2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

CBRRel: A Chinese Medical Entity Relationship Extraction Model Combining Location Aware Attention and Feature Fusion

Chuanxia Lin[1], Shudong Xia [2], and Jijun Tong[3(✉)]

[1] School of computer science and technology, Zhejiang Sci-Tech University, Hangzhou, 310018, Zhejiang, China
2023220603038@mails.zstu.edu.cn
[2] The Fourth Affiliated Hospital of School of Medicine, International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, 322000, Zhejiang, China
Shystone@zju.edu.cn
[3] School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou, 310018, Zhejiang, China
jijuntong@zstu.edu.cn

Abstract. Entity-relation extraction from Chinese electronic medical records (EMRs) is essential for constructing medical knowledge graphs and enabling intelligent diagnosis and clinical decision-making. However, the presence of complex sentence structures, overlapping entities, and sparse annotations poses significant challenges. To address these issues, we propose CBRRel, a joint extraction model optimized for Chinese EMRs. The model integrates a UNet-based semantic fusion module to enhance multi-scale representation learning and improve boundary detection for complex entities. To further strengthen structural understanding, we introduce a relative position attention mechanism that effectively captures positional dependencies between entity pairs. In addition, we apply the Fast Gradient Method (FGM) adversarial training to improve robustness against input perturbations. Experimental results on the CACMeD dataset show that CBRRel achieves 80.67% precision, 74.13% recall, and a 77.26% F1 score. On the DuIE public dataset, it achieves an F1 score of 76.74%, demonstrating strong capability in handling overlapping and complex relation scenarios. These results highlight the effectiveness of CBRRel and its potential for practical medical information extraction.

Keywords: Chinese EMRs, Entity-relation extraction, Relative position attention, Adversarial Training, UNet semantic fusion.

1 Introduction

With the continuous advancement of healthcare informatization, EMRs, as a core component of modern medical systems, have become a critical foundation for large-scale health data analysis [1],[2]. EMRs contain a vast amount of unstructured textual information related to diseases, symptoms, medications, examinations, and treatments [3], making them highly valuable for tasks such as clinical decision support, rare disease research, and the construction of medical knowledge graphs [4].

Automatically extracting structured medical entities and their semantic relationships from large volumes of unstructured clinical text has become a key research focus in the field of Medical Natural Language Processing (Medical NLP). However, relation extraction in Chinese medical texts remains in its early stages and faces multiple challenges, such as limited annotated corpora, complex language structures, and ambiguous entity boundaries. Compared to English clinical texts, Chinese texts pose additional difficulties for NLP tasks due to the lack of explicit word boundaries, which complicates fundamental tasks like word segmentation and named entity recognition, and ultimately affects the accuracy of relation extraction. Moreover, the lack of publicly available datasets specifically designed for clinical relation extraction in Chinese further limits model development.

To address these challenges, this study makes the following contributions:

1. We construct a manually annotated, small-scale dataset for clinical relation extraction using 573 coronary angiography reports from a Grade A tertiary hospital in Zhejiang Province. The dataset covers common clinical events and relationship types between medical entities, helping address the lack of annotated resources in Chinese clinical settings.
2. To enhance the modeling of semantic associations between entities, we propose an improved version of the CasRel framework [5], which integrates the CKBERT pre-trained language model, a Bi-directional Long Short-Term Memory network (BiLSTM), and a positional attention mechanism. This combination improves the representation of both syntactic and semantic information. In addition, we incorporate adversarial training to improve model robustness, particularly in handling one-to-many relation extraction scenarios.
3. Experimental results demonstrate that the proposed model achieves strong performance on two different datasets, validating its generalizability and practical value. This approach offers effective technical support for Chinese medical information extraction and knowledge graph construction, and contributes to the development of intelligent clinical applications.

2 Related work

In the task of information extraction, traditional pipeline approaches [6–8] treat entity recognition and relation extraction as two separate stages, first identifying entities and then classifying their relations. While these methods are intuitive and easy to implement, they often suffer from error propagation and lack of task-level coordination [9].

