad experiences for users in retrieval scenarios. Most of temporal information are ignored in the knowledge map construction
39	Zhang D. et al. 2020	Named Entity Recognition for Traditional Chinese Medicine Text using a Novel Back-Labeling Approach	2020	JA	Ancient literature	Chinese medical records of all famous records (10000 words)	Not specified	None	BiLSTM	Deep learning	Distantly supervised	Chinese medicine, symptom, medicine prescription, dose, tongue, pulse, pulse	Entity extraction	Precision=73.06%, Recall=68.79%, F1-measure=69.76%	✓	Compared with BiLSTM-CRF and dictionary matching method	Propose a method for entity recognition based on distant supervision, and complete the entity recognition task under the premise of only using the entity vocabulary related to the TCM field	The scheme can effectively improve the entity recognition on the basis of distant supervision	The limitation of using machine learning methods, which is need of a large amount of manually labeled data
43	Zhou S. et al. 2021	Feature engineering in deep learning for traditional Chinese medicine identification: Toward applications using a Novel Back-Labeling Approach	2021	JA	Journal publications	317 papers with 25007 sentences from human-curated Chinese literature resources derived from CNKI	Asthma	None	Structural Redefinition Long Short-Term Memory (SLSTM)	Deep learning	Supervised	Subjects, methods, results	Section extraction	SLSTM model achieve close to 90% performance in precision, recall, and F1-measures	✓	Compared with CRF, BiLSTM, HMM, LSTM	Studied the paper section identification problem in the context of Chinese medical literature analysis, where the subjects, methods, and results are more valuable from a physician's perspective	Design a novel deep learning model, the Structural Redefinition BiLSTM model, which outperforms the traditional machine learning methods and other deep learning methods	Due to natural language differences, the extracted features from English texts cannot be directly applied onto Chinese medical literature. Moreover, Chinese medical literature is often less structured than English literature, which makes the problem more difficult to tackle
44	Bai T. et al. 2021	Traditional Chinese medicine entity relation extraction based on CNN with segment attention	2021	JA	Web data	Annotated dataset with 4322 samples of herb-disease and 4666 samples of herbs chemicals combination extracted from PubMed	Not specified	None	SEGAT-CNN combined with segment attention	Deep learning	Supervised	Herb-disease relation, herb-disease relation, herb-disease relation	Entity extraction, relation extraction	Herb-disease relation: SVM combined SEGAT-CNN: P=93.43%, R=93.52%, F1=93.43; Herb-disease relation: SVM combined SEGAT-CNN: P=93.43%, R=93.52%, F1=93.43	✓	Compared with SVM, KNN, LR, RF	In order to extract Chinese medicine-specific entity relations more reliably and effectively from TCM literature	Propose a novel architecture with an improved layer for entity relation extraction. The method has an absolute advantage and is proved to be effective in solving the problem of herbal related relation extraction	(1) Although the deep learning technology has been applied to the relation extraction task with superlatives, recent works are often adapted to TCM text. (2) Language gap exists between TCM text from researches. (3) Classical machine learning algorithms improve the accuracy of relation extraction primarily through complicated manual feature engineering, which constrains the potential of accuracy improvement for classical machine learning
45	Ha D. et al. 2021	Named Entity Recognition of Traditional Chinese Medicine Patterns Based on BiLSTM-CRF	2021	JA	Patient tests	A total of 2000 TCM patients' abstract texts from intellectual property websites	Not specified	None	BiLSTM-CRF	Deep learning	Supervised	Herb name, disease name, symptom, therapeutic effect, property	Entity extraction	precision rate is 94.33%, recall rate is 94.47%, and F1 value is 94.48%	✓	Compared with HMM, LSTM, BiLSTM	The study aims to design a model to fill the technical gap in NER of TCM patients	A method combining Bidirectional Long Short-Term Memory neural network with Conditional Random Field (BiLSTM-CRF) is proposed to automatically recognize entities of herbs, herb names, disease names, symptoms, and therapeutic effects from the abstract texts of TCM patients' abstract texts. The experimental results show that the proposed model can effectively solve the problem of low recognition rate of new words, which is common in TCM field	(1) Due to the long history and wide geographical distribution of herbs. (2) Due to the different writing habits, TCM doctors and researchers tend to express a similar disease name in language similar with classical Chinese. (3) Due to the lack of labeled training data, TCM patients' tests, a proper training set has to be established from only unlabeled data
