

Article

# A Neural N-Gram-Based Classifier for Chinese Clinical Named Entity Recognition

Ching-Sheng Lin <sup>1,\*</sup>, Jung-Sing Jwo <sup>1,2</sup> and Cheng-Hsiung Lee <sup>1</sup> 

<sup>1</sup> Master Program of Digital Innovation, Tunghai University, Taichung 40704, Taiwan; jwo@thu.edu.tw (J.-S.J.); hsiung@thu.edu.tw (C.-H.L.)

<sup>2</sup> Department of Computer Science, Tunghai University, Taichung 40704, Taiwan

\* Correspondence: cslin612@thu.edu.tw

**Abstract:** Clinical Named Entity Recognition (CNER) focuses on locating named entities in electronic medical records (EMRs) and the obtained results play an important role in the development of intelligent biomedical systems. In addition to the research in alphabetic languages, the study of non-alphabetic languages has attracted considerable attention as well. In this paper, a neural model is proposed to address the extraction of entities from EMRs written in Chinese. To avoid erroneous noise being caused by the Chinese word segmentation, we employ the character embeddings as the only feature without extra resources. In our model, concatenated n-gram character embeddings are used to represent the context semantics. The self-attention mechanism is then applied to model long-range dependencies of embeddings. The concatenation of the new representations obtained by the attention module is taken as the input to bidirectional long short-term memory (BiLSTM), followed by a conditional random field (CRF) layer to extract entities. The empirical study is conducted on the CCKS-2017 Shared Task 2 dataset to evaluate our method and the experimental results show that our model outperforms other approaches.

**Keywords:** clinical named entity recognition; n-gram character embeddings; self-attention; BiLSTM



**Citation:** Lin, C.-S.; Jwo, J.-S.; Lee, C.-H. A Neural N-Gram-Based Classifier for Chinese Clinical Named Entity Recognition. *Appl. Sci.* **2021**, *11*, 8682. <https://doi.org/10.3390/app11188682>

Academic Editors: Julian Szymanski, Andrzej Sobocki, Higinio Mora and Doina Logofătu

Received: 7 July 2021

Accepted: 13 September 2021

Published: 17 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the rapid development of information technology, medical institutions have widely used electronic medical records (EMRs) to facilitate data collection which includes patient health information, diagnostic tests, procedures performed and clinical decision making. EMRs contain valuable clinical data and a large amount of medical information of patients which can have critical implications for future health care delivery. However, most of the EMRs are in the unstructured format which is difficult to extract to build intelligent biomedical systems and, most importantly, can hinder large-scale knowledge discovery. Therefore, it is urgent to explore effective approaches to convert EMRs into structured forms for improving the quality-of-care delivery.

The task of Information Extraction (IE) refers to identify and recognize the instances of the structured semantics (e.g., predefined classes of entities and relationships among entities) from unstructured or semi-structured text [1]. The continued expansion of EMRs has attracted researchers' interests and led to an active research topic called Biomedical Information Extraction (BioIE). BioIE aims to discover structured information from unstructured clinical notes and narratives that can be used by clinicians, researchers and applications. In general, there are three main subtasks in the BioIE: (1) Named Entity Recognition (NER) which aims to categorize entity names in clinical and biomedical domains, (2) Relation Extraction (RE) which targets the detection of semantic relations between entities and (3) Event Extraction (EE) which explores a more detailed alternative to produce a formal representation to extract the knowledge within the targeted documents [2]. BioIE is a hot and active research topic at the crossroads of Chemistry, Biology, Medicine, Computational Engineering and Natural Language Processing. A growing number of

workshops and conferences are testimony to its continuing importance and potential. IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM) and International Workshop on Health Text Mining and Information Analysis (LOUHI) are held to provide interdisciplinary forums for researchers interested in automated processing and information extraction of health documents. Two distinguished computational linguistics conferences, ACL and COLING, and affiliated workshops have always considered information extraction and clinical text modeling as topics of interest [3].