To address these issues, joint extraction methods [10] have emerged, aiming to model entity and relation extraction in an end-to-end manner. These approaches significantly enhance the accuracy and consistency of extraction results. Sui et al. [11] introduced a Set Prediction Networks approach for joint extraction in an end-to-end manner, which demonstrated strong performance but required substantial computational resources. To address efficiency concerns, Ma et al. [12] proposed a domain-adaptive generalized predictor that incorporates neural architecture search, effectively balancing performance and resource constraints. In contrast, Gao et al. [13] developed a lightweight joint extraction framework leveraging global entity matching to enhance efficiency. Despite these advancements, the practical deployment of joint methods remains hindered by their implementation complexity and scalability challenges. Li et al. [14] reformulated the task as a multi-turn question answering process by designing question templates with prior knowledge and using reinforcement learning to mitigate error accumulation. Dixit et al. [15] leveraged span alignment and filtering to generate candidate entities, though their relation classification stage is prone to redundancy, thereby limiting precision.

In the Chinese context, most research has focused on open-domain texts, with limited studies on Chinese EMRs. Traditional methods, including dependency parsing and rule-based approaches, have been used to extract relations among clinical entities [16]. In traditional Chinese medicine (TCM), co-occurrence statistics, topic modeling, and semi-supervised learning with external knowledge bases like MEDLINE have also been explored [17]. Recent graph-based and iterative learning methods have shown promise for low-resource settings [18].

Despite these efforts, relation extraction in Chinese EMRs remains constrained by limited annotated data, complex syntax, and ambiguous word boundaries, necessitating more robust and domain-adaptive models.

3 Model introduction

We propose CBRRel, a joint extraction model designed for complex Chinese medical texts. As shown in Figure 1, CBRRel employs CKBERT as the encoder, enhanced with BiLSTM for local dependency modeling. To improve robustness, it incorporates FGM adversarial training. A relative position attention mechanism is used to capture structural relationships, and an improved UNet-style fusion module integrates multi-scale features for more accurate entity and relation extraction.

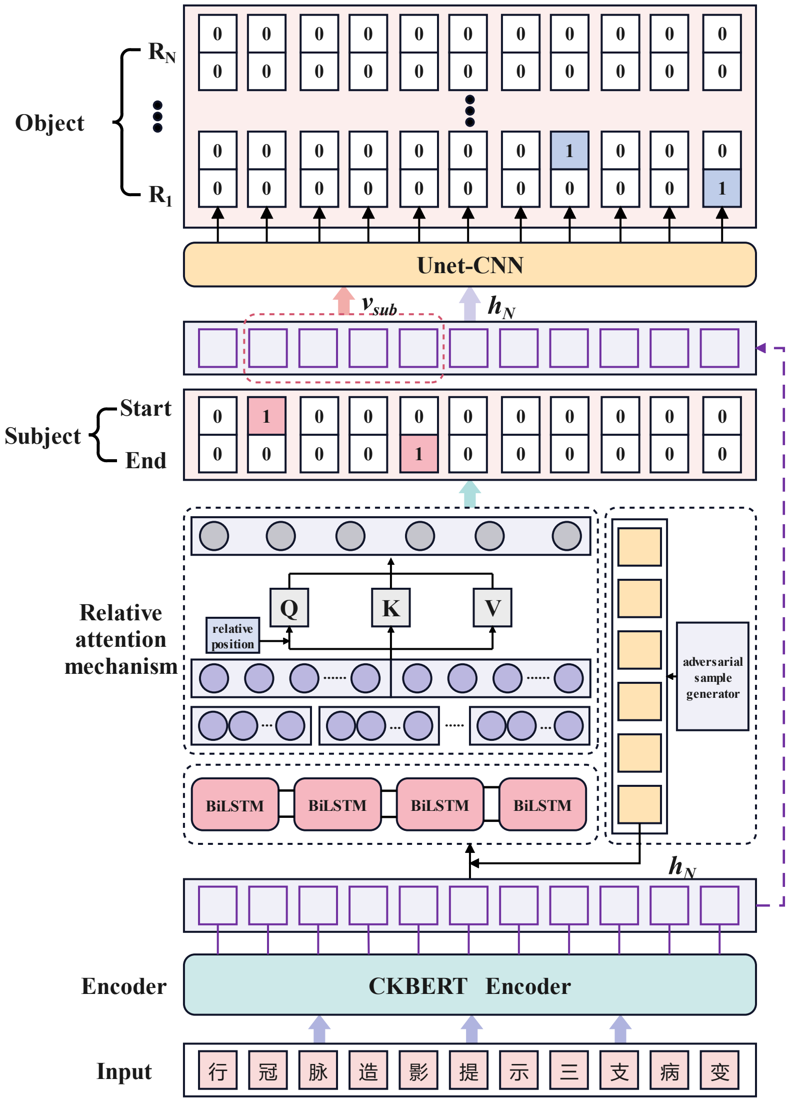

Fig. 1. CBRRel's model architecture.

