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

GafRel: A Joint Entity and Relation Extraction Framework for Chinese Electronic Medical Records with Multidimensional Semantic Enhancement

Chenyang He[1], Shudong Xia[2], Jijun Tong[3] (✉)

[1] School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou, 310018, China
[2] Department of Information Technology, the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu 322000, China
[3]School of information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou, 310018, China
jijuntong@zstu.edu.cn

Abstract. Relation extraction from electronic medical records (EMRs) is essential for advancing biomedical information systems, but it remains challenging due to nested entity structures and limited contextual representations. To address these issues, we propose GafRel, a novel joint entity and relation extraction framework for Chinese EMRs. GafRel extends CasRel by integrating a Global Pointer to better handle nested entity recognition. Additionally, we design a Multi-dimensional Feature Enhancement Layer (MFEL), which enables multi-scale contextual modeling through semantic fusion of both local and global features. This architecture enhances the capacity of the model to capture local continuity and long-range dependencies. To address the lack of relation extraction datasets in Chinese EMRs, we construct DiaRel, a new dataset derived from EMRs of 608 hospital patients. Experiments on CMeIE-v2, DiaKG, and DiaRel demonstrate the strong performance of our method, where GafRel outperforms existing baselines with F1 scores of 53.38%, 53.41%, and 83.98% on the three datasets, respectively. These results highlight the effectiveness of GafRel in extracting complex relations from EMRs and its potential for advancing biomedical information extraction.

Keywords: Chinese electronic medical records; Relation extraction; Global pointer; Multidimensional Feature Enhancement Layer.

1 Introduction

The digitization of healthcare systems has made Electronic Medical Records (EMRs) essential repositories of patient health data. While EMRs have significantly enhanced healthcare efficiency and diagnostic accuracy, their full potential remains underutilized due to the abundance of unstructured free-text data [1]. By locating medical entities and determining their semantic relationships within clinical narratives, relation extraction plays a crucial part in realizing this potential. This process facilitates the transformation of unstructured text into structured relational knowledge, thereby enabling downstream applications such as clinical decision-making [2], risk prediction [3], and knowledge graph construction [4].

Despite these advancements, EMR-based relation extraction faces persistent challenges. First, while international datasets exist, the scarcity of high-quality Chinese EMR datasets impedes research progress and model development in this domain. Second, medical entities in EMRs often exhibit complex nested structures characterized by localized density and long-range dependencies. These properties complicate the integration of fine-grained local features with global contextual information, particularly in scenarios involving multiple interrelated entities. The exponential growth of potential relational in such cases escalates extraction complexity [5], undermining semantic understanding and extraction accuracy, thereby compromising result reliability.

To address these challenges, we propose a relation extraction framework specifically adapted to Chinese EMRs, with the main contributions as follows:

1. To address the lack of domestic medical relation extraction datasets, this study used clinical EMRs from a tertiary hospital in Zhejiang Province as the data source. Based on the medical record information of 608 patients, a small-scale medical relationship extraction dataset was constructed, providing a foundation for subsequent research.
2. The Global Pointer [6] mechanism is integrated into the CasRel framework to enhance the recognition of nested and overlapping entities. Meanwhile, the original cascade decoding strategy is preserved, enabling subject entity information to guide the extraction of objects and their relations.
3. This study proposed a novel component called the Multidimensional Feature Enhancement Layer (MFEL) to address the challenge of effectively integrating fine-grained local features with global semantic context. MFEL introduces a stacked, multi-branch architecture that integrates 1-2-4 dilated convolutions, and a channel-wise attention mechanism. This design enables iterative cross-scale semantic fusion, allowing the model to capture and align features across multiple receptive field sizes. By dynamically selecting contextually relevant features while preserving local continuity, MFEL enhances the ability of the model to extract richer and more robust semantic representations.
4. Finally, the proposed relation extraction model was extensively validated across multiple datasets. On the public datasets CMeIE-v2 and DiaKG, the model achieved F1 scores of 53.38% and 53.41%, respectively. Notably, on the privately constructed dataset, the F1 score further improved to 83.98%, demonstrating the superior performance of the optimized model in medical relation extraction tasks.

2 Relative works

In this section, we will briefly review the current development status of relationship extraction, pre-trained models, and electronic medical record datasets.

2.1 Relationship extraction and Pre trained model

Entity and relation extraction (ERE), as a core task in natural language processing (NLP), aims to identify entities and the semantic relationships between them from unstructured text.

Early ERE methods depended on handcrafted rules and feature engineering, such as feature vector models [7] and SVMs integrating lexical and syntactic features [8], but faced issues with generalizability, maintenance, and expert dependence [9].

With the advancement of deep learning, neural network-based models have gradually replaced traditional methods. ERE evolved into a pipeline process, typically performing named entity recognition (NER) first, followed by relation classification. Representative works include Cheng et al. [10], who proposed a multimodal entity recognition method integrating text and images to enhance recognition accuracy, as well as shortest-path-enhanced architectures for medical relation extraction [11]. However, pipeline methods are prone to error propagation between stages and are inefficient when dealing with overlapping or nested entities.

