

Article

Not peer-reviewed version

Agricultural Meteorological Disaster Relation Extraction Technology Based on BERT and Adaptive Denoising Graph Network

[Yonglei Li](#), [Nengfu Xie](#)^{*}, Xin Wang, Qian Xu, [Xiaoli Wang](#)^{*}, Rui Man, Hailong Liu

Posted Date: 27 January 2026

doi: 10.20944/preprints202601.2029.v1

Keywords: relation extraction; knowledge extraction; multimodal fusion; graph convolutional network; adaptive denoising



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Agricultural Meteorological Disaster Relation Extraction Technology Based on BERT and Adaptive Denoising Graph Network

Yonglei Li ^{1,2}, Nengfu Xie ^{1,2,*}, Xin Wang ^{1,2}, Qian Xu ^{1,2}, Xiaoli Wang ^{1,2,*}, Rui Man ^{1,2} and Hailong Liu ^{1,2}

¹ Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China

² Key Laboratory of Agricultural Blockchain Application, Ministry of Agriculture and Rural Affairs, Beijing, China

* Correspondence: xienengfu@caas.cn (N.X.); wangxiaoli@caas.cn (X.W.); Tel.: +13970977281 (N.X.); +15201120079 (X.W.)

Abstract

The increasing frequency of extreme weather events has resulted in a rapid growth of unstructured meteorological texts, posing challenges for accurate disaster-related information extraction. Existing relation extraction models that integrate semantic and syntactic information via Graph Convolutional Networks (GCNs) often suffer from noise amplification and information redundancy. To address these limitations, we propose a domain-specific relation extraction framework that integrates Bidirectional Encoder Representations from Transformers (BERT) with an Adaptive Denoising Graph Network. A dual-tuning architecture is designed to jointly model semantic and syntactic features. Specifically, a Self-Attention Graph Convolutional Network (SA-GCN) dynamically prunes irrelevant dependency structures, while a Residual Shrinkage Network (RS-Net) performs fine-grained feature denoising through sample-adaptive threshold learning. Experiments on the public DUIE 2.0 dataset and a newly constructed Agricultural Meteorological Disaster dataset demonstrate that the proposed model achieves F1 scores of 91.51% and 92.84%, respectively, outperforming state-of-the-art methods. Ablation studies further confirm the effectiveness of adaptive denoising in handling complex linguistic structures in disaster reports. The proposed approach enhances the robustness and accuracy of meteorological relation extraction and provides technical support for knowledge graph construction and intelligent disaster early warning systems.

Keywords: relation extraction; knowledge extraction; multimodal fusion; graph convolutional network; adaptive denoising

1. Introduction

1.1. Research Background

Global climate change is accelerating, and the frequency of extreme weather events is increasing significantly. Agricultural meteorological disasters such as droughts, floods, typhoons, and frosts have become major factors threatening agricultural production and food security. Against the backdrop of global warming, both the frequency and intensity of meteorological disasters in China have risen markedly.

These disasters have a severe impact on the stable prosperity of natural ecosystems and the secure development of the political, economic, and social spheres. Therefore, research and work related to meteorological disaster warning and prevention have become particularly important. To effectively cope with agricultural meteorological disasters, China has formulated and introduced a series of related policies and measures in recent years. For example, the Central No. 1 Document(the

first and most prominent policy statement released annually by the Central Committee of the Communist Party of China and the State Council, which has prioritized agricultural and rural development for over two decades) has made agricultural meteorological disaster monitoring and early warning a key work point for many consecutive years (2023), and the Ministry of Agriculture and Rural Affairs has issued important documents such as the “Technical Guidelines for Risk Assessment and Zoning of Agricultural Meteorological Disasters” (2024). The aim is to enhance independent innovation capabilities in the meteorological field through cutting-edge technologies like artificial intelligence, big data, and quantum computing, and to build a meteorological disaster risk assessment and decision-making information support system with all-weather response capabilities.

Although the state has introduced multiple policies to encourage and support the monitoring, warning, and prevention of agricultural meteorological disasters, the massive growth of text data in the meteorological disaster field has led to low operational efficiency when relying on manual integration and analysis of meteorological data. This results in significant consumption of human and material resources, increasing operational costs. Furthermore, traditional methods in the meteorological disaster field struggle to efficiently extract key information from massive, fragmented, or heterogeneous data. This shortcoming can no longer meet the current objective demand for high-efficiency data processing capabilities.

To achieve data-intensive knowledge discovery and application, the Knowledge Graph (KG) has emerged as an important tool [1–3]. It can visually describe various entities and their corresponding relationships in the world and perform tasks such as knowledge discovery. In recent years, it has been widely used in fields like recommendation systems, intelligent question answering, and sentiment analysis [4].

Moreover, to extract entities and relationships from source texts, Natural Language Processing (NLP) techniques typically require domain experts to build rule templates and perform manual annotation for tasks like Named Entity Recognition (NER) and Relation Extraction (RE) [5–9]. With the rapid iteration of storage media and computing devices has led to a surge in computing power, feature learning methods represented by deep learning have begun to develop rapidly (Li Ying, Chen Huailiang, 2020) [10,11]. Combining deep learning and natural language processing to complete knowledge graph construction has become a research hotspot [12–16].

However, in the field of agricultural meteorological disasters, although intelligent processing methods and sufficient scientific data provide a basis for NER and RE of agricultural meteorological disasters, research on meteorological disaster knowledge graphs based on NLP is relatively scarce. Therefore, this paper, starting from the methodology of knowledge acquisition, integration, and analysis, integrates data from the meteorological disaster field. It builds the core technology of a knowledge graph based on entity relation extraction and named entity recognition processes, extracting entities associated with agricultural meteorological disasters from massive text data to form a high-quality knowledge graph. Subsequently, a knowledge-based question answering system for agricultural meteorological disasters will be developed to provide references for the decisions of agricultural producers, government departments, and scientific researchers, offering intelligent method support for promoting high-quality meteorological development.