46	Jin C. et al. 2021	Detection of Traditional Chinese Medicine Entity Recognition Based on BiLSTM-CRF	2021	JA	Clinical records	Case book of TCM compiled by the expert team of Professor Lu Zhaolin	Not specified	None	Span-level distant supervised named entity recognition (NER)	Deep learning	Distantly supervised	Symptom, medicine, prescriptions, dose, tongue, pulse, pulse	Entity extraction	Precision=78.78%, Recall=75.52, F1 score=77.24	✓	Compared with distant-LSTM-CRF, BiLSTM-CRF, BiLSTM-CRF	The idea is to utilize distant supervised method to conduct the named entity recognition (NER) task	Propose a span-level distant supervised named entity recognition (NER) approach to extract TCM-related entity. It utilizes the pre-trained language model, a simple multi-layer neural network, and a span-level distant supervised method to extract the named entity from the span-level model. The strategy randomly extracts negative samples in every epoch and then the possible loss negative sample proportion. It can be used to reduce the loss of the model	(1) Most of TCM clinical records are unstructured text. (2) The names in free-style and difficult to understand for modern medical practitioners. (3) The model of manually structuring and maintaining free-text clinical records is very expensive. (4) Lack of gold standard datasets for supervised learning methods
47	Zheng T. et al. 2021	Chinese medical entity relation extraction based on multi-head self-attention mechanism	2021	JA	Journal publications	4000 Chinese medical literature abstracts	Not specified	None	Multi-head self-attention mechanism-BiLSTM	Deep learning	Supervised	Therapeutic relation, causal relation	Relation extraction	Therapeutic relation: Precision=88.65, Recall=88.22, F1 score=89.19; Causal relation: Precision=89.23, Recall=79.53, F1 score=84.47	✓	Compared with CNN, Bi-LSTM, Bi-LSTM+Self-attention, Bi-LSTM+Multi-head attention	Propose an attention-based model to extract the multi-aspect semantic information for the Chinese medical entity relation by multi-head self-attention mechanism	The model could generate multiple weight vectors for the sentence through each attention layer. Therefore, we can generate the different semantic representation of a sentence, respectively. Experimental results show that the proposed model can effectively solve the problem of low recognition rate of new words, which is common in TCM field	(1) Complex semantic information in the sentence determines the relation between entities, the semantic information cannot be represented by one sentence vector. (2) Machine learning methods rely on manual selection of features, which is laborious and time-consuming
48	Xu H. et al. 2021	A Named Entity Recognition Model for Traditional Chinese Medicine	2021	CA	Clinical records	4000 sentences from 4000 Chinese medical records	Not specified	None	Named NER model based on LSTM-CRF	Deep learning	Supervised	Medicine, symptom, pulse, tongue, medicine prescription, dose, disease location, onset time and duration, severity, color, quantity, frequency	Entity extraction	Precision=85.64, Recall=86.11, F1 score=85.81	✓	Compared with LSTM-CRF, Attention hypergraph, Span-based model	Propose a named entity recognition method for traditional Chinese medicine records	This paper can word character-level embeddings to enable the model to better learn semantic and syntactic information. Two-layer labeling strategy allows the model to better extract hierarchical information. This work provides a good named NER method for TCM records, which can be used to extract the named entity from TCM records	Entity boundary problems caused by the difference between words and characters in TCM corpus, many methods that have good performance on the English datasets are not suitable for the NER task in TCM field
49	Guan Y. et al. 2021	A Traditional Chinese Medicine Terminology Recognition Model Based on Ontology Learning	2021	CA	Ancient literature	TCM prescription in 517 ancient TCM books	Not specified	None	BERT-BiLSTM-CRF	Deep learning	Supervised	Disease, pattern, symptom	Entity extraction	Precision=86.54, Recall=88.58, F1 score=87.55, when the size of training set is 10000	✓	Compared with BiLSTM, BiLSTM+CRF, BiLSTM+CRF	Presented a traditional Chinese medicine terminology recognition model named BERT-BiLSTM-CRF	Transferring a BERT model trained on other large-scale Chinese corpus as the pre-training model benefits the TCM terminology recognition, and CRF plays an important role in modeling the contextual semantic features of sequences in TCM term recognition	(1) There are some characteristics in Chinese text, such as fuzzy word boundary and lack of explicit morphological features of entities, which increase the problem of NER model for Chinese text. (2) How to apply the most advanced deep learning model to the task of TCM term recognition