To mitigate these issues, joint extraction models have been proposed, which integrate entity recognition and relation classification into a unified framework. For instance, Hong et al. [12] introduced a GCN-based model to jointly learn entity and relation features. CasRel [13], a cascade binary classification model, first identifies subject entities and then predicts object entities conditioned on each subject, showing strong performance in end-to-end triple extraction. Nonetheless, the subject-first strategy of CasRel may lead to missed entity pairs and does not explicitly model nested entities. More recent models have reframed ERE as token-pair tagging tasks, such as TPLinker [14] and OneRel [15], which unify entity and relation extraction through tagging schemes.

In addition, GPLinker significantly improves the performance of complex nested structures by jointly predicting entity boundaries and relationships through span matrices.

Despite these advancements, challenges remain in modeling cross-scale semantic dependencies, integrating multi-level contextual information, and adapting to domain-specific textual structures in medical scenarios.

Parallel advancements in pre-trained language models have further enhanced the performance of entity and relation extraction tasks. General-purpose models such as BERT [16] and RoBERTa [17] have established strong contextual representations across a wide range of NLP applications. Building upon these foundations, domain-specific variants tailored for biomedical contexts, such as ClinicalBERT [18] have demonstrated significant improvements in medical NLP tasks, particularly in extracting clinical entities and their relationships from electronic medical records.

2.2 Electronic medical record datasets

In the medical information domain, research in Western countries started earlier and has benefited from high-quality datasets. In 2010, the i2b2/VA Challenge introduced a clinical record entity and relation extraction task, which remains one of the most representative English-language datasets for clinical relation extraction [19].

In contrast, Chinese medical datasets have long been scarce, fragmented, and difficult to obtain. To address this gap, the China Health Information Processing Conference launched the Chinese Biomedical Language Understanding Evaluation (CBLUE) [20] initiative, which introduced the Chinese Medical Named Entity Recognition dataset and the Chinese Medical Entity Relation Extraction dataset (CMeIE). These efforts have drawn increasing attention to Chinese medical information extraction tasks.

However, despite the progress brought by CBLUE, large-scale and high-quality relation extraction datasets for Chinese EMRs remain extremely limited.

3 Model methods

In this section, we describe our relation extraction method, which aims to extract triples in the format of <subject, relation, object> from EMRs based on predefined relationships. Our approach follows a cascade framework.

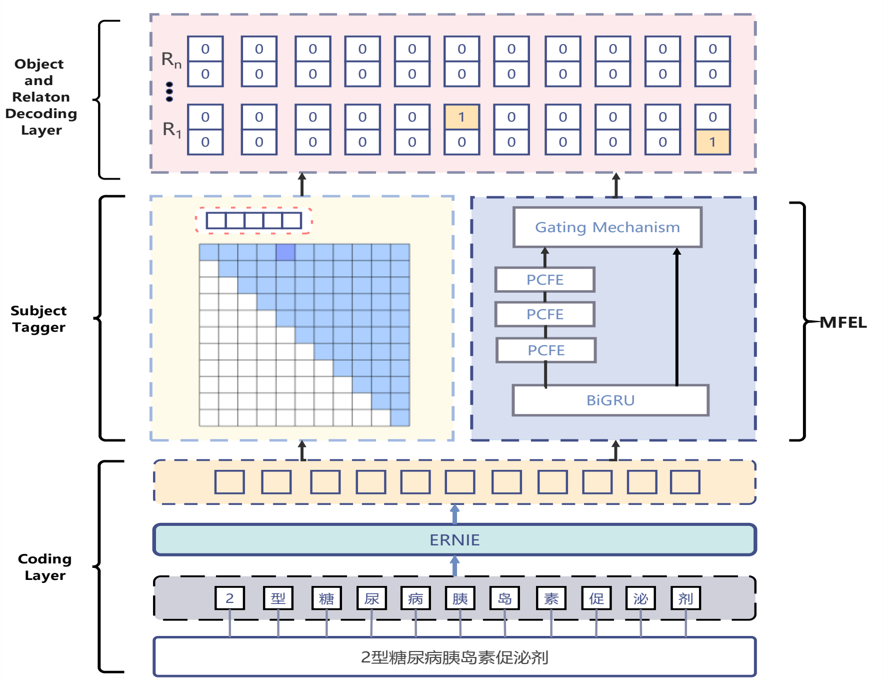


Fig. 1. Main architecture of the proposed GafRel.

In previous studies, training samples were usually generated by randomly selecting a single instance along with its corresponding triples from the given dataset. However, in EMRs containing a high number of triples, this approach often results in substantial missing extractions. To address this issue, our method includes all instances from each data input, ensuring that they are distributed together for training. This not only reduces the number of epochs but also shortens the overall training time. Our method can be summarized in two steps: Extracting the subject, and then extracting the corresponding relations and objects based on the subject.