3.1 Coding module

We employ CKBERT [19], a knowledge-enhanced pre-trained model, as the encoder. CKBERT enhances the standard MLM objective by introducing knowledge-aware inputs and auxiliary tasks such as entity alignment and relation prediction, enabling joint modeling of text and knowledge.

To adapt it to the medical domain, we fine-tune CKBERT using EasyNLP [20], leveraging structured triples from the QABasedOnMedicaKG knowledge graph [21] and 40% of samples from the CASMeD EMRs corpus.

CKBERT is composed of multiple Transformer layers, each consisting of multi-head self-attention and feed-forward neural network components. The input text is first processed with word and positional embeddings to generate the initial representation. Given a one-hot encoded token sequence S after Chinese word segmentation, the embedding is computed as：

$$h_0 = SW_S + W_p \tag{1}$$

where W_S is the word embedding matrix and W_p is the positional embedding matrix. The embedded sequence is then passed through N Transformer layers, with the representation at layer a defined as:

$$h_a = Trans(h_{a-1}), a \in [1, N] \tag{2}$$

Here, h_a denotes the hidden representation at the a -th layer, and $Trans$ represents the Transformer encoding operation. The final output integrates both semantic and structural knowledge, providing deep contextualized features to support downstream relation extraction tasks.

3.2 Adversarial training

In this work, we incorporate the FGM into the CKBERT model. Originally proposed by Goodfellow et al. [22], FGM is an efficient adversarial training strategy designed to improve model robustness. It achieves this by generating adversarial examples through the addition of small perturbations in the direction of the input gradients. This enhances the model's ability to resist noise and adversarial attacks. The optimization objective of FGM is formulated as follows:

$$min_\theta E_{(x,y)\sim} D[max_{\Delta x \in \Omega} L(x + \Delta x, y; \theta)] \tag{3}$$

where D denotes the training dataset, x and y are the input and corresponding label respectively, θ represents the model parameters, $L(x, y; \theta)$ is the loss function for a single sample, Δx denotes the adversarial perturbation, and Ω is the perturbation space. By maximizing the loss under perturbations and minimizing the expected loss, FGM effectively improves the generalization and stability of the model.

3.3 BiLSTM layer

To enhance the model's capability in capturing global semantic information, this study integrates a BiLSTM network into the encoding layer. BiLSTM consists of two LSTM networks operating in opposite directions, which enables it to capture both forward and backward contextual dependencies. This characteristic makes it particularly suitable for relation extraction tasks that are sensitive to context semantics. The update at each time step in the BiLSTM is governed by the following equations:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \tag{4}$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \tag{5}$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \tag{6}$$

$$\tilde{C}_t = tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \tag{7}$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \tag{8}$$

$$H_t = O_t \odot \tanh(C_t) \tag{9}$$

Here, I_t, F_t, and O_t denote the input gate, forget gate, and output gate, respectively; $\tilde{C}_t$ is the candidate cell state; C_t and H_t represent the current cell state and hidden output; and $\odot$ denotes element-wise multiplication.

3.4 Relative positional attention mechanism

The standard self-attention mechanism [23] allows tokens to attend globally but lacks explicit modeling of relative positional relationships, limiting structural understanding. To address this, Shaw et al. [24] proposed Relative Positional Encoding, which integrates token distance biases into attention computation. Formally, the attention scores are modified as:

$$e_{ij} = \frac{(x_i W^Q)(x_j W^K + a_{ij}^K)^T}{\sqrt{d_k}} \tag{10}$$

and the output is:

$$z_i = \sum_j softmax(e_{ij})\,(x_j W^V + a_{ij}^V) \tag{11}$$

In this formulation, a_{ij}^K and a_{ij}^V represent the relative positional bias vectors between positions i and j, which are derived from learnable embedding parameters. This enhancement enables the model to capture both content and structural dependencies, thereby improving its sequential modeling capabilities.

3.5 Subject tagger

The subject tagger decodes subject boundaries from encoder hidden states using two binary classifiers that predict the probability of each token being a subject start or end:

$$p_i^{start\_sub} = \sigma(W_{start} . x_i + b_{start}) \tag{12}$$

$$p_i^{end\_sub} = \sigma(W_{end} . x_i + b_{end}) \tag{13}$$

where x_i denotes the input representation of the i-th token; W and b are learnable parameters; σ denotes the sigmoid activation function. Tokens with probabilities above a threshold are labeled 1, others 0. Subject spans are identified by matching start and end labels.

