

# Named Entity Recognition of PCI Surgery Information Based on BERT+BiLSTM+CRF



Yuhang Zheng, Li Wang, Feng Li, Hongzeng Xu, and Jiancheng Ge

**Abstract** Percutaneous coronary intervention (PCI) is a vital treatment method for coronary artery disease, but the unstructured nature of its clinical data makes it challenging to utilize directly. The data for this study was obtained from the Cardiovascular Treatment Center of the People's Hospital of Liaoning Province, China. A representative dataset of 5.8% of PCI patients' surgical records was selected for labeling, and a language model-based PCI surgical information entity recognition model was developed. First, Encoder Representations from Transformers (BERT) was employed to express the semantic relationship between characters accurately. Then, BiLSTM was used as a feature extractor to extract contextual relations, and finally, conditional random field (CRF) was applied to optimize the prediction results. Experimental results demonstrated that the F1 score in the PCI surgical information entity recognition model reached 85.49%, which is 25.66% higher than the traditional HMM and 0.94% higher than BiLSTM in deep learning.

**Keywords** Percutaneous coronary intervention · Natural language processing · Clinical named entity recognition

---

Y. Zheng · L. Wang (✉) · J. Ge  
College of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China  
e-mail: [liwang2002@dlnu.edu.cn](mailto:liwang2002@dlnu.edu.cn)

F. Li  
School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

H. Xu  
The People's Hospital of China Medical University, The People's Hospital of Liaoning Province, No. 33 Wenyi Road, Shenhe District, Shenyang 110011, China

Key Laboratory of Cardiovascular Imaging and Research of Liaoning Province, Shenyang 110016, China

## 1 Introduction

Percutaneous Coronary Intervention (PCI) is a cardiology procedure that involves balloon dilation or stent implantation to alleviate symptoms of coronary artery stenosis or occlusion. Although PCI operation information records the complete PCI process, clinical texts related to PCI are unstructured, which makes it difficult for clinicians to effectively utilize the information. Therefore, the use of artificial intelligence to process PCI information in a structured manner is necessary to uncover the full potential value of PCI operation information.

In 1996, the term Named Entity Recognition (NER) was introduced as a fundamental task of Natural Language Processing (NLP) at MUC-6 [1]. NER has demonstrated excellent performance in medical data mining. Early approaches relied on rule-based and dictionary-based methods [2] that used templates based on contextual semantic structures. However, these methods could not effectively summarize difficult-to-extract information, and were relatively expensive. To address these limitations, scholars have applied machine learning methods [3–5] such as Hidden Markov Models (HMM), which outperformed traditional methods but could not effectively use contextual semantics for named entity recognition. Currently, deep learning-based NER methods [6, 7] have become more prevalent than the previous two methods. The most popular method is the Bidirectional Long Short-Term Memory (BiLSTM) method. However, BiLSTM cannot constrain the relationship between predicted labels. The Conditional Random Field (CRF) can better constrain the relationship between labels through the emission probability matrix and transition probability matrix. For example, Li et al. [8] proposed an LSTM-CRF-based named entity recognition method.

The intersection of computer science and medicine has led to the emergence of Clinical Named Entity Recognition (CNER) as an important research field. CNER has been successfully utilized to identify body parts, diseases, drugs, and more in various medical fields [9]. However, its specific application in PCI clinical texts remains largely unexplored. The structure of PCI clinical information is complex, involving a combination of Chinese, English, and symbols. Clinical records in this field are relatively incomplete and disorganized, with a high degree of grammatical errors and context ambiguity. These factors pose significant challenges for the identification of named entities related to PCI clinical and surgical information within China's electronic medical records.

## 2 Methods

The BERT-BiLSTM-CRF model's structure diagram is presented in Fig. 1. This model consists of an input layer, an LSTM layer, a Linear layer, and a CRF layer. The input layer is responsible for converting the input corpus into character vector embedding matrices to facilitate subsequent global feature extraction using Bi-LSTM. The