Taking the sentence from **Fig. 1.** as an example, the model first identifies the potential subject entity, "Type 2 diabetes". It then predicts the possible object under the predefined relationship. This process results in the extracted triple:<Type 2 diabetes, Drug treatment for disease, Insulin secretagogues>.

3.1 Encoding layer

We employ ERNIE 3.0 [21] as the base encoder, which introduces a key innovation by incorporating both lexical and knowledge-level semantics into representation learning.

Unlike BERT, which applies a local subword-level masking strategy, ERNIE employs dynamic entity-level masking. In this approach, entire named entities are masked during pretraining, compelling the model to capture global semantic dependencies among entities, which is especially beneficial for complex medical texts. Furthermore, ERNIE is specifically optimized for Chinese-language processing, making it a more suitable choice for Chinese medical information extraction tasks. The corresponding equation is illustrated in Eq 1:

$$H_1 = ERNIE(t_1, t_2, \dots \dots t_n) = (h_1, h_2, \dots \dots h_n) \tag{1}$$

Where $(t_1, t_2, \dots \dots t_n)$ represents the input token sequence and $(h_1, h_2, \dots \dots h_n)$ represents the feature vector of the corresponding token.

3.2 Subject tagger

The accuracy of head entity recognition is crucial in the CasRel framework, as it directly impacts tail entity and relation extraction. To address the limitations of CasRel in handling nested head entities, this study introduced a Global Pointer mechanism that models entity boundaries as span-level pairs. Unlike the original character-by-character decoding, Global Pointer leverages a boundary matrix to detect overlapping entities more accurately. This enhancement improves head entity recognition and subsequently boosts overall relation extraction performance. The subject extraction of the model is shown in the **Fig. 2.**

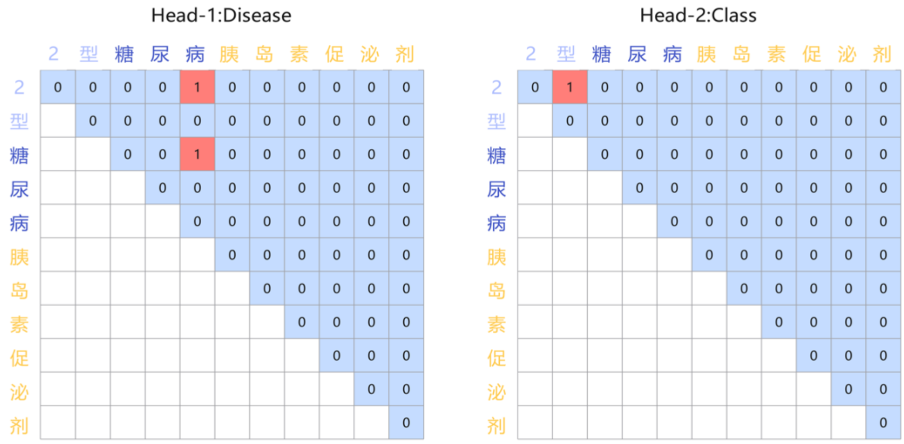
Specifically, the approach can be described as follows: After encoding, the obtained vector sequences $(h_1, h_2, \dots \dots h_n)$ transformed through decoding, resulting in the sequence vectors: $[q_{1,a}, q_{2,a} \dots \dots q_{n,a}]$ and $[k_{1,a}, k_{2,a} \dots \dots k_{n,a}]$, At this stage, the model assigns a score to the position at row i, column j in the matrix for entity type α. The equation is as follows:

$$q_{i,\alpha} = w_{q\alpha} h_i + b_{q\alpha} \tag{2}$$

$$k_{i,\alpha} = w_{k\alpha} h_i + b_{k\alpha} \tag{3}$$

$$s_\alpha(i,j) = {q_{i,\alpha}}^{\mathrm{T}} k_{j,\alpha} \tag{4}$$

Where i represents the position of the i^{th} word in the sentence after linear transformation, and α represents the alpha attribute in the attribute set, b_{qa} and b_{ka} are trainable parameters. $s_a(i,j)$ represents the score where the attribute of the entity in the continuous segment from i to j in the sentence is α.



Head-1:Disease

	2	型	糖	尿	病	胰	岛	素	促	泌	剂
2	0	0	0	0	1	0	0	0	0	0	0
型		0	0	0	0	0	0	0	0	0	0
糖			0	0	1	0	0	0	0	0	0
尿				0	0	0	0	0	0	0	0
病					0	0	0	0	0	0	0
胰						0	0	0	0	0	0
岛							0	0	0	0	0
素								0	0	0	0
促									0	0	0
泌										0	0
剂											0

Head-2:Class

	2	型	糖	尿	病	胰	岛	素	促	泌	剂
2	0	1	0	0	0	0	0	0	0	0	0
型		0	0	0	0	0	0	0	0	0	0
糖			0	0	0	0	0	0	0	0	0
尿				0	0	0	0	0	0	0	0
病					0	0	0	0	0	0	0
胰						0	0	0	0	0	0
岛							0	0	0	0	0
素								0	0	0	0
促									0	0	0
泌										0	0
剂											0