NER occupies a very important position in the field of natural language processing (NLP) and has been largely studied for decades especially in the general domain like news articles. It focuses on identifying and classifying all mentions of a subset of nouns such as places, persons, and organizations [4]. NER plays an essential role as a pre-processing step for downstream tasks which include question answering, information retrieval and relation extraction. With experiencing an ever-increasing explosion of EMRs, clinical named entity recognition (CNER) is of interest for researchers and many competitions have been organized in order to promote the development and stimulate the community [5]. Unlike the NER in the general domain, CNER, which consists of a large number of clinical terms and professional designations, aims to recognize entities in EMR (such as disease, symptom and body part) and benefit other intelligent clinical systems for health monitoring, prognosis, diagnostics, and treatment. In addition to the research of CNER in English, other languages have also gained prominence and Chinese is one of the core research topics. Compared with the CNER of alphabetic languages represented by English, Chinese CNER faces the following challenges [6–8]:

- The ambiguous chunk of text corresponds to the same character sequence but to different named entities. For example, “泌尿道感染” (urinary tract infection) could refer to the disease entity or symptom entity depending on the context;
- There are no clear word boundaries in Chinese text and the effect of word segmentation will significantly impact the performance of the NER. For example, “小腸切除術” (small bowel resection) is a treatment entity if it is considered as one segmentation unit. However, if the word segmentation model splits it into “小腸” (small intestine) and “切除術” (resection), their entity types will become body part and treatment, respectively;
- Because of the casual use of Chinese abbreviations for clinical entities written by doctors, it may result in multiple expressions of the same entity. For example, “盲腸炎” and “闌尾炎” could all refer to appendicitis.

In this paper, we propose an n-gram based neural network to model the Chinese CNER task as a sequence labelling problem. Given a sentence, we represent text at character level to avoid erroneous noise being caused by the Chinese word segmentation technique. We employ the character embeddings as the only feature. More specifically, adjacent character embeddings are integrated into n-gram features (unigrams, bigrams and trigrams). They are then given to a self-attention mechanism to learn long-term dependencies. Finally, a Bidirectional Long Short-Term Memory (BiLSTM) is applied to encode the sequential structure and capture contextual features, followed by a Conditional Random Field (CRF) layer to consider the correlations between adjacent tags for predicting the label sequence.

There are two main contributions in this paper:

- We propose an Att-BiLSTM-CRF model to perform the Chinese CNER task based on combinations of n-gram character embeddings of different lengths without using external knowledge. Unlike other approaches in the literature which rely on domain-specific resources and may limit the ability of generalization, our model will be scalable to other datasets.
- We assess the effectiveness of the proposed model on the CCKS-2017 Shared Task 2 dataset. Our model obtains an F-score of 89.33% and performs better than other competitive methods including CNN, BiLSTM and BERT based models which have F-scores in the range 87.75% to 88.51%.

The remainder of this paper is organized as follows. Section 2 reviews several techniques related to the work of this paper. The proposed model is described in Section 3. We explain the experimental setup and report the results in Section 4. In Section 5, we present conclusions and also discuss future research avenues.

## 2. Related Work

Since the growth of EMRs has increased considerably over recent decades, the CNER problem has drawn much interest and a great deal of research effort has been devoted to the study and development as well. Basically, four representative types of methods are proposed to perform the task which include rule-based, dictionary-based, machine learning and deep learning approaches [9–11].

In the early stage, rule-based approaches were the dominant approaches to solve the CNER problem. The adoption of heuristic information, handcrafted features [12,13] and lexical resources [14,15] is designed to detect clinical entities. Although playing a critical role before, rule-based approaches heavily require expert domain knowledge, resulting in being difficult to transfer to different fields.

Traditionally, dictionary-based methods take advantage of existing clinical vocabularies to extract entities and were widely applied due to their simplicity [15,16]. Several clinical ontologies and vocabularies, such as MeSH [17] and SNOMED-CT [18], have been proposed. However, the performance is limited by the size of the lexicon and could lead to low recall when the input data contains a high number of out-of-dictionary entities. Additionally, similar to the rule-based approaches, dictionary-based approaches also lack generalizability and require tremendous human effort to build the lexicons.