1.2. Our Contributions

This paper presents a multimodal relation extraction model that fuses syntactic pruning with adaptive denoising to facilitate the construction of a domain-specific Knowledge Graph. The core contributions are as follows:

A Dual-Path Multimodal Fusion Framework: We propose a parallel architecture that extracts and fuses semantic and syntactic features. This approach ensures that the model captures both the linguistic meaning and the structural dependencies essential for accurate relation extraction in complex meteorological texts.

Self-Attention Graph Convolutional Network (SA-GCN) for Dynamic Pruning: To resolve the issue of information redundancy, we innovate upon the standard GCN by integrating a self-attention mechanism. Unlike traditional methods that aggregate neighbors equally, the SA-GCN dynamically calculates neighbor weights to prune irrelevant dependencies. This focuses the model on core syntactic paths, significantly improving computational efficiency and extraction precision.

Residual Shrinkage (RS) for Adaptive Denoising: We introduce a Residual Shrinkage Network into the NLP domain to address noise propagation. By learning adaptive thresholds, the RS module suppresses weak noise features while retaining strong signals. This enhances the model's robustness against the inherent noise found in automated text processing.

2. Related Work

2.1. Information Extraction in Agricultural Meteorological Disasters

With the development of precision agriculture, applying Natural Language Processing (NLP) technology to extract knowledge from unstructured meteorological texts has become a research hotspot. Early research mostly relied on manual rule construction or statistical machine learning methods to extract disaster information. For instance, Qiu et al. (2018) [2] constructed a meteorological agricultural knowledge graph based on semi-structured data, validating the feasibility of knowledge organization for disaster warning. However, these traditional methods rely heavily on manual feature engineering and struggle to generalize to massive, heterogeneous texts.

In recent years, deep learning has been gradually applied to this field. Named Entity Recognition (NER) in agriculture has achieved significant progress. Zeng et al. (2025) [17] released the AgriCHN dataset and proposed a cross-domain NER model, effectively identifying agricultural entities. Specifically for meteorological disasters, Ye et al. (2024) [18] proposed a state information extraction method for typhoon disasters, utilizing deep neural networks to capture disaster attributes from news reports. Qiu et al. (2025) [19] further introduced a character-word fusion method to improve the accuracy of agricultural meteorological disaster event extraction.

Despite these advancements, existing research still faces two major unresolved problems:

Lack of Relation Extraction (RE) Studies: Most current works focus on entity recognition (NER) or simple event detection [1,20]. Research on extracting semantic relationships between disaster entities (e.g., "Drought" causes "Yield Reduction") is relatively scarce, limiting the reasoning capability of downstream knowledge graphs.

Sensitivity to Noise and Long Sequences: Disaster reports often contain long sentences with complex syntactic structures and redundant information (noise). Existing agricultural IE models [17,21] often lack effective mechanisms to filter out irrelevant syntax or denoise the features, leading to performance degradation when processing complex real-world reports [22–24].

2.2. Neural Relation Extraction Techniques

To address the aforementioned challenges in agricultural texts, recent advancements in general domain relation extraction provide potential solutions.

Semantic Encoding and Pre-training: Pre-trained language models like Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) [25] have revolutionized relation extraction by capturing rich contextual semantics through large-scale unsupervised training. Variants such as RoBERTa (Robustly optimized BERT approach) further optimized the training strategy [26]. Other models like ALBERT [27], DeBERTa [28], and ELECTRA [29] have also been explored [30–34]. However, BERT-based methods mainly focus on semantic sequences and often struggle to capture long-distance dependencies between entities in complex sentences [35,36].

Syntactic Enhancement via Graph Neural Networks (GNNs): To capture long-distance dependencies, Graph Convolutional Networks (GCNs) are widely used to model the dependency tree of sentences [37–44]. Zhang et al. (2018) [45] proposed Contextualized Graph Convolutional Network (C-GCN) to facilitate information propagation over dependency paths. To reduce the noise

from erroneous dependency parsing, Guo et al. (2019) [46] proposed Attention Guided Graph Convolutional Network (AGGCN), which uses a “soft pruning” strategy to weigh neighbor nodes. Nevertheless, these models typically use equal-weight aggregation or simple attention, which is insufficient to completely filter out the structural noise found in disaster reports.

Feature Denoising Mechanisms: Dealing with noisy data is a critical challenge in deep learning [47]. The “soft thresholding” mechanism, originally from signal processing, has been successfully applied to fault diagnosis. Zhao et al. (2019) [48] proposed the Residual Shrinkage Network (RSN), which uses an attention mechanism to learn adaptive thresholds and suppress noise features in vibration signals. Similar deep residual shrinkage networks have been applied in other domains, such as lithology identification[49] and anomaly detection of rotating machines [50]. In the NLP domain, although some works have attempted to use gating mechanisms for denoising, applying adaptive residual shrinkage to suppress syntactic noise in relation extraction remains an unexplored area.

3. Model Design and Implementation

The relation extraction model proposed in this paper is centered on an innovative Dual-Path Feature Fusion Architecture, aiming to organically combine the deep semantic understanding capabilities endowed by pre-trained language models with structured syntactic analysis based on graph neural networks. To clearly demonstrate the model’s internal workflow, the following figure presents the complete processing flow from data input to relationship prediction.

3.1. Overall Model Architecture

The entire model can be deconstructed into four core modules, as shown in Figure 1.

(1) **Dual-Path Feature Extraction Module:** As the model’s data processing front-end, this module receives raw text sequences and works in parallel along two independent paths. The semantic path utilizes the powerful BERT pre-trained model to capture the deep, implicit contextual semantics within the text. The syntactic path uses a dependency parser to explicitly convert the text’s grammatical structure into a graph structure, laying the foundation for subsequent structured information processing.

(2) **Syntactic Enhancement Module:** This module is the first core innovation of the model, designed to address the “one-size-fits-all” limitation of standard graph networks. With the Self-Attention Graph Convolutional Network (SA-GCN) at its core, it receives semantic features and the syntactic graph from the dual-path module. By introducing a self-attention mechanism and a dynamic pruning strategy during the graph convolution process, this module can intelligently identify and amplify key syntactic dependencies while filtering out redundant information, achieving preliminary syntactic enhancement of the original semantic features.