Fig. 2. Subject tagger model architecture

Then, the relative positional information using ROPE [22] is integrated into the entity representation. The relative position of Rope satisfies ${R_i}^{\mathrm{T}} R_j = R_{j-i}$. The equation is as follows:

$$s_\alpha(i,j) = \left(R_i q_{i,\alpha}\right)^{\mathrm{T}} \left(R_j k_{j,\alpha}\right) = {q_{i,\alpha}}^{\mathrm{T}} {R_i}^{\mathrm{T}} R_j k_{j,\alpha} = {q_{i,\alpha}}^{\mathrm{T}} R_{j-i} k_{j,\alpha} \tag{5}$$

3.3 Feature enhancement layer

To enhance the representation of complex and nested semantic structures in electronic medical records (EMRs), we propose a Multidimensional Feature Enhancement Layer (MFEL), which is designed to effectively model both global context and multiscale local features. The MFEL module consists of two core components: a Bidirectional GRU (BiGRU) for capturing global semantic dependencies and a stacked Parallel Convolutional Feature Extraction (PCFE) structure for multiscale contextual representation.

First, the input token sequence is encoded using a BiGRU layer to model long-distance dependencies in both forward and backward directions. The output hidden states from BiGRU are denoted as:

$$H = BiGRU(H_1) \tag{6}$$

The output H is subsequently fed into a stack of three PCFE blocks, each of which is designed to extract and refine semantic features at multiple scales. The feature extraction process of PCFE is illustrated in **Fig. 3.**

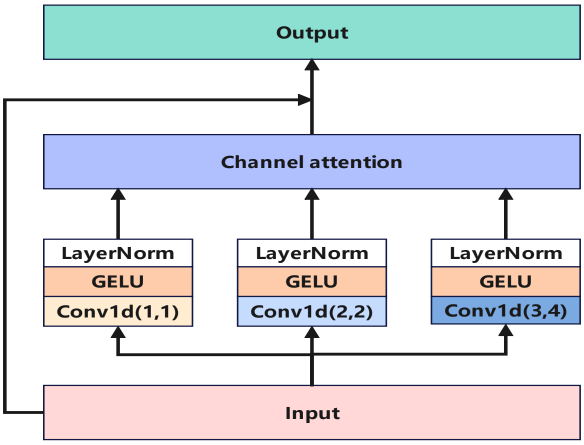


Fig. 3. Construction of PCFE module

Each PCFE block contains three parallel dilated convolution [23] branches with kernel sizes of 1×1, 1×2, and 1×3, and corresponding dilation rates of 1, 2, and 4, respectively, enabling the model to simultaneously capture fine-grained and long-range patterns. The equation is as follows:

$$F_1 = GLUE\big(Conv1D(x, k = 1, d = 1)\big) \tag{7}$$

$$F_2 = GLUE\big(Conv1D(x, k = 2, d = 2)\big) \tag{8}$$

$$F_3 = GLUE\big(Conv1D(x, k = 3, d = 4)\big) \tag{9}$$

Where k represents the kernel size, and d denotes the dilation rate.

To further enhance feature relevance and suppress noise in the inherently redundant EMR text, we apply a channel attention mechanism after convolution. This allows the model to assign higher weights to more informative features across branches. Following attention-based feature selection, we incorporate residual connections to preserve essential semantic information and improve training stability. Equations are as follows:

$$\alpha = \sigma\left(W_2\delta\big(W_1 \cdot GAP(F_{concat})\big)\right) \tag{10}$$

$$F_{att} = \alpha \otimes F_{concat} \tag{11}$$

$$H_{l+1} = H_l + F_{att} \tag{12}$$

Where GAP is global average pooling, δ is ReLU activation, σ is Sigmoid function, and $\alpha \in [0,1]$ is channel weight vector.

This design is to address the grid effect, which often arises in dilated convolutions with larger dilation rates. This effect can cause semantic discontinuity and the loss of important local details. To mitigate this problem, the model adopts a 1-2-4 dilation strategy across three parallel branches and further enhances the receptive field through the stacking of multiple PCFE blocks. This approach enables the model to gradually expand its receptive field while preserving local continuity and ensuring the integrity of semantic information.

To effectively integrate global contextual information captured by the BiGRU and the multi-scale semantic features extracted by the stacked PCFE blocks, we employ a gating mechanism for adaptive feature fusion. The formulation is provided below:

$$g = \sigma\left(W_g[H_3 : H]\right) \tag{13}$$

$$H_{out} = LayerNorm\left(g \odot F_{enhanced} + (1 - g) \odot F_{gru}\right) \tag{14}$$

3.4 Relation-specific object tagger

Following the feature enhancement layers, the model proceeds to the object-relation tagging stage, which leverages the previously extracted subject to identify corresponding objects for each predefined relation type. During inference, subject information is explicitly incorporated to guide the object extraction process. To accurately determine relational boundaries, two binary classifiers are employed to predict the start and end positions of each object. The formal formulation is presented below.