Since machine learning methods have been successfully used for sequence labelling tasks such as POS tagging, NER and chunking, the CNER task has also been transformed into a sequence labelling problem and is solved by various machine learning algorithms. Typically, feature engineering is performed on the input sentence to convert the data into numerical representation. Three typical supervised sequence tagging models (HMM, MEMM, and CRF) based on n-gram and position features are evaluated for the name recognition in traditional Chinese clinical records where CRF achieves better performance than the other two classifiers [19]. Support Vector Machine (SVM) with word shape and part-of-speech features is applied to recognize biomedical named entity and obtains a precision rate of 84.24% and a recall rate of 80.76% [20]. However, most of these machine learning methods rely on pre-defined features (such as lexical, syntactic, and semantic features) and are difficult to generalize to different datasets.

In recent years, since deep learning techniques are growing fast and have achieved significant success across various applications, the prevailing approaches have shifted to the employment of deep learning methods. The Long Short-Term Memory (LSTM) network is suitable for learning temporal relations and has been widely used in NLP tasks [21]. A BiLSTM-CRF approach, which is a neural network system based on bidirectional LSTMs and CRF, is proposed to solve the Chinese CNER problem using specialized word embeddings as feature representations and the external health domain lexicons as the knowledge base [22]. The system reports 87.95% in F-score on the CCKS-2017 (task 2) CNER dataset. Another bidirectional RNN-CRF model for recognizing Chinese CNER adopts concatenated n-gram embeddings and also includes word segmentation information, part-of-speech tagging and medical entity vocabulary as additional features [23]. Unlike the previous research relying on other miscellaneous information, in this paper, we present a neural n-gram based classifier without external resources.

## 3. The Proposed Approach

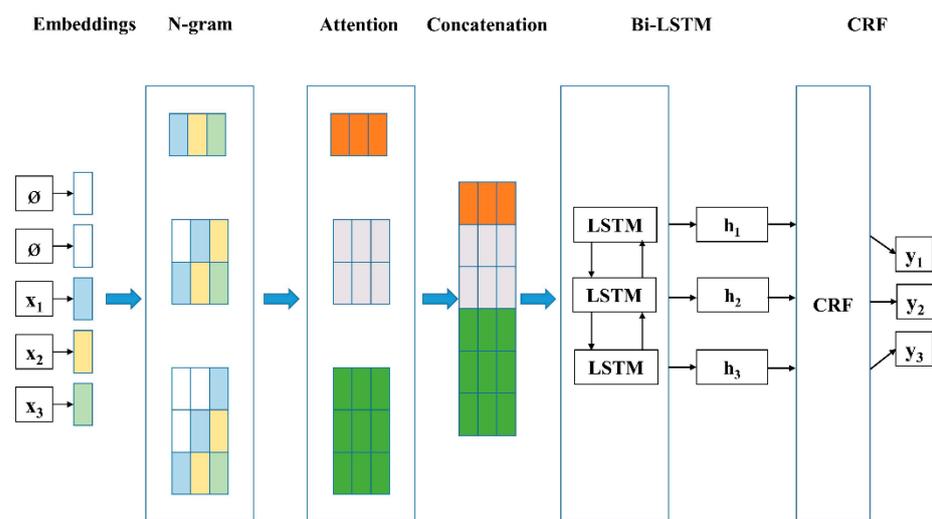
The Chinese CNER task is modeled as a sequence labelling problem in this work. Given an input sequence  $X$  with  $t$  characters (i.e.,  $X = (x_1, x_2, \dots, x_t)$ ), the goal is to label each character  $x_i$  with a predefined tag based on the tagging scheme to obtain an output sequence  $Y = (y_1, y_2, \dots, y_t)$ . We use BIOES as the annotation strategy where B denotes the

beginning of an entity, I denotes the middle of an entity and O denotes not an entity. In addition, the B and I tags need to be followed by an entity type such as B-BODY and I-BODY for the Body entity type. A tagging result of an input sentence “左側髖部正常” (the left hip is normal) is displayed in Table 1.

**Table 1.** An entity tagging example of “左側髖部正常”.

	左	側	髖	部	正	常
BIO	B-BODY	I-BODY	I-BODY	I-BODY	O	O

The proposed Att-BiLSTM-CRF model shown in Figure 1 is composed of six building blocks: Embeddings, N-gram, Attention, Concatenation, BiLSTM and CRF layers. The Embeddings Layer converts each input character into an embeddings vector and the N-gram Layer applies n-gram techniques on embeddings to form n-gram embeddings (n from 1 to 3). The Attention Layer employs self-attention on n-gram embeddings and the Concatenation Layer combines the self-attention representations. The BiLSTM Layer captures the features of the previous concatenated results and then the CRF Layer takes the BiLSTM Layer output to decode the tag sequence. The method of n-gram character embeddings and details of the neural entity recognition model will be discussed in the following sections.