For training, the model is optimized using the binary cross-entropy loss. The objective function for subject tagging is defined as:

$$p_\theta(s|x) = \prod t \in \{start\_sub, end\_sub\} \prod_{i=1}^{L} (p_i^t)^{I=\{y_i^t=1\}} (1-p_i^t)^{I=\{y_i^t=0\}} (14)$$

where L is the length of the input sequence, and $y_i^{start\_sub}$, $y_i^{end\_sub}$ are the binary ground truth labels for the start and end positions of the subject.

3.6 U-Net feature enhancement layer

To enhance the model's ability to recognize entity boundaries and overlapping relations, this study integrates an improved U-Net-based feature enhancement module following the subject extraction component . The module adopts an encoder–decoder architecture, which leverages multi-layer convolution operations and skip connections to fuse semantic information across different hierarchical levels, thereby improving the model's capacity to capture both local and global contextual features.

Specifically, the convolution operations are applied in a two-dimensional fashion over the textual feature maps, enabling richer and more expressive semantic representations. This design facilitates the aggregation of multi-scale information, which is crucial for handling the structural complexity of entity mentions in clinical texts.The detailed convolutional structure is illustrated in Figure 2.

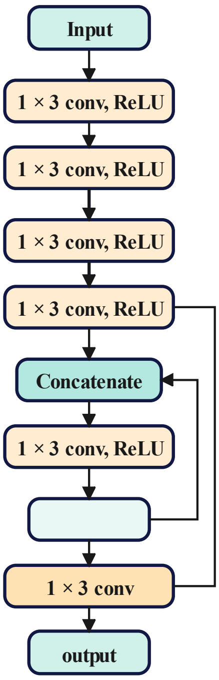

Fig. 2. U-Net convolution process.

3.7 Relation-specific object tagger

In the CasRel framework, subject entity extraction provides essential contextual cues for subsequent relation prediction. To further enhance the interaction between entities and relations, the object entity recognition process incorporates features from the identified subject entities. Specifically, the model predicts the start and end positions of object entities using the following formulations:

$$p_i^{start\_obj} = \sigma(W_{start}^r(x_i + v_{sub}^k) + b_{start}^r) \quad (15)$$

$$p_i^{end\_obj} = \sigma(W_{end}^r(x_i + v_{sub}^k) + b_{end}^r) \quad (16)$$

Here, $p_i^{start\_obj}$ and $p_i^{end\_obj}$ represent the probabilities of the i-th token being the start and end of an object entity, respectively. The vector v_{sub}^k denotes the semantic representation of the k-th subject entity identified by the subject tagging module. This vector is integrated into the computation to strengthen the association between the subject and object entities within the given relation.

Based on the sentence vector enriched with subject-specific information, the probability of an object entity is further computed as:

$$p_{\phi r}(o|s,x) = \prod t \in \{start\_obj, end\_obj\} \prod_{i=1}^{L} (p_i^t)^{y_i^t} (1 - p_i^t)^{1-y_i^t} \quad (17)$$

where L is the sentence length, and $y_i^{start\_obj}$, $y_i^{end\_obj}$ are binary labels indicating whether the i-th token marks the start or end of an object.

The final loss function K for model optimization is defined as:

$$K = \sum_{i=1}^{|D|} [\sum_{s \in T_i} logp_\theta^{sub}(s|x_i) + \sum_{s \in T_r} logp_\theta^{obj}(o|s,x_i)] \quad (18)$$

where $|D|$ denotes the total number of training samples, T_i represents the set of subject entities in the i-th sentence, and T_r corresponds to the set of relations associated with each subject. During training, the model maximizes the likelihood of correctly extracting both subject and object entities via the respective probability functions p_θ^{sub} and p_θ^{obj}.

4 Experiments

4.1 Dataset

Private dataset. The CASMeD private dataset consists of 573 coronary angiography surgical records from a Class III Grade A hospital in Zhejiang Province, China. Each record includes detailed diagnostic procedures, patient history, surgical steps, and postoperative evaluations. All data were anonymized to ensure patient privacy, with ethical approval obtained from the hospital's ethics committee. To improve model performance, we preprocessed the text by removing irrelevant symbols and correcting typographical errors. The dataset was split into training and testing sets in an 8:2 ratio. A statistical overview is provided in Table 1.

Table 1. Statistics on types of entity relationships.