(3) **Feature Denoising Module:** This is the second core innovation, aimed at overcoming the inherent noise amplification effect in graph network information propagation. This module receives features from the syntactic enhancement module and introduces the Residual Shrinkage Network (RS-Net) for the first time. Through an innovative adaptive threshold learning mechanism, RS-Net can “tailor” a denoising solution for each sample, precisely suppressing noise features to output highly purified syntactically enhanced features.

(4) **Fusion and Classification Module:** As the model’s back-end, this module is responsible for integrating information and making the final decision. It deeply fuses the original semantic features from the first module with the refined syntactic features [51] from the third module. To enhance the model’s robustness and generalization, we designed a dual-classifier joint training strategy, optimizing through an interpolation loss function to finally output the relationship category between entity pairs. Data flows sequentially through it, forming a complete processing pipeline from “feature extraction” to “fusion and classification.”

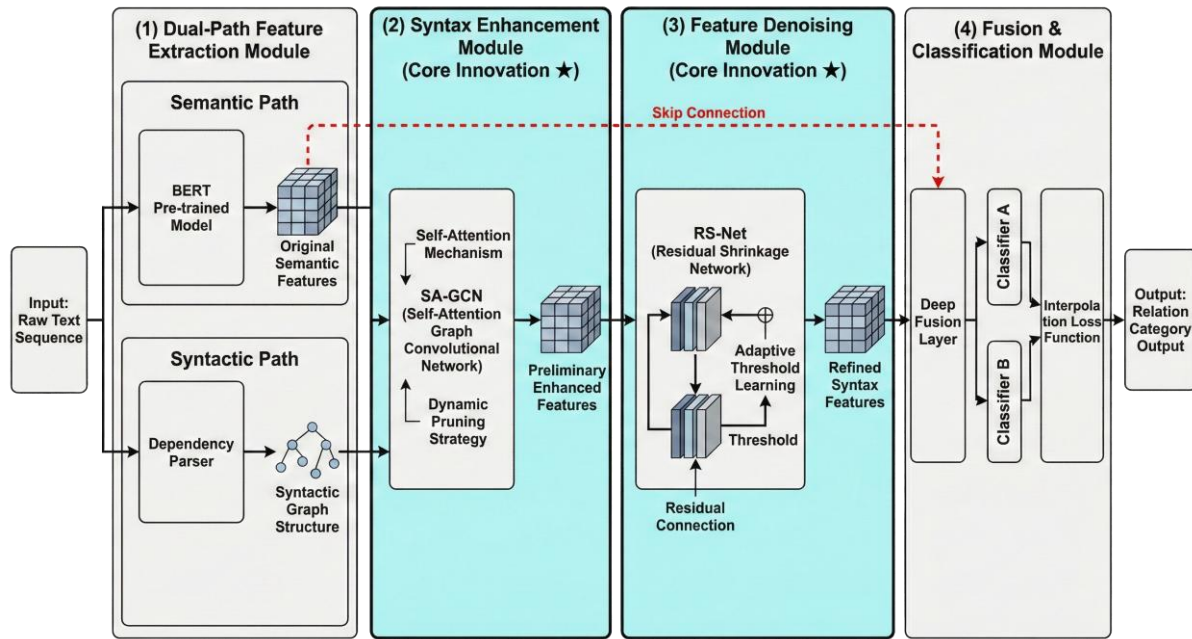


Figure 1. Model architecture diagram.

3.2. Dual-Path Feature Extraction Module

The BERT model, with its Transformer-based bidirectional encoding capability, has become a cornerstone in the field of natural language processing [25]. In this model, BERT undertakes the key task of converting discrete text symbols into dense, continuous semantic vector representations. To fully leverage BERT's pre-trained knowledge while better adapting it to the specific downstream task of relation extraction, we adopted a Hierarchical Freezing fine-tuning strategy. For a BERT-based model with 12 Transformer layers, the parameters of the first 8 layers are kept frozen during training, while the parameters of the last 4 layers participate in gradient updates.

BERT's lower layers (1-8) learn more universal lexical and basic grammatical knowledge. Freezing these layers effectively prevents "catastrophic forgetting" when fine-tuning on small-scale domain datasets, thus maximally preserving its generalization ability. The higher layers (9-12), on the other hand, learn more task-related abstract semantic representations. Finetuning them guides the model to focus its general language knowledge on the specific needs of relation extraction. After processing by this module, an input text sequence is encoded into a semantic feature matrix $H_{bert} \in R^{n \times d_{bert}}$, where n is the sequence length after BERT tokenization, and d_{bert} is the hidden layer dimension (typically 768). Each row vector in this matrix contains rich contextual semantic information for the corresponding token, serving as the initial node features for subsequent modules.

Although BERT can implicitly learn some syntactic information, explicit syntactic structures (like dependency trees) provide unambiguous, structured word-to-word dominance and modification relationships, acting as a powerful supplement to semantic information. For this, we designed a parallel syntactic path. Using the spacey dependency parsing tool, we parse the raw text into a dependency graph $G = (V, \varepsilon)$.

To construct a graph structure that both reflects core syntactic dependencies and possesses good connectivity, we adopted an enhanced adjacency matrix $A \in R^{n \times n}$ construction strategy:

$$A_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E(\text{preserve core dependencies}) \\ 1 & \text{if } i = j(\text{add self-loops for self-information}) \\ 1 & \text{if } |i - j| = 1(\text{connect adjacent tokens for robustness}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The considerations for this strategy are as follows: (1) Core dependency edges are the backbone of syntactic information; (2) Adding self-loops ensures each node can pass its own information to the

next layer during graph convolution; (3) Connecting adjacent tokens is an effective robustness measure, compensating for potential “broken chains” when the parser handles complex or non-standard sentences.