**Figure 1.** The system architecture of our approach.

### 3.1. N-Gram Character Embeddings

In an n-gram model which is a widely used concept in the NLP field, each sentence is represented by a sequence of n consecutive units. To reduce the ambiguity of segmentation for Chinese words, we use the character as the basic unit rather than a word. For example, given an input Chinese sentence “胃部疼痛” (stomach-ache), the unigram is {胃, 部, 疼, 痛}, the bigram is {胃部, 部疼, 疼痛} and the trigram is {胃部疼, 部疼痛}.

For a sentence  $X = (x_1, x_2 \dots, x_t)$  with t characters, the embedding process is designed to transform each character into a distributed and dense vector representation  $R^d$ , where d is the size of the character embedding. The Embeddings Layer, a part of the neural network, is initialized with random vectors and learns to represent all the characters in the training set during the training stage. Each character will be mapped to an embeddings vector once the training is completed. To better encode the input sentence, we use the n-gram character embeddings model rather than the n-gram character model. An n-gram character embeddings model is represented by concatenating the embeddings of n characters. As seen in the N-gram Layer of Figure 1, we show unigram, bigram and trigram character embeddings.

### 3.2. Neural Entity Recognition Model

In this section, we discuss the proposed approach to deal with Chinese CNER in the EMRs. The neural model adopted in this research mainly relies on Attention, BiLSTM and CRF layers to obtain a more semantic representation of Chinese characters.

**Attention Layer:** The attention method has been widely used in many tasks, especially NLP applications to capture the context information and dependency between tokens for the given sentence [24,25]. The mechanism computes attention weights between every two tokens and uses a summation operation to obtain the representation [26]. The calculation of attention on our n-gram character embeddings is described as follows. Given an input sequence  $E_n = (e_1, e_2 \dots, e_t)$  where  $e_i \in R^{1 \times dn}$  is the n-gram character embeddings ( $n$  from 1 to 3) and  $d_n$  is the size of the embeddings,  $E_n$  is converted to query  $Q_n$ , key  $K_n$  and value  $V_n$  through linear transformations by the following expression:

$$Q_n, K_n, V_n = E_n W_q, E_n W_k, E_n W_v \quad (1)$$

where  $W_q, W_k, W_v$  are learnable parameters, and  $Q_n \in R^{t \times dn}, K_n \in R^{t \times dn}, V_n \in R^{t \times dn}$ . The attention score is, then, calculated as follows:

$$A_n(Q_n, K_n, V_n) = \text{softmax}\left(\frac{Q_n K_n^T}{\sqrt{d_n}}\right) V_n \quad (2)$$

In this paper, we use a special kind of attention, self-attention, to learn the feature of one unit in a sentence by attending to all units within the same sentence. In the self-attention mechanism, the query  $Q_n$ , key  $K_n$  and value  $V_n$  are all the same.

**BiLSTM Layer:** After the above self-attention calculation, we concatenate  $A_1, A_2$  and  $A_3$  to obtain the final embedding matrix  $A \in R^{t \times (d1+d2+d3)}$  as shown in the Concatenation Layer of Figure 1. We then pass  $A$  into a BiLSTM layer. The LSTM is proposed to capture long-term dependencies by introducing gated memory units to address the gradient problems and control the information flow [27]. At each timestamp  $t$  for the given input  $A = (a_1, a_2 \dots, a_t)$ , the LSTM updates its hidden state  $h_t$  based on the current input  $a_t$  and the previous hidden state  $h_{t-1}$  by computing the following equations:

$$i_t = \text{sigmoid}(W_i[a_t, h_{t-1}]) \quad (3)$$

$$f_t = \text{sigmoid}(W_f[a_t, h_{t-1}]) \quad (4)$$

$$o_t = \text{sigmoid}(W_o[a_t, h_{t-1}]) \quad (5)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_g[a_t, h_{t-1}]) \quad (6)$$

$$h_t = o_t \tanh(c_t) \quad (7)$$

where  $i_t, f_t, o_t$  and  $c_t$  are input gate, forget gate, output gate and cell vector.  $W_i, W_f, W_o$  and  $W_g$  are the corresponding weight matrices. Since the output of each time step in the LSTM only considers the previous state, BiLSTM is carried out to use both forward and backward information. The output is created by concatenating the two hidden vectors from both directions as  $h_t = [h_t^f; h_t^b]$ .