Relation	Meaning	Count
BpSy	Symptoms of Body Parts	5379
BpIn	Examination of Body Parts	72
BpTm	The Way Body Parts Are Treated	2671
TmDo	The Drug and Dosage Used in the Treatment	1199
TmMc	Medical Consumables for Treatment	970
InSy	Symptoms Found by Medical Examination	829
InMc	Medical Consumables for Medical Examinations	528
McBp	Medical Consumables for Body Parts	444
Total		12092

Public dataset. We conducted experiments on the publicly available DuIE dataset released by Baidu. This large-scale dataset features high-quality annotations and comprises a total of 211,529 samples, covering 49 distinct relation types. Specifically, it includes 82,413 SEO instances, 22,174 EPO instances, and 106,942 instances of other relation types. Evaluating our model on this dataset helps to assess its robustness and generalization capability in large-scale and diverse Chinese relation extraction tasks. To further illustrate the fundamental characteristics of both datasets, Table 2 provides a statistical summary of the CASMeD and DuIE datasets.

Table 2. Data set statistics.

Statistcians	DuIE	Ours
Quantity	17924	573
Relational triad	75348	12092
Relationship number	49	8
Normal	106942	7940
SEO	82413	4064
EPO	22174	88

4.2 Implementation details

The experimental environment for this study consists of an Intel 13600KF CPU and an NVIDIA RTX 4070Ti GPU with 12GB of memory. In terms of software, Python 3.8.20 and PyTorch 1.8.0 were used as the primary deep learning framework. All experiments were conducted under this configuration to ensure the stability and reproducibility of results. In addition, the detailed hyperparameter settings used throughout the experiments are summarized in Table 3.

Table 3. Hyperparameter settings.

Parameter name	Parameter value
Bert_hidden_size	768
BiLSTM_hidden_size	768
Optmizer	EMA
Batch_size	4
Max_seq_length	300
Learning rate	1e-5
Dropout	0.1
FGM_epsilon	3e-5
Epoch	50

4.3 Evaluation metrics

The key to triple extraction lies in accurately identifying the subject (s), object (o), and their corresponding relation (r). To evaluate the model's performance, we employ Precision, Recall, and F1-Score as evaluation metrics, which are computed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{21}$$

where TP represents the number of correctly extracted triples, FP denotes the number of incorrectly extracted triples, and FN refers to the number of triples that exist in the dataset but are not correctly identified. A predicted triple (s,r,o) is considered correct only if the subject, object, and relation all match the ground truth.

5 Results and analysis

5.1 Comparative experiment

To validate the effectiveness of the proposed CBRRel model for entity-relation extraction from Chinese EMRs, we conducted comparative experiments on the private CASMeD dataset and the public DuIE dataset. We compared CBRRel against several representative baselines:

SpERT [25] is a span-based joint extraction model that incorporates negative sample features to improve entity and relation representations. OneRel [26] unifies entity-relation extraction using a scoring-based relation classifier combined with relation-specific tagging. BiRTE [27] models bidirectional interactions between entities and relations at the token level via a shared encoder. TPLinker [28] uses a handshaking tagging mechanism to effectively detect overlapping entities of multiple types. EmRel [29] captures relational triplets by jointly modeling entities, their context, and the relation. AGT [30] employs a Transformer-based encoder-decoder with a pointer network for end-to-end triplet extraction. CasRel [5] utilizes a two-stage binary tagging strategy for subject recognition, relation classification, and object extraction, excelling in handling overlapping triplets.

As shown in Table 5, CBRRel consistently outperformed mainstream joint extraction baselines on both the DuIE and private CASMeD datasets. On CASMeD, it achieved an F1-score of 76.76%, surpassing the best baseline CasRel by 5.69%, with improvements of 6.66% in precision and 4.82% in recall. This superior performance results from key architectural enhancements: a BiLSTM encoder for better local context modeling; FGM-based adversarial training to improve robustness; relative positional attention to capture structural entity relationships; and an enhanced U-Net module for multi-scale feature fusion, which improves entity boundary detection and complex relation recognition. These innovations enable stronger triplet extraction and greater stability, especially in structurally complex, relation-overlapping medical texts.

On the DuIE dataset, CBRRel achieved an F1-score of 79.96%, outperforming most existing models and demonstrating strong generalization. However, the improvement margin was smaller than on the private dataset, mainly because DuIE features simpler sentence structures, shorter lengths, and more balanced entity-relation distributions. Under these conditions, CBRRel's hierarchical modules and adversarial training offer less advantage. Thus, while competitive on general datasets, CBRRel's strengths are more pronounced in complex, domain-specific tasks like medical relation extraction.

Table 4. Comparative experiments.