A key technical challenge that must be addressed is Token Alignment. Due to the mismatch between BERT’s sub-word tokenization granularity and the parser’s word-level granularity, we implemented a precise alignment algorithm. This algorithm lossless maps word-level dependency relationships onto the sub-word sequence’s adjacency matrix, ensuring dimensional consistency and semantic correspondence between the semantic feature matrix H_{bert} and the syntactic adjacency matrix A .

3.3. Syntactic Enhancement Module

Standard Graph Convolutional Networks (GCNs) use Laplacian smoothing when aggregating neighbor information [44], assigning the same weight to all neighbors. This ignores the varying contribution of different syntactic dependencies to judging a specific relationship. To address this “equal-weight aggregation” limitation, we designed the Self-Attention Graph Convolutional Network (SA-GCN).

In each layer of information propagation, SA-GCN introduces a self-attention mechanism [52–55]. First, the model still aggregates neighborhood information through a basic graph convolutional operation:

$$H^{(1)} = A^{(1-1)}H^{(1-1)}W^{(1)} \quad (2)$$

Subsequently, instead of directly passing $H^{(1)}$ to an activation function, it is used as input to a multi-head self-attention module. This module computes a context-aware attention weight distribution for each node in the graph. This allows the model to “focus” on neighbors with more information and higher relevance when updating each node’s representation, thereby achieving differentiated, weighted information aggregation.

In real-world text, a complete dependency tree contains many dependencies that are redundant or even irrelevant for judging a specific relationship. Retaining these edges not only increases the computational burden but may also introduce noise. To this end, we proposed a dynamic graph pruning strategy. After each SA-GCN layer’s computation, we use a small feed-forward network to calculate an “importance score” $s_i^{(1)}$ for each node. This score reflects the node’s potential contribution to the final task in the current semantic and structural context.

Based on this score, we iteratively remove the nodes (and all their connections) ranking in the bottom 20% of scores in each layer. This pruning threshold was determined empirically based on the performance on the validation set. Experiments demonstrated that setting the ratio to 20% achieved the optimal trade-off between suppressing structural noise and preserving meaningful semantic connections. After three pruning layers, approximately 49% of the connections in the graph are removed, significantly reducing computational complexity. More importantly, this pruning mechanism can be viewed as a form of Structured Regularization. It forces the model to discard minor syntactic paths during information propagation and focus on the backbone structure that truly defines the core semantic relationship (such as the “subject-verb-object” path), thus effectively improving the model’s signal-to-noise ratio and generalization ability, as shown in Figure 2.

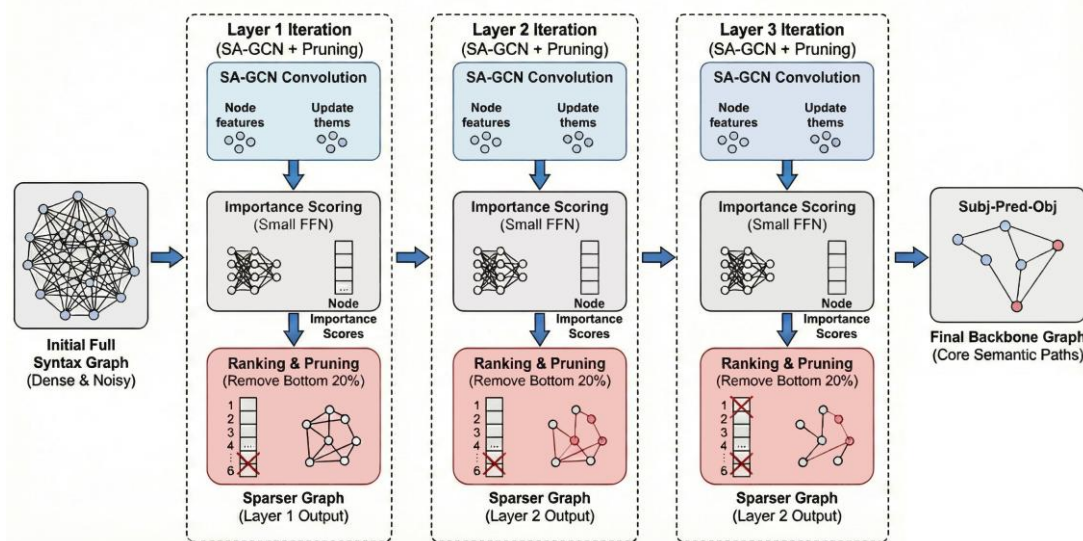


Figure 2. Syntactic enhancement module.

3.4. Feature Denoising Module: Residual Shrinkage Network

Although SA-GCN can focus on important information, the multi-layer propagation nature of graph convolution can still cause noise, present in the initial features (originating from BERT representation biases or parsing errors) to be amplified and diffused layer by layer. To precisely suppress this type of noise, we innovatively introduced the Residual Shrinkage Network (RS-Net), as shown in Figure 3.

Traditional denoising methods often use fixed thresholds that require manual tuning, making it difficult to adapt to the vastly different noise distributions of different samples. The core breakthrough of RS-Net is its adaptive threshold learning mechanism. It contains a lightweight sub-network dedicated to learning the threshold. This sub-network first compresses the input feature matrix X into a global vector representing the entire sentence's feature distribution via Global Average Pooling (GAP). Subsequently, this global vector is fed into a three-layer MLP. The MLP's output, activated by a Sigmoid function, generates a denoising threshold τ , which is specific to the current sample and falls between $[0, 1]$.

$$\tau = \text{Sigmoid} \left(\text{MLP}(\text{GAP}(X)) \right) \quad (3)$$

This mechanism endows the model with the ability to dynamically adjust the denoising strength according to the specific situation of each sample, achieving highly intelligent and personalized noise suppression.

After obtaining the adaptive threshold τ , RS-Net employs the soft thresholding function as its core denoising operator. It is defined as:

$$\text{SoftThreshold}(x, \tau) = \text{sign}(x) \cdot \max(|x| - \tau, 0) \quad (4)$$

The working principle of this function, derived from signal processing theory, is as follows: it treats parts where the feature's absolute value is less than τ as noise and sets them directly to zero; for parts where the absolute value is greater than τ (considered valid signal), it subtracts the threshold amount, retaining but "shrinking" them. This approach is smoother than the hard thresholding method, which simply clips, and can better preserve the relative strength of the signal.