**CRF Layer:** In the CNER task, there exist several constraints and dependencies in the BIO tagging scheme. For instance, the I tag must follow the B tag. It is, therefore, important to take the above factors into consideration and we adopt CRF to predict a label sequence by learning the correlations between the current label and its neighbors [28]. Given the input sequence obtained from the output of the BiLSTM Layer  $h = (h_1, h_2 \dots, h_t)$ , we use  $y = (y_1, y_2 \dots, y_t)$  to represent a sequence of labels for  $h$ . The CRF model defines

the conditional probability distribution over all label sequences  $y$  given  $h$  and uses the following equation [29]:

$$p(y|h; W, b) \propto \exp\left(\sum_{i=1}^t W_{y_{i-1}, y_i}^T h_i + b_{y_{i-1}, y_i}\right) \quad (8)$$

where  $W$  denotes the weight and  $b$  the bias term corresponding to the neighboring pair  $(y_{i-1}, y_i)$ . To train a CRF model for a given training dataset  $\{h^{(i)}, y^{(i)}\}$  where superscript  $i$  represents the  $i$ -th data, the parameter estimation is performed by maximizing the conditional log-likelihood below:

$$(W^*, b^*) = \underset{W, b}{\operatorname{argmax}} \sum_{j=1}^t \log p(y^{(j)}|h^{(j)}; W, b) \quad (9)$$

In the process of inference, the optimal output label sequence  $y^*$  of a test input  $z$  is derived based on the maximization of the conditional probability by the Viterbi algorithm [30]:

$$y^* = \underset{y}{\operatorname{argmax}} p(y|z; W, b) \quad (10)$$

## 4. Experiments

### 4.1. Dataset and Evaluation Metrics

We conduct the empirical evaluation on CCKS-2017 Shared Task 2 benchmark dataset [31]. This dataset contains 400 EMRs in total where 300 of them are used as the training set and the remaining 100 are regarded as the testing set. Each EMR has four sections which are general items, medical history, diagnosis and treatment and discharge summary. There are five categories of clinical entities including body, exam, disease, symptom and treatment. Table 2 lists the statistics of clinical named entities for each category. There are 29,866 entities used for training and 9493 entities for testing.

**Table 2.** Statistics of named entities for training/testing sets on CCKS 2017.

	Training Set	Testing Set
Body	10,719	3021
Exam	9546	3143
Disease	722	553
Symptom	7831	2311
Treatment	1048	465
Total	29,866	9493

In this research, we use the character-level “BIO” annotation mode where “B” means that the character is at the beginning of an entity, “I” means that the character is at the middle of an entity and “O” means that the character does not belong to any entity. Since there are 5 clinical entity categories on CCKS 2017 dataset, this will result in ten annotation labels and an “O” label, yielding eleven labels in total.

The evaluation measures for entity recognition are three standard performance indicators, namely Precision (P), Recall (R) and F-score (F). Precision determines how capable the proposed method is for predicting entity categories, while Recall reflects how well it is for retrieving entity categories. F-score is defined to be the harmonic mean of Precision and Recall as an overall measure. The calculation formulas of the three metrics are defined as follows:

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \quad (12)$$

$$F = \frac{2 \times P \times R}{P + R} \quad (13)$$

where TP is the true positive, FP is the false positive and FN is the false negative.

#### 4.2. Experiment and Results

To study the effectiveness of the proposed model, two experiments are conducted. The first one evaluates the model against other competitive algorithms. In the second experiment, we test our model on different lengths of n-gram character embeddings. The experiments are carried out on the Windows system with an Intel(R) Core(TM) i7-8750H CPU, 8 GB RAM, and NVIDIA GeForce GTX 1050 Ti GPU. The neural network model is composed of one character embedding lookup table, a self-attention layer, a BiLSTM layer and a final CRF layer. The hyper-parameter setting of the model is shown in Table 3.