Method	DuIE			Ours		
	Precision	Recall	F1	Precision	Recall	F1
SpERT	80.45	64.52	71.61	62.84	68.35	65.48
BiRTE	83.03	67.29	74.34	71.35	69.64	70.48
TPLinker	77.92	81.05	79.45	71.82	73.26	72.53
CasRel	74.32	71.81	73.04	73.02	69.23	71.07
OneRel	76.28	73.71	74.97	80.37	56.42	66.30
ATG	78.34	74.02	76.12	77.18	73.62	75.35
EmRel	77.40	75.29	76.33	72.63	74.41	73.51
Ours	79.58	74.09	76.74	80.67	74.13	77.26

5.2 Analysis of Model Effectiveness for Overlapping Relationship Extraction

To further validate CBRRel's effectiveness in extracting overlapping triplets, we conducted a fine-grained analysis on three sentence types with varying structural complexity and relation overlap: SEO, EPO, and Normal. Table 6 compares CBRRel's performance against baselines under different overlapping scenarios. CBRRel showed significant gains, outperforming AGT by 3.89% and CasRel by 2.79% in SEO, and surpassing AGT by 6.09% and CasRel by 0.96% in EPO, demonstrating superior ability in distinguishing and extracting triplets from complex sentences with multiple overlapping relations.

In contrast, the AGT model, due to its autoregressive generation mechanism, shows limitations in capturing global interaction information among entities, making it less effective for overlapping relation extraction tasks. CBRRel, by incorporating a relative positional attention mechanism and a UNet-based encoding architecture, significantly enhances the modeling of long-distance dependencies between entities, thereby improving the acquisition of global semantic and structural information. Although the EmRel model leverages a Tucker decomposition-based triplet representation alignment strategy to model high-order interactions among entities and relations, it falls short in fine-grained feature modeling, resulting in limited precision when handling complex overlapping cases. Similarly, OneRel maps different relations to separate sub-matrices without a unified perspective for relational discrimination, leading to fragmented representations and reduced accuracy in overlapping relation scenarios.

Overall, the multi-scale semantic fusion structure and decoupled decoding mechanism employed by CBRRel significantly enhance the model's adaptability to various types of overlapping relations. In particular, CBRRel exhibits stronger consistency and robustness when a single entity is associated with multiple relations, ultimately delivering superior performance and stability in the extraction of complex and diverse relational triplets.

Table 5. Results of extraction performance with or without overlapping triples.

Method	KPIs	Overlap scenarios		
		EPO	Normal	SEO
ATG	Precision	67.43	84.57	70.93
	Recall	65.47	80.51	66.48
	F1	66.43	82.49	68.63
CasRel	Precision	69.72	77.33	71.74
	Recall	71.39	73.68	67.91
	F1	70.55	75.41	69.73
CBRRel	Precision	73.66	81.77	74.89
	Recall	69.49	79.73	70.30
	F1	71.51	80.74	72.52
EmRel	Precision	66.31	79.35	67.28
	Recall	64.63	78.55	67.88
	F1	65.45	78.94	67.57
OneRel	Precision	69.46	79.31	71.68
	Recall	68.11	73.34	69.75
	F1	68.77	76.20	70.70
SpERT	Precision	61.46	73.52	58.35
	Recall	60.68	72.22	57.57
	F1	61.06	72.86	57.95

5.3 Entity Recognition

To evaluate the performance of the CBRRel model in medical entity recognition tasks, we conducted experiments on six categories of entities: Bp, Do, In, Mc, Sy, and Tm. As shown in Table 6, CBRRel consistently achieved higher F1-scores across all entity types compared to CasRel, with particularly notable improvements in low-resource categories such as “In” and “Mc,” where it outperformed CasRel by 4.10% and 1.63%, respectively. These results underscore the model’s enhanced adaptability in low-resource scenarios.

The observed performance gains can be attributed to several key design choices within CBRRel. The BiLSTM encoder effectively captures local contextual information, while the relative position attention mechanism strengthens the modeling of structural relationships among entities. Additionally, the improved UNet module allows for more precise identification of entity boundaries. Furthermore, the incorporation of FGM adversarial training enhances the model's robustness to noise and perturbations, thereby contributing to greater overall stability in entity recognition.

Taken together, these advances enable CBRRel to deliver superior entity recognition performance and generalization capabilities, particularly in complex, overlapping, and heterogeneous medical text environments.

Table 6. Results of model recognition performance for different entities.