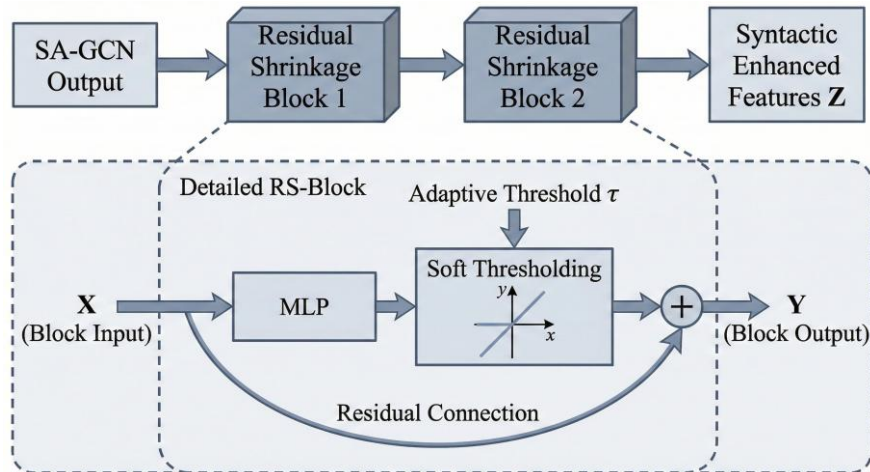


Figure 3. Feature denoising module.

To ensure the stability of the information flow during the denoising process and to avoid gradient vanishing, we embed the soft thresholding operation within a Residual Block. The processing flow of the entire residual shrinkage block is:

$$Y = \text{SoftThreshold}(\text{MLP}(X), \tau) + X \quad (5)$$

The model cascades two such residual shrinkage blocks to perform progressive, deep feature purification on the output of the SA-GCN, finally obtaining a high-quality syntactically enhanced feature matrix Z .

3.5. Fusion and Classification Module

After obtaining the clean semantic features Bert and the refined, syntactically enhanced features Z , the model extracts the representations for the entity pair, e_{bert} and e_{gcn} , by applying average pooling to the token vectors at the entity locations. Subsequently, an Early Fusion strategy is adopted, concatenating these two modal feature vectors and feeding them into a two-layer MLP for deep interaction and fusion. This MLP is not only used for dimensionality reduction but, more importantly, its non-linear transformation capability allows the model to learn the complex complementary and dependent relationships between semantic and structural features, as shown in Figure 4.

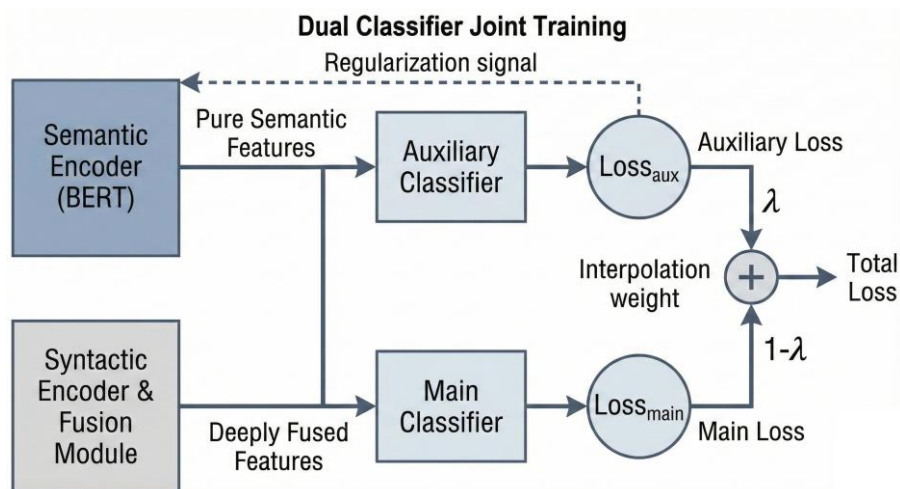


Figure 4. The overall architecture of the proposed model.

Main Classifier: Receives the deeply fused features fused and undertakes the primary classification task. Auxiliary Classifier: Receives only the pure semantic features bert for classification.

The motivation for this design is: the auxiliary classifier provides an additional supervisory signal, forcing the semantic encoding part of BERT to maintain its own strong semantic discrimination ability throughout the joint training of the entire model. This can be seen as a regularization technique, effectively preventing the model's semantic representation from "drifting" while learning syntactic features, thereby enhancing the model's baseline performance when syntactic information is missing or unreliable. The total loss function is a weighted sum of their respective cross-entropy losses, using an interpolation weight λ :

$$L_{total} = \lambda L_{main} + (1 - \lambda) L_{bert} \quad (6)$$

In this way, the model balances the advantages of multimodal fusion with the purity of the core semantic representation during optimization, achieving an overall performance improvement.

4. Experiments and Analysis

4.1. Datasets and Evaluation Metrics

This paper evaluates the performance of the proposed multimodal fusion relation extraction model based on BERT and an adaptive denoising graph network using two datasets: an agricultural meteorological disaster corpus collected from authoritative public sources and the widely used public dataset DUIE 2.0.

The agricultural meteorological disaster dataset is constructed by collecting and organizing textual data from official and authoritative channels, including governmental meteorological agencies, agricultural departments, and publicly released disaster reports. The dataset primarily covers information related to various agricultural meteorological disasters, such as droughts, floods, frost events, and extreme weather impacts on crops. It is used to assess the model's robustness and generalization ability in a specialized agricultural disaster domain.

DUIE 2.0, as a large-scale open-domain relation extraction benchmark dataset, is adopted to evaluate the proposed model's performance on general relation extraction tasks and to facilitate fair comparison with existing mainstream methods. The scale and key characteristics of the two datasets are summarized in Table 1.

Table 1. Statistical Comparison of Experimental Datasets.