**Table 3.** Hyper-parameter settings of the proposed approach.

Parameter	Value
n-gram	1, 2, 3
character embedding size	100
LSTM hidden units	100
batch size	16
dropout rate	0.5
learning rate	0.001

To verify the proposed approach, we conduct the first experiment and compare our results with the following methods:

- I. LSTM-CRF: A LSTM neural network model with a CRF layer.
- II. BiLSTM-CRF: A bidirectional LSTM model with a CRF layer [32].
- III. RD-CNN-CRF: A residual dilated Convolutional Neural Network with CRF where dictionary features are utilized according to the drug information in Shanghai Shuguang Hospital and some medical literature [33].
- IV. ID-CNN-CRF: A Convolutional Neural Network-based model with iterated dilated convolutions and a domain-specific lexicon for word embeddings matching [34,35].
- V. BERT-BiLSTM-CRF: A pre-trained language model BERT to enhance the semantic representation, a BiLSTM network and a CRF layer [36].

The comparison results are demonstrated in Table 4. Our model obtains the best F-score among all competitors. We push the F-score to 89.33% and outperform the second-best system (RD-CNN-CRF) by 0.82%. In general, LSTM based approaches have better results in the Recall metric while CNN based methods perform better in the Precision score. RD-CNN-CRF achieves the best Precision score (88.64%) and our approach is the second-best (88.53%). The best Recall score is reported by BERT-BiLSTM-CRF (90.48%) and our system is ranked second (90.13%).

**Table 4.** Performance comparison results of each model.

Models	P	R	F
LSTM-CRF	83.59	85.28	84.42
BiLSTM-CRF	88.22	88.53	88.37
RD-CNN-CRF	<b>88.64</b>	88.38	88.51
ID-CNN-CRF	88.30	87.21	87.75
BERT-BiLSTM-CRF	86.50	<b>90.48</b>	88.45
Our model	88.53	90.13	<b>89.33</b>

We perform the second experiment to further investigate the effect of different lengths of n-gram embeddings. Table 5 demonstrates the experimental results of 1-g, 2-g, 3-g and our approach which combines all of them. Our model achieves the highest F-score of 89.33% and Precision of 88.53%. In terms of Recall, the 2-g method is the best (90.47%) and our model achieves the second-highest score (90.13%).

**Table 5.** Performance comparison between different n-gram lengths with our model.

Models	P	R	F
1-g	87.88	89.98	88.92
2-g	87.30	<b>90.47</b>	88.86
3-g	87.21	89.84	88.50
Our model	<b>88.53</b>	90.13	<b>89.33</b>

In addition to the overall performance evaluation discussed above, we also show the detailed results for all five entity categories in Table 6. Our model achieves the highest F-scores in 4 entity categories except for the “Body” category. The most challenging entity types are Disease and Treatment where the F-scores for all models are less than 80%.

**Table 6.** Detailed F-score comparison for all entity categories.

Models	Body	Exam	Disease	Symptom	Treatment
1-g	<b>84.47</b>	92.87	75.76	95.04	74.75
2-g	84.06	93.22	75.55	95.13	73.71
3-g	83.39	92.99	75.13	94.82	74.27
Our model	83.99	<b>93.82</b>	<b>77.89</b>	<b>95.23</b>	<b>75.88</b>

Though our approach is generally applicable, there are several limitations that have to be addressed. First, since our model uses a Bi-LSTM layer that is not able to fully utilize GPU for parallel processing, it is important to handle this issue to ensure both high performance and high computational efficiency. Second, we incorporate an embeddings layer to learn the distributed representation without applying any pre-trained embeddings. We expect that pre-trained embeddings learned from large Chinese medical corpora can help in the Chinese CNER task.

## 5. Conclusions

In this study, we propose a neural model based on n-gram character embeddings to learn more semantic information of Chinese characters and address the problem of Chinese clinical named entity recognition. This method avoids relying on external resources and knowledge bases. We conduct the experiments on CCKS-2017 Shared Task 2 dataset with five categories of clinical named entities. The empirical studies show that our approach performs better than other CNN and LSTM based baselines.