Method	KPIs	Entity					
		Bp	Do	In	Mc	Sy	Tm
CBRRel	Precision	86.19	95.85	86.59	85.91	83.74	89.13
	Recall	79.33	84.62	74.02	76.28	81.43	79.44
	F1	82.62	89.89	79.81	80.81	82.57	84.01
CasRel	Precision	85.44	94.42	80.42	82.11	84.20	85.98
	Recall	78.92	83.17	71.52	76.45	79.93	75.06
	F1	82.05	88.44	75.71	79.18	82.01	80.15

5.4 Ablation experiment

To assess the contribution of each key component within the CBRRel model, we conducted an ablation study on both the DuIE dataset and the private medical dataset. As shown in Table 7, the complete model achieved the best performance across both datasets, with F1-scores of 76.74 and 77.26, respectively, demonstrating the synergistic effect of the integrated modules on overall performance. Notably, the removal of the relative position attention mechanism led to F1-score drops of 0.87 and 1.69, respectively, indicating its critical role in modeling the structural dependencies between entities and relations. Eliminating FGM adversarial training resulted in a substantial performance degradation of 1.61 F1 on the private dataset, highlighting its effectiveness in enhancing model robustness. Similarly, removing the BiLSTM encoder caused noticeable performance declines, affirming its value in capturing local semantic dependencies.

Moreover, replacing the structural encoding module with a conventional CNN or relying solely on the CKBERT encoder led to more significant drops in performance, particularly on the private dataset, where F1-scores declined to 75.25 and 71.07, respectively. These results underscore the necessity of multi-layer semantic modeling and structure-aware mechanisms when dealing with complex, entity-overlapping clinical narratives. In conclusion, each component of the CBRRel model plays a pivotal role in enhancing relation extraction performance in medical texts, and its architectural design is particularly well-suited for structurally intricate and semantically dense healthcare scenarios.

Table 7. Ablation experiment.

Methodologies	DuIE			Ours		
	Precision	Recall	F1	Precision	Recall	F1
Ours	79.58	74.09	76.74	80.67	74.13	77.26
-Attention	78.15	73.71	75.87	78.01	73.27	75.57
-FGM	78.23	73.96	76.04	78.38	81.16	75.65
-BiLSTM	77.92	73.61	75.70	78.94	73.53	76.13
-CNN	78.74	73.23	75.89	77.91	72.76	75.25
Only CKBERT	76.41	71.13	73.68	73.02	69.23	71.07

6 Conclusion

This paper proposes CBRRel, a joint entity and relation extraction model optimized through a multi-module structural design, aiming to enhance the extraction of relational triples in structurally complex and semantically dense texts such as Chinese electronic medical records. The architecture of CBRRel integrates several key components: it adopts CKBERT as the base encoder and incorporates a BiLSTM layer to strengthen local contextual representation. To improve model robustness against input perturbations, an FGM-based adversarial training mechanism is employed. A relative position attention mechanism is further introduced to better capture semantic and positional dependencies between entities. In addition, an enhanced UNet-based feature fusion module is designed to enable multi-scale semantic integration, thereby improving the accuracy of both entity boundary detection and complex relation identification.

Experimental results on both the publicly available DuIE dataset and a proprietary medical dataset demonstrate that CBRRel achieves strong performance in relation extraction tasks, particularly in challenging scenarios involving overlapping entities and multiple relations. The model exhibits superior generalization and extraction stability under such complex conditions. Future work will focus on more efficient modeling of complex and rare relation types, exploring cross-modal information to further support relation extraction, and improving the model's adaptability to long-text and multi-sentence inputs, with the ultimate goal of advancing the application of medical knowledge graph construction in clinical decision support systems.

Acknowledgments. This study was funded by the Key Research and Development Program of Zhejiang (Grant numbers 2025C01135)