Dataset	Language	Classes	Train	Test	Validation	Main Features
Agri-Disaster	Chinese	21	8,997	1,126	1124	Domain-specific, complex sentence
DUIE 2.0	Chinese	49	13,669	3,000	2996	General domain, large-scale

The agricultural meteorological disaster dataset's data is primarily sourced from authoritative literature such as the "Yearbook of Meteorological Disasters in China." We used the doccano annotation tool, and domain experts performed manual annotation. The final dataset contains approximately 11,247 sentences, 110,450 entities, and 67,590 relation instances.

DUIE 2.0 is a large-scale Chinese relation extraction dataset released by Baidu. Its text sources are extensive, including encyclopedias, news, and forums. The language style is colloquial, and it contains complex relationships, making it suitable for testing model performance on large-scale, general-domain data. The dataset includes 170,000 data entries; we selected 13,669 as the training set and 3,000 as the test set.

Evaluation Metrics: To assess model performance, we use the standard evaluation metrics for relation extraction tasks: Precision (P), Recall (R), and F1-score (F1). Their calculation formulas are as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (9)$$

where TP (True Positives) is the number of positive instances correctly predicted, FP (False Positives) is the number of negative instances incorrectly predicted as positive, and FN (False Negatives) is the number of positive instances incorrectly predicted as negative.

4.2. Experimental Environment and Parameters

The experimental research environment for this paper is based on an Intel(R) Xeon(R) Gold 6226R processor (CPU @ 2.90GHz), an NVIDIA RTX A6000 GPU, and the Microsoft Windows 10 IoT operating system. The experiments use the PyTorch deep learning framework and PyCharm as the Python development environment. Specific environment configurations are shown in Table 2.

Table 2. Experimental Environment.

Parameter Item	Parameter Value
Operating System	Microsoft Windows 10 IoT
CPU	Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz
GPU	NVIDIA RTX A6000
Memory	128GB
Programming Env	Python 3.11 + PyTorch 2.1.1

4.3. Comparative Experiments and Result Analysis

To verify the superiority of the proposed model, we selected the following representative baseline models for comparison:

1. R-BERT: A classic BERT-based relation classification model that combines entity information with marker representations for prediction, representing the baseline for pure semantic methods.
2. BERT-LSTM: Adds an LSTM layer on top of BERT, aiming to better capture the sequential features of the text [56,57]; a representative of sequence enhancement methods.
3. AGGCN: An attention-guided graph convolutional network, representing advanced technology for “soft pruning” on the complete dependency tree; a key baseline for comparison with our model’s structured information processing.
4. BiLSTM-CRF: A classic model widely used in NER and sequence labeling tasks, used to compare the effect without using pre-trained models and graph structures [58].

Implementation Details: All experiments in this study were completed within the PyTorch framework. The model’s core hyperparameter settings were determined through grid search on the validation set to ensure experimental reproducibility. (Note: Table 3 mentioned in the original text for hyperparameters is provided as “Experimental Results” instead. We present the results table as given.)

Table 3. Experimental Results.

Model	Agri-Disaster			DUIE 2.0		
	F1 (%)	R (%)	P (%)	F1 (%)	R (%)	P (%)
R-BERT	84.48	85.92	83.10	91.86	90.72	90.08
BERT-LSTM	91.28	90.57	90.47	90.28	89.57	89.47
AGGCN	91.87	91.32	92.59	91.87	91.32	92.59
Ours	92.84	91.02	90.05	91.51	91.35	90.97

4.4. Ablation Study

To verify the actual contribution of each innovative component in the model, we conducted a series of ablation experiments on the Agri-Disaster dataset. The results are shown in Table 4.

The ablation experiment results clearly reveal the value of each component:

1. Contribution of GCN: After adding a standard GCN to the BERT baseline, the F1 score slightly increased (+2.74%), demonstrating the preliminary effectiveness of introducing syntactic structure information.
2. Contribution of SA-GCN: Adding self-attention pruning (SAG) on top of GCN further improved performance (+3.30%), indicating that dynamically pruning irrelevant nodes in the syntactic graph can effectively reduce structural noise interference, thus aggregating information more precisely.
3. Contribution of RSN: Adding the Residual Shrinkage Network (RSN) to the model led to a significant performance leap (+2.32%). To verify that this improvement is not due to random chance, we conducted a paired t-test over 5 runs with different random seeds. The results show that our model statistically outperforms the baseline (p -value < 0.05). This strongly proves that feature-level noise is a key bottleneck affecting model performance. The adaptive soft thresholding denoising mechanism of RSN can extremely effectively purify feature representations, greatly enhancing classification accuracy. This validates the synergistic effect of the “dual denoising” architecture.
4. Contribution of Interpolation Loss: After replacing the full model’s interpolation loss with a standard cross-entropy loss, performance dropped (-0.17%). This shows that the training strategy of using an auxiliary classifier for regularization can improve the model’s stability and generalization ability, preventing it from overfitting to the noise in the syntactic branch.

Impact of Sentence Length on Performance: To verify the model’s advantage in handling long-distance dependencies, we compared the F1 score performance of our model and a pure sequential model (BERT) on different sentence lengths. On the DUIE 2.0 and SemEval 2010 datasets, as sentence length increased, the performance advantage of our model over BERT became more pronounced. Especially when the sentence length exceeded 100, the F1 score improvements reached 1.34% and 2.17%, respectively. This intuitively demonstrates that by establishing “shortcuts” on the syntactic structure via GCN and using the dual denoising mechanism to ensure pure information transmission, our model can more effectively capture dependencies between distant entities.

Table 4. Ablation Study Results.

Model Variant	F1 (%)	Δ
BERT (Baseline)	84.48	-
+ GCN (Introduce syntactic graph)	87.22	+2.74
+SA-GCN (Pruning)	90.52	+3.3
+ RSN (Full Model)	92.84	+2.32

Visualization of Noise Suppression Effect: intuitively demonstrated the effect of the dual denoising module through confusion matrices. The standard GCN model (left) had more lightly colored squares in the off-diagonal regions, representing a higher misclassification rate (noise). After adding the SA-GCN and RSN modules (right), the colors in the off-diagonal regions clearly deepened, indicating a significant reduction in misclassifications, with the model’s predictions becoming more concentrated on the diagonal. This provided qualitative evidence for the effectiveness of the dual denoising mechanism.