$$p_i^{start0} = sigmoid\left(W_{start}^r\left(x_i{}' + v_{sub}\right) + b_{start}^r\right) \tag{15}$$

$$p_i^{end0} = sigmoid\left(W_{end}^r\left(x_i{}' + v_{sub}\right) + b_{end}^r\right) \tag{16}$$

Where $p_i{}^{start0}$ and $p_i{}^{end0}$ present the probabilities of identifying the i^{th}word in the input vector as the start and end positions of the object, respectively.v_{sub} represents the vector representation of the encoding detected as the k^{th}subject-object in the subject labeling module. Similarly, object o can be denoted as:

$$p_{\Phi_r}(o|s,x) = \prod_{t \in (start_0, end_0)} \prod_{i=1}^{L} (p_i^t)^{I\{y_i^t=1\}} (1 - p_i^t)^{I\{y_i^t=0\}} \tag{17}$$

Where $start_0$ and end_0 denote binary tags for the start and end positions of the object, and Φ_r represents relation-specialized object decoder parameters.

4 Analysis of experimental results

4.1 Dataset

Private dataset The dataset used in this study is from the private medical data set of Diarel, including 608 electronic medical records of diabetes patients collected from a top hospital in Zhejiang Province, China. Each electronic medical record covers detailed clinical information such as the patient's main complaint, current medical history, past medical history, examination results, and preliminary diagnosis. To standardize the definition of entities and relationships, we referred to the "Commonly Used Clinical Medical Terms (2019 Edition)" released by the National Health Commission and the annotation methods for entity relationship extraction tasks in the CBLUE, combined with real clinical contexts, and finally defined six types of entity relationships. The data annotation work is completed by professional doctors with medical backgrounds, using the open-source annotation tool doccano to annotate entities and relationships, ensuring the consistency of annotation quality and clinical semantics. The entire dataset is divided into training and testing sets in an 8:2 ratio, as shown in Table 1 for specific statistical information.

Table 1. Statistics of the relation type.

Relation	Meaning	Example	Count
Sydi	Symptoms induced by the disease	<Foamy urine, Sydi, Diabetes>	1451
Drdi	Medications required for disease	<Metformin, Drdi, Diabetes>	1704
Dodr	Dosage of medication	<0.5g, Dodr, Metformin >	1055
Medr	Drug usage methods	< Subcutaneous injection, Medr, Insulin >	487
Evex	Results of the examination	<6mmoL/L, Evex, Fasting blood glucose >	1225
Exdi	The necessary examinations for the disease	<Postprandial Blood Glucose, Drdi, Diabetes>	1350
Total		7272	

Public dataset CMeIE-V2 is a large-scale benchmark dataset for Chinese medical information extraction, constructed from electronic medical records (EMRs) spanning multiple clinical departments, including pediatrics, obstetrics and gynecology, laboratory medicine, and pathology. The dataset defines 53 types of medical entity relationships, encompassing a wide range of clinical knowledge such as disease diagnosis, symptom manifestation, treatment procedures, and examination items, thereby supporting fine-grained relation extraction research. CMeIE-V2 consists of 14339 training instances, and 3585 validation instances, providing sufficient data volume for robust model development and evaluation. Its multi-source, heterogeneous EMR composition and rich relationship schema make it a valuable benchmark for assessing the cross-domain generalization capabilities of Chinese medical information extraction models.

DiaKG [24] focuses on the construction of medical diagnosis knowledge graphs, specifically targeting diabetes. It is compiled from 41 expert consensus documents written in Chinese, and contains 15 types of medical entity relations, with a total of 8,641 relation triples. This dataset supports structured knowledge modeling and relation extraction in the diabetes domain.

4.2 Evaluation index and experimental environment

Triple extraction requires identifying the entity scope and the corresponding relation type. Therefore, a triple is considered accurate only when the subject, relation, and object are correct. We use the commonly adopted precision (P), recall (R), and F1 score (F1) as evaluation metrics. The formula is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{20}$$

Where TP denotes the number of correctly extracted triplets, FP refers to the number of incorrectly extracted triplets, and FN indicates the number of cases where the triplet exists but was not correctly identified by the model.

The experiments were conducted on a server configured with an INTEL 13600KF CPU, an NVIDIA GeForce RTX 4070tisupr GPU, and 32 GB of memory. The learning rate is 3e-5; The batch size is 4; The dropout is 0.5; the GPU hidden size is 384; the Epoch is 150, and the Optimizer is Adam.

4.3 Comparative experiments

In order to evaluate the effectiveness of the GafRel model, we conducted experiments on the CMeIE-v2, DiaKG, and DiaRel datasets. A comparative analysis was performed against several state-of-the-art relation extraction models, including:

CasRel [13]: A cascade decoding model that first extracts the subject and then predicts its associated relations and objects.

Gplinker [6]: A model that treats every word pair as a potential relation candidate and efficiently screens valid relationships using matrix calculations.

OneRel [15]: A framework that simplifies relation extraction into a label classification task for each entity, directly annotating both entities and their relationships.