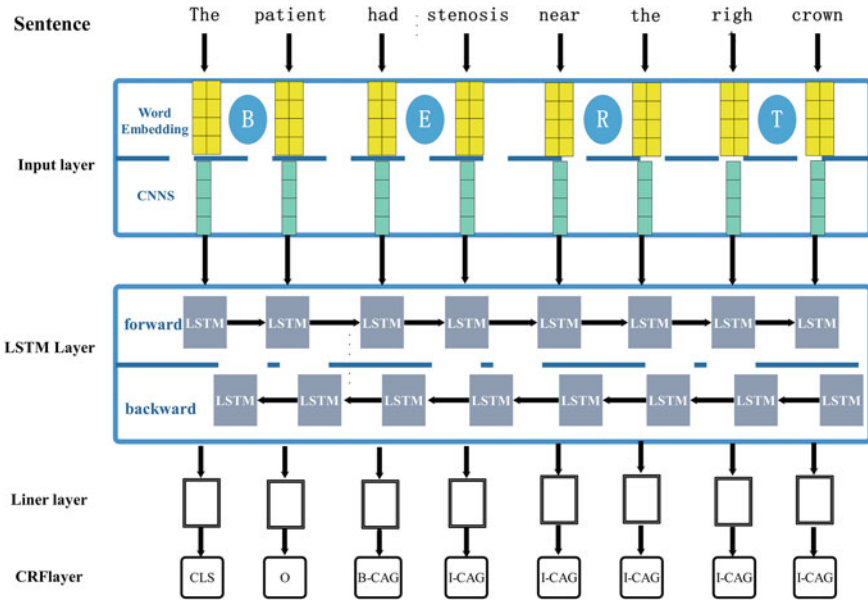


Fig. 1 BERT-BiLSTM-CRF model structure diagram

LSTM layer is tasked with extracting the global features and contextual semantic relationships of the time series. The Linear layer functions as a classifier to assign each entity a probability matrix based on the number of entities. The CRF layer then employs the probability matrix to constrain the relationship between the labels and determine the most probable label sequence.

## 2.1 BERT

In 2018, Devlin et al. [10] introduced BERT, a pre-trained language model based on the Transformer architecture. Unlike traditional language models, BERT employs a 12-layer Transformer Encoder for learning, with each Encoder consisting of a multi-head attention mechanism and a feedforward network. The multi-head attention mechanism calculates the relationship between words using Query, Key, and Value, and adjusts the weight to extract essential features from the text. Compared to previous pre-training models, BERT captures contextual information more accurately and learns the relationship between consecutive text fragments.

Traditional embedding methods, such as Word2vec [11], Glove [12], and FastText [13], represent all possible word-level meanings in vector form. However, the resulting embeddings are often limited in their ability to express the semantic and distance relationships between words. Additionally, in PCI operation information, there is

no clear boundary between Chinese, English, and symbols, which can lead to poor recognition performance. To address these issues, this paper proposes a model that utilizes BERT as an embedding method through transfer learning. Unlike traditional methods, BERT is capable of capturing contextual information and understanding relationships between consecutive text fragments, which can improve performance in recognizing PCI operation information.

## 2.2 LSTM

The LSTM layer's individual unit receives the output from the Input layer and extracts the sequence information of the text to learn the contextual features of the corpus. It can consider the semantic relationship between each character before and after the sentence simultaneously and combine them to create a more comprehensive representation. The LSTM layer takes the WordEmbedding of the corpus as an input sequence and produces another output sequence vector  $h = (h_1, h_2, \dots, h_n)$  that represents the sequence at each time step in the input sequence. LSTM addresses the challenges of long sentence dependencies and gradient explosion in RNNs by incorporating three gate control units that regulate the retention and forgetfulness of specific information. The computation formula for the LSTM hidden layer output representation at a given input time  $X_t$  is expressed as Formula (1)–(5).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f) \quad (2)$$

$$C_t = f_t C_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + b_o) \quad (4)$$

$$h_t = o_t \tanh(C_t) \quad (5)$$

In the equation,  $W$  represents the weight matrix and  $b$  is the offset vector.  $C$  represents the state of the memory unit,  $\sigma$  is the sigmoid activation function, and  $\tanh$  is another activation function. The input gate, forget gate, and output gate are represented by  $i_t$ ,  $f_t$ , and  $o_t$ , respectively. The hidden state output,  $H_t$ , includes the text information in the clinical information. The gate's frequency threshold mechanism is capable of effectively filtering out irrelevant information and retaining important information that needs to be preserved.