Figure 5. Confusion matrices on the Agri-Disaster test set. (a) The baseline model shows confusion between similar classes due to noise. (b) The proposed model (Ours) demonstrates a sharper diagonal, indicating effective noise suppression and higher classification accuracy.

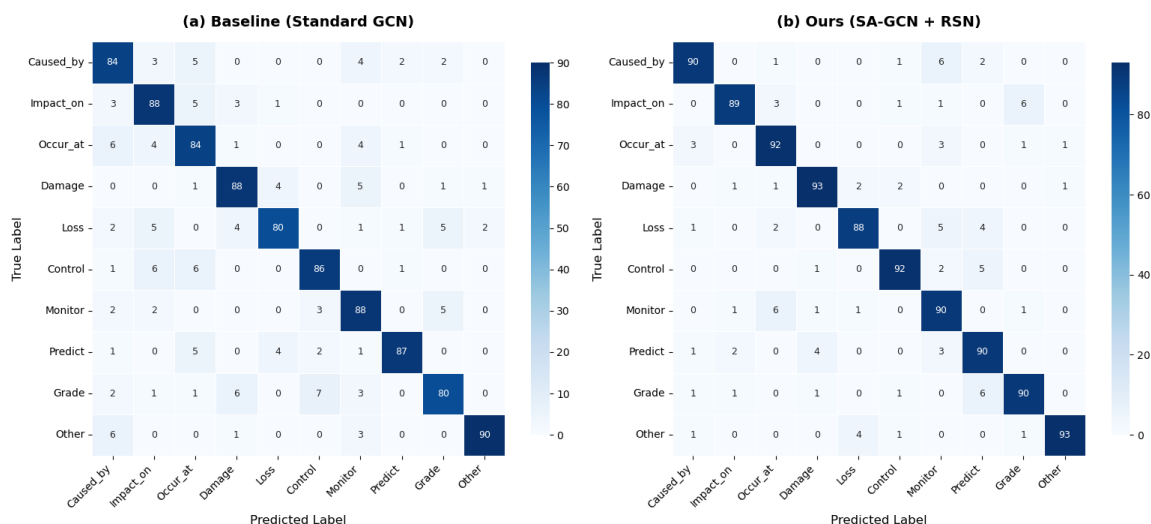


Figure 5. Confusion matrices on the Agri-Disaster test set.

4.5. Qualitative Analysis

To concretely demonstrate how the proposed modules (SA-GCN and RSN) improve performance in real-world scenarios, we selected a representative complex sample from the Agri-Disaster dataset for a case study.

Case Sample: "In late July, the continuous heavy rainfall caused severe waterlogging in the low-lying maize fields of Zhoukou, leading to a significant reduction in yield." (Entity 1: heavy rainfall; Entity 2: waterlogging; Relation: Cause)

- BERT (Baseline) Prediction: None (Incorrect).
- Analysis: The baseline model failed to capture the relationship because the two entities are separated by 6 words, and the sentence contains interfering adjectives like "continuous" and "low-lying". The attention mechanism of BERT was distracted by the surrounding noise.
- Ours (SA-GCN + RSN) Prediction: Cause (Correct).
- Analysis: SA-GCN contribution: The dependency syntactic tree explicitly connected "rainfall" and "waterlogging" through the path rainfall (subj) -> caused (root) -> waterlogging (obj). The SA-GCN module successfully shortened the semantic distance between the two entities.

RSN contribution: The adaptive denoising mechanism identified "continuous" and "low-lying" as low-information noise features. The soft thresholding operation suppressed these features, allowing the model to focus on the core "disaster-consequence" semantic link.

5. Conclusions and Future Work

This study proposes a relation extraction model for agricultural meteorological disaster texts that integrates BERT, a Self-Attention Graph Convolutional Network (SA-GCN), and a Residual Shrinkage Network (RSN). By combining dynamic graph pruning with adaptive feature denoising, the model forms a dual-denoising architecture that effectively alleviates long-distance dependency and noise interference issues commonly encountered in disaster-related texts. Experimental results on a self-constructed agricultural meteorological disaster dataset and two public benchmark datasets demonstrate that the proposed approach consistently outperforms existing mainstream methods across all evaluation metrics, validating its effectiveness and robustness.

From an application perspective, the proposed model provides a practical solution for automatically extracting high-precision structured relations from large-scale agricultural meteorological disaster reports. It can serve as a core component for downstream applications such as disaster knowledge base construction, intelligent monitoring of disaster evolution, risk assessment, and decision support in agricultural management. By enabling accurate and efficient

transformation of unstructured disaster texts into structured knowledge, the model supports timely analysis and informed decision-making in real-world disaster prevention and mitigation scenarios.

Future work will focus on extending the model to document-level relation extraction [59,60] to better capture cross-sentence and cross-paragraph dependencies in complex disaster reports. In addition, exploring few-shot and zero-shot learning paradigms will further enhance its applicability in data-scarce or emerging disaster scenarios. Finally, integrating the proposed model with large language models and dynamically constructed knowledge graphs is expected to provide a scalable and intelligent framework for continuous disaster knowledge acquisition and advanced intelligent services.

Acknowledgments: This research was supported by the National Key R&D Program of China(2022ZD0119500) and the Chinese Academy of Agricultural Sciences Innovation Project (No. CAAS-ASTIP-2026-AII). We also gratefully acknowledge the Data Hub, Chinese Agrosystem Long-Term Observation Network and the National Agriculture Science Data Center for providing data resources and technical support.