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Zhao, Q., Li, J.: Knowledge-enhanced relation extraction for Chinese EMRs. IT Professional **22**(4), 57-62 (2020)
2. Shickel, B., Tighe, P.: Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics **22**(5), 1589-1604 (2017)
3. Yang, J. F., Guan, Y.: Corpus construction for named entities and entity relations on Chinese electronic medical records. Ruan Jian Xue Bao/J Softw **27**(11), 2725–2746 (2016)
4. Murali, L., Gopakumar, G.: Towards electronic health record-based medical knowledge graph construction, completion, and applications: A literature study. Journal of biomedical informatics **143**, 104403 (2023)
5. Wei, Z., Su, J.: A novel cascade binary tagging framework for relational triple extraction. arXiv preprint arXiv:1909.03227 (2019)
6. Zelenko, D., Aone, C.: Kernel methods for relation extraction. Journal of machine learning research **3**, 1083-1106 (2003)
7. Chan, Y. S., Roth, D.: Exploiting syntactico-semantic structures for relation extraction. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, pp. 551-560. ACL, Portland, Oregon, USA (2011)
8. Gormley, M. R., Yu, M.: Improved relation extraction with feature-rich compositional embedding models. arXiv preprint arXiv:1505.02419 (2015)
9. Li, Q., Ji, H.: Incremental joint extraction of entity mentions and relations. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 402-412. ACL, Baltimore, Maryland, USA (2014)
10. Tuo, M., Yang, W.: Review of entity relation extraction. Journal of Intelligent & Fuzzy Systems **44**(5), 7391-7405 (2023)
11. Sui, D., Zeng, X.: Joint Entity and Relation Extraction With Set Prediction Networks. In: IEEE Transactions on Neural Networks and Learning Systems, pp. 12784-12795. IEEE, Piscataway, NJ, USA (2024)
12. Ma, L., Kang, H.: Single-Domain Generalized Predictor for Neural Architecture Search System. In: IEEE Transactions on Computers, pp. 1400-1413. IEEE, Piscataway, NJ, USA (2024)
13. Gao, C., Zhang, X.: ERGM: A multi-stage joint entity and relation extraction with global entity match. Knowledge-Based Systems **271**, 110550 (2023)
14. Li, X., Yin, F.: Entity-relation extraction as multi-turn question answering. arXiv preprint arXiv:1905.05529 (2019)
15. Dixit, K., Al-Onaizan, Y.: Span-level model for relation extraction (2019)
16. Hassan, M., Makkaoui, O.: Extracting disease-symptom relationships by learning syntactic patterns from dependency graphs. In: BioNLP 15, pp. 184. ACL, Beijing, China (2015)
17. Zhou, X., Liu, B.: Integrative mining of traditional Chinese medicine literature and MEDLINE for functional gene networks. Artificial intelligence in medicine **41**(2), 87-104 (2007)

18. Ryan, R. J.: Groundtruth budgeting: a novel approach to semi-supervised relation extraction in medical language (Doctoral dissertation, Massachusetts Institute of Technology) (2011)
19. Zhang, T., Dong, J.: Revisiting and advancing chinese natural language understanding with accelerated heterogeneous knowledge pre-training. arXiv preprint arXiv:2210.05287 (2022)
20. Wang, C., Qiu, M.: Easynlp: A comprehensive and easy-to-use toolkit for natural language processing. arXiv preprint arXiv:2205.00258 (2022)
21. Guo, Q., Cao, S.: A medical question answering system using large language models and knowledge graphs. International Journal of Intelligent Systems **37**(11), 8548-8564 (2022)
22. Dong, Y., Liao, F.: Boosting adversarial attacks with momentum. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9185-9193. IEEE, Salt Lake City, Utah, USA (2018)
23. Vaswani, A., Shazeer, N.: Attention is all you need. Advances in neural information processing systems, **30** (2017)
24. Shaw, P., Uszkoreit, J.: Self-attention with relative position representations. arXiv preprint arXiv:1803.02155 (2018)
25. Eberts, M., Ulges, A.: Span-based joint entity and relation extraction with transformer pre-training. In: ECAI 2020, pp. 2006-2013, IOS Press, Santiago de Compostela, Spain (2020)
26. Shang, Y. M., Huang, H.: Onerel: Joint entity and relation extraction with one module in one step. In: Proceedings of the AAAI conference on artificial intelligence, pp. 11285-11293. ACL, Online (2022)
27. Ren, F., Zhang, L.: A simple but effective bidirectional framework for relational triple extraction. In: Proceedings of the fifteenth ACM international conference on web search and data mining, pp. 824-832. ACM, New York, NY, USA (2022)
28. Wang, Y., Yu, B.: TPLinker: Single-stage joint extraction of entities and relations through token pair linking. arXiv preprint arXiv:2010.1341 (2020)
29. Xu, B., Wang, Q.: EmRel: joint representation of entities and embedded relations for multi-triple extraction. In: Proceedings of the 2022 conference of the north american chapter of the association for computational linguistics: human language technologies, pp. 659-665. ACL, Seattle, Washington, USA (2022)
30. Zaratiana, U., Tomeh, N.: An autoregressive text-to-graph framework for joint entity and relation extraction. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 38, No. 17, pp. 19477-19487. AAAI, Washington, D.C., USA (2024)