However, unidirectional LSTM can only capture information in a forward direction. As sentences become longer, local features at the beginning of the sentence make up a smaller proportion. Therefore, it is necessary to use a Bidirectional LSTM to capture information both forward and backward, and concatenate the outputs. This enhances the information captured from both the beginning and end of the sentence and improves the ability to capture semantic dependencies within the context.

### 2.3 CRF

While BiLSTM is effective at extracting semantic features from the entire text, it does not impose constraints on the label relationships. During the entity output prediction stage, the softmax function is typically used as a classifier to address multi-classification problems, resulting in incorrect outcomes and affecting model performance. To address this issue, this paper employs the CRF model in decoding.

The Conditional Random Field (CRF) is a type of undirected probabilistic graphical model that can constrain the relationship between labels and improve the accuracy of entity predictions. By defining a starting probability matrix and a transition probability matrix, the constraints between tags and the prediction of entities can be enforced. Specifically, because the entity labels starting with “B” must be followed by labels of the same type starting with “I”, the launch probability matrix and transition probability matrix effectively model these constraints and reduce errors. Given an input sequence  $H = h_1, h_2, h_3, \dots, h_n$  and the corresponding output from the LSTM layer, the CRF model calculates the conditional probability distribution of the output sequence  $Y = y_1, y_2, y_3, \dots, y_n$  and assigns a score to each label. The label with the highest score is then selected as the final prediction label using the calculation formula (6):

$$s(h, y) = \sum_{i=1}^N P_{i,y_i} + \sum_{i=1}^N T_{y_{i-1},y_i} \tag{6}$$

When decoding, the Viterbi algorithm [14] is used to obtain the optimal output sequence  $y_R$ . This algorithm computes the score of each label in the sequence based on the transition matrix T and the score vector P. The optimal path is determined by selecting the label with the highest score for each position in the sequence, taking into account the constraints between labels. The calculation formula for the Viterbi algorithm is given by Eq. (7):

$$y_R = \arg \max_{y \in Y} s(h, y) \tag{7}$$

## 3 Experiments and Results

### 3.1 Dataset

The clinical data utilized in this research was sourced from the People’s Hospital of Liaoning Province (Shenyang, China), a prominent and comprehensive third-class hospital in China. The clinical data relates to PCI (percutaneous coronary intervention) procedures involving coronary angiography, balloon dilatation, and stent implantation. Specifically, the clinical texts pertaining to three distinct categories of

**Table 1** Distribution of eight entities in the dataset

Entity	Training set	Validating set	Testing set
Guding	1030	194	264
GuideBrand	1501	307	396
BalloonModel	2136	524	628
SupportModel	1558	273	398
SupportBrand	1578	273	389
Anticoagulants	1012	166	223
ContrastMedium	936	163	225
CAGResult	2900	476	635
All	12,651	2376	3158

fine-grained PCI surgical information were extracted. By conducting a comprehensive analysis of various clinical and surgical records, this study chose to use data from 1340 inpatients who were admitted to the Cardiology Department as the corpus for analysis. This study was finally approved by the Ethics Committee of the People's Hospital of Liaoning Province (Ethics number: (2023)K021).

For the experimental study, a corpus of 1340 cases of PCI surgery information was utilized. From this, 930 cases were selected as the training set, while 180 cases were assigned as the validation set. Following model training, the remaining 230 cases were utilized as the test set. Table 1 displays the distribution of the eight types of entities in the training, validation, and test sets.

### 3.2 Result Analysis

Table 2 shows that traditional HMM achieved excellent results due to the simpler format of anticoagulants than the normative structure, but it performed poorly in identifying entities that require contextual semantics. Bi-LSTM showed the best performance in Chinese and English entity recognition, but it did not perform well in identifying complex entities such as contrast results. However, after adding CRF, the relationship between labels was effectively constrained, resulting in significant improvements in angiography results and entities with complex stent signal structures. This greatly reduced false recognition. Additionally, with the addition of BERT, sensitivity to entities such as the combination of numbers and symbols, such as the bracket model, was increased, and the accuracy of other entities was improved.

The overall performance of a model can be evaluated by its ability to recognize all types of entities correctly. The F1 score for each model is calculated based on the proportion of each entity in the PCI operation information. The results are presented in Table 3.