UIE [25]: A unified information extraction system that concurrently models tasks like entity, relation, event, and sentiment extraction to enhance generalization.

BiRTE [26]: An enhanced cascade approach employing a bidirectional extraction mechanism to capture triples from both directions and mitigate error propagation.

Chatgpt-O3 mini: The lightweight version of the ChatGPT series models proposed by OpenAI has reasoning ability and performs well in language understanding tasks.

Table 2 compares model performance on the DiaRel dataset, showing that GafRel outperforms all competing approaches in relation extraction. To assess generalization, we further evaluated GafRel on CMeIE-v2 and DiaKG against CBlue, CasRel, GPLinker, OneRel, and BiRTE. The results in Table 3 confirm its robustness and effectiveness across diverse datasets.

Table 2. Results of the models on DiaRel

Method	P (%)	R (%)	F1 (%)
Casrel	**93.78**	62.02	74.66
Birte	33.55	26.27	29.61
Gplinker	88.68	69.77	78.09
Onerel	87.37	71.42	78.59
UIE	64.97	**98.64**	78.34
Chatgpt-O3mini	79.56	76.88	78.20
GafRel(Our)	91.74	77.44	**83.98**

Table 3. Results of the models on CMeIE-V2 and DiaKG

Method	CMeIE-v2			DiaKG		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
Cblue	51.96	48.06	49.93	51.22	45.31	48.08
Casrel	57.29	39.11	46.48	63.56	30.96	41.64
Birte	50.44	**51.03**	50.73	19.46	6.83	10.11
Gplinker	55.72	50.12	52.77	58.78	**47.38**	52.47
Onerel	58.19	48.00	52.60	62.24	45.80	52.76
GafRel(Our)	**61.17**	47.35	**53.38**	**66.18**	44.77	**53.41**

Table 2 shows that the proposed GafRel model outperforms the baseline CasRel on the Diarel dataset, improving the F1 score by 9.32%, and recall by 15.42%. It also demonstrates competitive results against other models, confirming its superior triple extraction capabilities.

Further Analysis, CasRel adopts a subject-first strategy, meaning that errors in subject recognition propagate to relation and object extraction, reducing overall performance. Due to the presence of nested entities in EMRs, CasRel struggles with subject recall, leading to a low relation extraction recall rate. BiRTE attempts to improve the extraction framework of CasRel, but in small datasets, it does not show significant advantages in bidirectional extraction or mutual enhancement effects. OneRel and GPLinker represent two of the most advanced joint extraction architectures to date, leveraging sophisticated encoding, decoding schemes and a larger parameter footprint compared to earlier models. When entity pairs are far apart, The global attention mechanism of GPLinker struggles to attend to the relevant spans, resulting in missed extractions and a low recall on cross-sentence relations within EMRs. OneRel secured the second-highest performance, attributable to its unified labeling strategy that mitigates error propagation and its multi-head attention mechanism which adaptively prioritizes salient semantic spans. UIE, which employs a generative paradigm to enumerate all potential relations, exhibits elevated false positive rates in clinical settings due to the absence of a specialized filtering mechanism for fine-grained, domain-specific terminology.

The limited performance of ChatGPT-O3 Mini in the few-shot setting stems mainly from its insufficient capacity to model long-range dependencies and nested structures, which are prevalent in complex medical texts. Additionally, it struggles with accurately identifying overlapping entity boundaries in EMRs, leading to information loss. In contrast, our model benefits from multi-level contextual representation and enhanced nested entity recognition, which effectively address the challenges of long-range semantic dependencies and entity overlap in medical texts, thereby achieving superior relation extraction performance.

4.4 Ablation experiments

An ablation study was conducted to further ascertain the effects of the model components. The experimental results are shown in Table 4.

Table 4. Results of the ablation studies

Method	DiaKG			DiaRel		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
GafRel	66.18	44.77	53.41	91.74	77.44	83.98
-Global pointer	62.79	42.11	50.41	89.84	73.27	80.71
-Channel attention	64.53	44.87	52.93	91.63	76.35	83.30
-PCFE	64.56	42.40	51.19	91.36	72.47	80.82
-MFEL	62.53	40.19	48.93	89.16	71.90	79.60
Only Ernie	63.56	30.96	41.64	93.78	62.02	74.66

As shown in Table 4, removing the Global pointer led to missing nested entities and reduced accuracy in head entity recognition, resulting in notable declines in F1 scores: 3.00% on DiaKG, and 3.27% on DiaRel. These results underscore the importance of the global pointer in entity recognition and demonstrate its strong adaptability across different domains. Similarly, removing the MFEL module impaired the capacity of model to capture multi-scale contextual features, resulting in notable F1 score declines of 4.48% and 3.94% on DiaKG and DiaRel, respectively. These findings highlight the critical role of multi-scale semantic representation in effectively modeling complex medical relationships.