Future work will investigate to obtain a more contextualized representation for named entity recognition. Joint learning trains a single model to handle multiple tasks for the purpose of improving performance on all tasks. We plan to exploit the recognition of EMR sections (general items, medical history, diagnosis and treatment, discharge summary) and jointly train with the Chinese CNER to boost the performance. In addition to the Embeddings and N-gram layers which are used to preprocess Chinese characters, other layers of our approach are expected to be applicable to different languages. Another possible avenue of future work might be to extend the model to other languages in order to maximize the usefulness of our method.

**Author Contributions:** Supervision, J.-S.J.; methodology, J.-S.J., C.-S.L. and C.-H.L.; investigation, C.-S.L. and C.-H.L.; writing—review and editing, C.-S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Skounakis, M.; Craven, M.; Ray, S. Hierarchical hidden markov models for information extraction. *IJCAI* **2003**, *2003*, 427–433.
- Kang, T.; Zhang, S.; Tang, Y.; Hruby, G.W.; Rusanov, A.; Elhadad, N.; Weng, C. EliIE: An open-source information extraction system for clinical trial eligibility criteria. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 1062–1071. [[CrossRef](#)]
- Yadav, V.; Bethard, S. A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 21–25 August 2018; pp. 2145–2158.
- Wang, X.; Yang, C.; Guan, R. A comparative study for biomedical named entity recognition. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 373–382. [[CrossRef](#)]
- Hu, J.; Shi, X.; Liu, Z.; Wang, X.; Chen, Q.; Tang, B. *HITSZ\_CNER: A Hybrid System for Entity Recognition from Chinese Clinical Text*; CEUR Workshop Proceedings: Aachen, Germany, 2017; Volume 1976, pp. 25–30.
- Li, L.; Zhao, J.; Hou, L.; Zhai, Y.; Shi, J.; Cui, F. An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 235. [[CrossRef](#)]
- Gong, L.; Zhang, Z.; Chen, S. Clinical Named Entity Recognition from Chinese Electronic Medical Records Based on Deep Learning Pretraining. *J. Healthc. Eng.* **2020**, *2020*, 8829219. [[CrossRef](#)]
- Wu, G.; Tang, G.; Wang, Z.; Zhang, Z.; Wang, Z. An Attention-Based BiLSTM-CRF Model for Chinese Clinic Named Entity Recognition. *IEEE Access* **2019**, *7*, 113942–113949. [[CrossRef](#)]
- Zhu, Q.; Li, X.; Conesa, A.; Pereira, C. GRAM-CNN: A deep learning approach with local context for named entity recognition in biomedical text. *Bioinformatics* **2018**, *34*, 1547–1554. [[CrossRef](#)]
- Wang, Q.; Zhou, Y.; Ruan, T.; Gao, D.; Xia, Y.; He, P. Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition. *J. Biomed. Inform.* **2019**, *92*, 103133. [[CrossRef](#)]
- Han, X.; Zhou, F.; Hao, Z.; Liu, Q.; Li, Y.; Qin, Q. MAF-CNER: A Chinese Named Entity Recognition Model Based on Multifeature Adaptive Fusion. *Complexity* **2021**, *2021*, 6696064. [[CrossRef](#)]
- Zeng, Q.T.; Goryachev, S.; Weiss, S.; Sordo, M.; Murphy, S.N.; Lazarus, R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: Evaluation of a natural language processing system. *BMC Med. Inform. Decis. Mak.* **2006**, *6*, 30. [[CrossRef](#)]
- Savova, G.K.; Masanz, J.J.; Ogren, P.V.; Zheng, J.; Sohn, S.; Kipper-Schuler, K.C.; Chute, C.G. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. *J. Am. Med. Inform. Assoc.* **2010**, *17*, 507–513. [[CrossRef](#)] [[PubMed](#)]
- Rindflesch, T.C.; Tanabe, L.; Weinstein, J.N.; Hunter, L. EDGAR: Extraction of drugs, genes and relations from the biomedical literature. *Bioinformatics* **2000**, *2000*, 517–528.
- Aronson, A.R. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. In Proceedings of the AMIA Symposium, Washington, DC, USA, 3–7 November 2001; American Medical Informatics Association: Bethesda, MD, USA, 2001; p. 17.
- Gaizauskas, R.; Demetriou, G.; Humphreys, K. Term recognition and classification in biological science journal articles. In Proceedings of the Computational Terminology for Medical and Biological Applications Workshop of the 2nd International Conference on NLP, Patras, Greece, 2–4 June 2000.
- McDonald, C.J.; Overhage, J.M.; Tierney, W.M.; Dexter, P.R.; Martin, D.K.; Suico, J.G.; Zafar, A.; Schadow, G.; Blevins, L.; Glazener, T.; et al. The Regenstrief medical record system: A quarter century experience. *Int. J. Med. Inform.* **1999**, *54*, 225–253. [[CrossRef](#)]
- Donnelly, K. SNOMED-CT: The advanced terminology and coding system for eHealth. *Stud. Health Technol. Inform.* **2006**, *121*, 279.
- Wang, Y.; Yu, Z.; Chen, L.; Chen, Y.; Liu, Y.; Hu, X.; Jiang, Y. Supervised methods for symptom name recognition in free-text clinical records of traditional Chinese medicine: An empirical study. *J. Biomed. Inform.* **2014**, *47*, 91–104. [[CrossRef](#)] [[PubMed](#)]
- Ju, Z.; Wang, J.; Zhu, F. Named entity recognition from biomedical text using SVM. In Proceedings of the 2011 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China, 10–12 May 2011; pp. 1–4.
- Yin, W.; Kann, K.; Yu, M.; Schütze, H. Comparative study of CNN and RNN for natural language processing. *arXiv* **2017**, arXiv:1702.01923.
- Li, Z.; Zhang, Q.; Liu, Y.; Feng, D.; Huang, Z. *Recurrent Neural Networks with Specialized Word Embedding for Chinese Clinical Named Entity Recognition*; CEUR Workshop Proceedings: Aachen, Germany, 2017; Volume 1976, pp. 55–60.