**Table 2** F1 values of eight entities in each model

Entity	HMM	BiLSTM	BiLSTM-CRF	BERT-BiLSTM-CRF
Guding	25.7	<b>89.2</b>	85.9	86.8
GuideBrand	52.0	89.2	89.9	<b>91.2</b>
BalloonModel	73.9	91.2	91.6	<b>93.1</b>
SupportModel	50.6	75.9	76.7	<b>77.2</b>
SupportBrand	56.7	90.9	93.1	<b>93.2</b>
Anticoagulants	<b>91.9</b>	86.4	90.1	90.5
ContrastMedium	79.5	94.2	<b>94.2</b>	89.7
CAGResult	15.5	68.5	70.5	<b>71.6</b>

Bold indicates the best performing of all models

**Table 3** Performance between models

Model	Strict			Relaxed		
	Precision	Recall	F1	Precision	Recall	F1
HMM	52.08	72.89	59.83	88.98	81.48	82.79
BiLSTM	84.25	84.91	84.55	97.43	97.36	97.35
BiLSTM-CRF	84.71	85.1	84.9	97.43	97.36	97.35
BERT-BiLSTM-CRF	85.27	85.49	<b>85.49</b>	97.43	97.37	<b>97.37</b>

Bold indicates the best performing of all models

To ensure the rigor of the experiment, the overall performance of each model is evaluated under both strict and loose standards. As shown in the table, the traditional HMM performed poorly, with an F1 score of only 59.83 under the strict standard, due to its limited ability to extract features and combine contextual semantics, resulting in a large number of recognition errors and limitations in identifying entities. Although BiLSTM achieved higher performance than HMM, its performance on complex imaging entities was not satisfactory, with incorrect label order leading to a reduced F1 score. The addition of CRF effectively constrained label relationships and improved accuracy for some complex entities, such as those with mixed numbers and symbols, resulting in an overall F1 score improvement of 0.35. In comparison, BERT-BiLSTM-CRF performed better, as the WordEmbedding of the former two was randomly initialized and could not accurately reflect the relative distance of the word vector space in character information representation. This improved accuracy in the recognition of most entities.

## 4 Discussion and Conclusion

This study focuses on the extraction of clinical information from PCI surgery information based on Chinese EMRs. Determine and identify eight entities including angiography result, catheter type, guidewire brand, stent model, stent brand, balloon

model, contrast agent, and anticoagulant from three clinical surgical records. The performance of the BERT-BiLSTM-CRF model for extracting clinical PCI operation information has been further improved, and the accuracy rate basically meets the needs of clinical applications. The results demonstrate that deep learning methods can be used to automatically extract PCI surgical information from EMRs for clinical named entity recognition.

## References

1. Grishman R, Sundheim BM (1996) Message understanding conference-6: a brief history. In: COLING 1996: the 16th international conference on computational linguistics, vol 1
2. Song M, Yu H, Han WS (2015) Developing a hybrid dictionary-based bio-entity recognition technique. *BMC Med Inf Decis Making* 15(1):1–8
3. McCallum A, Li W (2003) Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons
4. Chieu HL, Ng HT (2002) Named entity recognition: a maximum entropy approach using global information. In: COLING 2002: the 19th international conference on computational linguistics
5. Bender O, Och FJ, Ney H (2003) Maximum entropy models for named entity recognition. In: Proceedings of the seventh conference on natural language learning at HLT-NAACL 2003, pp 148–151
6. Wang X, Zhang Y, Ren X, Zhang Y, Zitnik M, Shang J, Langlotz C, Han J (2019) Cross-type biomedical named entity recognition with deep multi-task learning. *Bioinformatics* 35(10):1745–1752
7. Cai X, Dong S, Hu J (2019) A deep learning model incorporating part of speech and self-matching attention for named entity recognition of chinese electronic medical records. *BMC Med Inf Decis Making* 19(2):101–109
8. Li L, Jiang Y (2018) Integrating language model and reading control gate in BLSTM-CRF for biomedical named entity recognition. *IEEE/ACM Trans Comput Biol Bioinform* 17(3):841–846
9. Chen X, Ouyang C, Liu Y, Bu Y (2020) Improving the named entity recognition of Chinese electronic medical records by combining domain dictionary and rules. *Int J Environ Res Publ Health* 17(8):2687
10. Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
11. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Process Syst* 26
12. Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543
13. Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Trans Assoc Comput Linguist* 5:135–146
14. Viterbi AJ (2006) A personal history of the Viterbi algorithm. *IEEE Sig Process Mag* 23(4):120–142