To further examine the internal structure of MFEL, we conducted a component-level ablation study. Removing the channel attention led to moderate performance declines of 0.48% and 0.86% on DiaKG and DiaRel. These results indicate that the attention module plays a vital role in filtering redundant information within EMRs by adaptively reweighting features extracted from each convolutional kernel. Additionally, eliminating the PCFE caused F1 scores to decrease by 2.22% and 3.16%, respectively. This indicates that this module is able to capture both local fine-grained features and global contextual dependencies. This multi-level receptive field design enhances the model’s understanding of complex medical semantics.

Finally, in the ablation setting where only the ERNIE is retained and all additional modules are removed, the model’s performance significantly declines across all datasets. It demonstrates that all modules contribute meaningfully.

4.5 Impact of the number of PCFE on result

Additionally, we explored the effect of varying the number of the PCFE. As shown in Table 5, the model's performance consistently improved as the number of layers increased from 0 to 3. However, beyond three layers, performance began to degrade due to increased model complexity.

The initial improvement is attributed to the expanded receptive field, which evolved from word-level features to cross-sentence contextual representations, enabling the model to better capture long-range dependencies. This resulted in notable gains in extracting long-distance relations. For instance, Drdi improved by 5.11%, Exdi by 2.59%, and Sydi by 8.15%. The overall F1 score increased by 3.16%, highlighting the effectiveness of multi-layer dilated convolutions in enhancing relational extraction.

Table 5. Impact of the number of the PCFE

Layers of number	Relation F1(%)						
	Sydi	Drdi	Dodr	Medr	Evex	Exdi	Total
0	84.99	66.23	90.34	74.35	96.40	55.45	80.82
1	87.50	68.87	89.58	74.14	95.21	54.34	81.24
2	88.48	70.22	90.81	72.57	97.09	55.91	82.41
3	93.14	71.34	89.76	79.50	95.96	58.04	83.98
4	89.42	71.21	90.88	72.32	96.38	55.75	82.62

However, increasing the number of layers beyond three introduced several challenges. As the network deepened, the channel attention mechanism became less effective at identifying and emphasizing key information, leading to a more dispersed allocation of attention weights. At the same time, important semantic features were gradually weakened during propagation through deeper layers, reducing the model's ability to preserve critical relational signals. Additionally, for short-range relations, excessive depth introduced unnecessary contextual noise, which further undermined extraction accuracy.

Consequently, the best performance was achieved using three PCFEs, which balance contextual feature extraction and attention stability. The first layer enhances entity features, the second captures potential relationships, and the third extracts global features to model long-range dependencies.

4.6 Efficiency analysis

In EMR relation extraction, both training efficiency and GPU use are critical. To this end, we compared the efficiency of our model with the best-performing OneRel model on the above dataset. To ensure a fair comparison, all experiments use a batch size of 4 and a maximum sequence length of 512. "Train" represents the training speed (it/s), "Inference" denotes inference speed (it/s), "Dedicated GPU" and "Sharing GPU" represent GPU memory usage. The results are shown in Table 6.

Table 6. The cost of the model on the DiaRel

Model	Train(it/s)	Inference(it/s)	Dedicated GPU(G)	Sharing GPU(G)
Gafel	1.43	17.87	9.7	0.1
Onerel	1.02	20.21	15.3	7.7

As shown in Table 6, our model achieves a training speed of 1.4 times faster than OneRel, while requiring only 42% of the GPU memory used by OneRel. Although the inference speed of our model is slightly lower than OneRel, it produces a higher F1 score, indicating a more accurate extraction performance. This compromise is acceptable in clinical applications.

We analyzed that this is because our model uses subject extraction and corresponding relationship extraction at the decoding end, reducing the dimensionality of the output tensor and parameter redundancy. The OneRel model, on the other hand, has a higher dimensionality and a greater computational burden due to the single-step decoding method, which results in a score matrix that contains both entity and relation components. This unified modeling strategy requires computing interaction scores for all possible entity pairs across all relation types simultaneously, leading to significantly increased memory consumption and training time.

Overall, our model demonstrates strong computational efficiency, achieving both lower resource consumption and improved performance. Such characteristics make it especially suitable for deployment in clinical settings, where both computational capacity and accuracy are critical.

5 Conclusion and future work

In this paper, we proposed GafRel, a novel joint entity-relation extraction framework tailored for Chinese EMRs. By integrating a Global Pointer mechanism into the CasRel backbone, GafRel effectively addresses the challenge of nested entity recognition in head entity identification. Furthermore, we introduced the Multidimensional Feature Enhancement Layer (MFEL), which combines parallel dilated convolutions with varying kernel sizes and dilation rates to capture multi-granular semantic patterns. The incorporation of channel attention further enhances the model's capacity for feature refinement and stable learning.

We also constructed a new dataset, DiaRel, based on the EMR of 608 hospital patients to support relation extraction. Experimental results on the CMeIE v2, DiaKG, and DiaRel datasets show that GafRel consistently obtains better F1 scores compared to the existing baseline.