23. Ouyang, E.; Li, Y.; Jin, L.; Li, Z.; Zhang, X. *Exploring N-Gram Character Presentation in Bidirectional RNN-CRF for Chinese Clinical Named Entity Recognition*; CEUR Workshop Proceedings: Aachen, Germany, 2017; Volume 1976, pp. 37–42.
24. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is All you Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
25. Tan, Z.; Wang, M.; Xie, J.; Chen, Y.; Shi, X. Deep semantic role labeling with self-attention. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
26. Ma, Q.; Yan, J.; Lin, Z.; Yu, L.; Chen, Z. Deformable Self-Attention for Text Classification. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 1570–1581. [[CrossRef](#)]
27. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
28. Lafferty, J.; McCallum, A.; Pereira, F.C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning, Williamstown, MA, USA, 28 June–1 July 2001.
29. Alzaidy, R.; Caragea, C.; Giles, C.L. Bi-LSTM-CRF sequence labeling for keyphrase extraction from scholarly documents. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2551–2557.
30. Ma, X.; Hovy, E. End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv* **2016**, arXiv:1603.01354.
31. Li, X.; Zhang, H.; Zhou, X.H. Chinese clinical named entity recognition with variant neural structures based on BERT methods. *J. Biomed. Inform.* **2020**, *107*, 103422. [[CrossRef](#)] [[PubMed](#)]
32. Unanue, I.J.; Borzeshi, E.Z.; Piccardi, M. Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition. *J. Biomed. Inform.* **2017**, *76*, 102–109. [[CrossRef](#)]
33. Qiu, J.; Wang, Q.; Zhou, Y.; Ruan, T.; Gao, J. Fast and accurate recognition of Chinese clinical named entities with residual dilated convolutions. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 935–942.
34. Strubell, E.; Verga, P.; Belanger, D.; McCallum, A. Fast and accurate entity recognition with iterated dilated convolutions. *arXiv* **2017**, arXiv:1702.02098.
35. Zhao, S.; Cai, Z.; Chen, H.; Wang, Y.; Liu, F.; Liu, A. Adversarial training based lattice LSTM for Chinese clinical named entity recognition. *J. Biomed. Inform.* **2019**, *99*, 103290. [[CrossRef](#)] [[PubMed](#)]
36. Jiang, S.; Zhao, S.; Hou, K.; Liu, Y.; Zhang, L. 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), Xiangtan, China, 26–27 October 2019; pp. 166–169.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.