In future work, we plan to extend this study in several directions. Specifically, we will continue to expand our private EMR dataset by annotating additional clinical records and incorporating a broader range of disease types. This effort aims to enhance the model's generalization ability across diverse medical scenarios. In addition, we intend to explore lightweight model architectures and optimization techniques to facilitate real-time inference and deployment in resource-constrained clinical environments. We also plan to investigate the integration of external medical knowledge to further improve relation extraction accuracy and support more interpretable clinical decision-making.

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

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

References

1. Lamy, M., Pereira, R.: Extracting clinical information from electronic medical records. In: Ambient Intelligence–Software and Applications–, 9th International Symposium on Ambient Intelligence, pp. 113-120. Springer International Publishing (2019)
2. Yue, H., Feng, X., Xu, H.: Enhancing Nested Named Entity Recognition in Chinese EMRs: A Collaborative Approach. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine, pp. 2821-2828. IEEE (2024)
3. Nelson, C. A., Bove, R.: Embedding electronic health records onto a knowledge network recognizes prodromal features of multiple sclerosis and predicts diagnosis. Journal of the American Medical Informatics Association 29(3), 424-434 (2022)
4. Zhang, Y. F., Gou, L., Zhou, T. S.: An ontology-based approach to patient follow-up assessment for continuous and personalized chronic disease management. Journal of biomedical informatics, 72, 45-59 (2017)

5. Sousa, D.: Biomedical relation extraction with knowledge graph-based recommendations. IEEE Journal of Biomedical and Health Informatics, 26(8), 4207-4217 (2022)
6. Su, J., Murtadha, A., Pan, S.,Liu, Y.: Global pointer: Novel efficient span-based approach for named entity recognition. arxiv preprint arxiv:2208.03054 (2022)
7. Huang, K., Singh, A., Chen, S.: Clinical XLNet: Modeling sequential clinical notes and predicting prolonged mechanical ventilation. arxiv preprint arxiv:1912.11975 (2019)
8. Zhou, G., Su, J., Zhang, J., & Zhang, M.: Exploring various knowledge in relation extraction. In Proceedings of the 43rd annual meeting of the association for computational linguistics, pp. 427-434 (2005)
9. Liao, G., Tang, X.: A Review of Relationship Extraction Based on Deep Learning. In: International Artificial Intelligence Conference, pp.73-84. Springer Nature Singapore. (2023)
10. J. Cheng, K. Long, S. Zhang, T. Zhang, L.: Text-image scene graph fusion for multi-modal named entity recognition, IEEE Trans. Artif. Intell (2023)
11. Li, Z., Yang, Z., Shen, C.: Integrating shortest dependency path and sentence sequence into a deep learning framework for relation extraction in clinical text. BMC medical informatics and decision making, 19, 1-8 (2019)
12. Hong, Y., Liu, Y., Zhang, K.: Improving graph convolutional networks based on relation-aware attention for end-to-end relation extraction. IEEE Access, 8, 51315-51323 (2020)
13. Wei, Z., Su, J., Wang, Y., Tian, Y.,Chang, Y.: A novel cascade binary tagging framework for relational triple extraction. arxiv preprint arxiv:1909.03227 (2019)
14. Wang, Y., Yu, B., Zhang, Y., Liu, T., Zhu, H.: TPLinker: Single-stage joint extraction of entities and relations through token pair linking. arxiv preprint arxiv:2010.13415 (2020)
15. 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, Vol. 36, No. 10, pp. 11285-11293 (2022)
16. Devlin, J., Chang, M. W., Lee, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1, pp. 4171-4186 (2019)
17. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arxiv preprint arxiv:1907.11692 (2019)
18. Alsentzer, E., Murphy, J. R., Boag, W., Weng, Naumann, T.: Publicly available clinical BERT embeddings. arxiv preprint arxiv:1904.03323 (2019)
19. Uzuner, Ö., South, B. R., Shen, S.:2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association, 18(5), 552-556 (2011)
20. Zhang, N., Chen, M., Bi, Z., Liang, X., Li, L.: Cblue: A chinese biomedical language understanding evaluation benchmark. arxiv preprint arxiv:2106.08087 (2021)
21. Sun, Y., Wang, S., Li, Y., Feng, S., Chen, X., Zhang, H.: Ernie: Enhanced representation through knowledge integration. arxiv preprint arxiv:1904.09223 (2019)
22. Su, J., Ahmed, M., Lu, Y., Pan, S.: Roformer: Enhanced transformer with rotary position embedding. Neurocomputing, 568, 127063 (2024)
23. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arxiv preprint arxiv:1511.07122 (2015)
24. Chang, D., Chen, M.: Diakg: An annotated diabetes dataset for medical knowledge graph construction. In: Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction: 6th China Conference, CCKS 2021, Proceedings 6 (pp. 308-314). Springer Singapore. Guangzhou, China (2021)

25. Lu, Y., Liu, Q., Dai, D., Wu, H.: Unified structure generation for universal information extraction. arxiv preprint arxiv:2203.12277 (2022)
26. Ren, F., Zhang, L., Zhao, X., Yin, S.: 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 (2022)