References

1. Qiu, M.; Xie, N.; Ji, L.; Huang, Y. Research on the Construction of Knowledge Graphs for Agricultural Meteorological Disasters: A Review. *Chin. J. Agrometeorol.* 2024, 45, 1216. <https://doi.org/10.3969/j.issn.1000-6362.2024.10.005>
2. Qiu, C.; Shen, Q.; Zhang, P.; Li, Y. Cn-MAKG: China Meteorology and Agriculture Knowledge Graph Construction Based on Semi-Structured Data. In *Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*; IEEE: Singapore, 2018; pp. 692 - 696. <https://doi.org/10.1109/ICIS.2018.8466389>
3. Wang, H.; Zhao, R. Knowledge Graph of Agricultural Engineering Technology Based on Large Language Model. *Displays* 2024, 85, 102820. <https://doi.org/10.1016/j.displa.2024.102820>
4. Yuan, L.; Wang, J.; Yu, L.C.; Wu, X.; Xue, L. Pruning Self-Attention With Local and Syntactic Dependencies for Aspect Sentiment Triplet Extraction. *IEEE Trans. Audio Speech Lang. Process.* 2025, 33, 2527 - 2538. <https://doi.org/10.1109/TASLPRO.2025.3577380>
5. Bach, N.; Badaskar, S. A Review of Relation Extraction; Technical Report; Carnegie Mellon University: Pittsburgh, PA, USA, 2007.
6. Bassignana, E.; Plank, B. What Do You Mean by Relation Extraction? A Survey on Datasets and Study on Scientific Relation Classification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*; Association for Computational Linguistics: Dublin, Ireland, 2022; pp. 67 - 76. <https://doi.org/10.18653/v1/2022.acl-srw.7>
7. Delaunay, J.; Tran, H.T.H.; González-Gallardo, C.E.; Ermakova, L. A Comprehensive Survey of Document-Level Relation Extraction (2016 - 2023). *arXiv* 2023, arXiv:2309.16396. <https://doi.org/10.48550/arXiv.2309.16396>
8. Zheng, H.; Wang, S.; Huang, L. A Comprehensive Survey on Document-Level Information Extraction. In *Proceedings of the Workshop on the Future of Event Detection (FuturED)*; Association for Computational Linguistics: Bangkok, Thailand, 2024; pp. 58 - 72. <https://doi.org/10.18653/v1/2024.futurED-1.6>
9. Smirnova, A.; Cudré-Mauroux, P. Relation Extraction Using Distant Supervision: A Survey. *ACM Comput. Surv.* 2018, 51, 1 - 35. <https://doi.org/10.1145/3241741>
10. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* 2021, 32, 4 - 24. <https://doi.org/10.1109/TNNLS.2020.2978386>
11. Zhao, X.; Deng, Y.; Yang, M.; Wang, Y.; Wang, J. A Comprehensive Survey on Relation Extraction: Recent Advances and New Frontiers. *arXiv* 2023, arXiv:2306.02051. <https://doi.org/10.48550/arXiv.2306.02051>
12. Zhu, H.; Lin, Y.; Liu, Z.; Fu, J.; Chua, T.S.; Sun, M. Graph Neural Networks with Generated Parameters for Relation Extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*; Association for Computational Linguistics: Florence, Italy, 2019; pp. 1331 - 1339. <https://doi.org/10.18653/v1/P19-1128>

13. Du, J.; Liu, G.; Gao, J.; Han, T. Graph Neural Network-Based Entity Extraction and Relationship Reasoning in Complex Knowledge Graphs. In *Proceedings of the 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*; IEEE: Chengdu, China, 2024; pp. 679 - 683. <https://doi.org/10.1109/ICICML60459.2024.10958048>
14. Liu, D.; Zhang, Y.; Li, Z. A Survey of Graph Neural Network Methods for Relation Extraction. In *Proceedings of the 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*; IEEE: Chongqing, China, 2022; Volume 10, pp. 2209 - 2223. <https://doi.org/10.1109/ITAIC54216.2022.9836511>
15. Wu, T.; You, X.; Xian, X.; Wang, S. Towards Deep Understanding of Graph Convolutional Networks for Relation Extraction. *Data Knowl. Eng.* **2024**, *149*, 102265. <https://doi.org/10.1016/j.datak.2023.102265>
16. Li, J.; Feng, S.; Chiu, B. Few-Shot Relation Extraction With Dual Graph Neural Network Interaction. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 14396 - 14408. <https://doi.org/10.1109/TNNLS.2023.3278938>
17. Zeng, L.; Tong, Y.; Guo, W.; Wang, Y. AgriCHN: A Comprehensive Cross-domain Resource for Chinese Agricultural Named Entity Recognition. *arXiv* 2025, arXiv:2506.17578. <https://doi.org/10.48550/arXiv.2506.17578>
18. Ye, P.; Zhang, C.; Chen, M.; Sun, Z.; Wan, H. Typhoon Disaster State Information Extraction for Chinese Texts. *Sci. Rep.* **2024**, *14*, 7925. <https://doi.org/10.1038/s41598-024-58537-x>
19. Qiu, M.; Jiang, L.; Xie, N.; Li, M.; Hu, Y. Research on Agricultural Meteorological Disaster Event Extraction Method Based on Character - Word Fusion. *Agronomy* **2025**, *15*, 2135. <https://doi.org/10.3390/agronomy15092135>
20. Wu, D.; Liu, X.; Zai, S.; Zhang, H. Analysis of Agrometeorological Hazard Based on Knowledge Graph. *Agriculture* **2024**, *14*, 1130. <https://doi.org/10.3390/agriculture14071130>
21. Efeoglu, S.; Karagoz, P. A Continual Relation Extraction Approach for Knowledge Graph Completeness. *arXiv* 2024, arXiv:2404.17593. <https://doi.org/10.48550/arXiv.2404.17593>
22. Fei, H.; Zhang, Y.; Ren, Y.; Dong, X. A Span-Graph Neural Model for Overlapping Entity Relation Extraction in Biomedical Texts. *Bioinformatics* **2021**, *37*, 1581 - 1589. <https://doi.org/10.1093/bioinformatics/btaa993>
23. Park, S.; Kim, H. Effective Sentence-Level Relation Extraction Model Using Entity-Centric Dependency Tree. *PeerJ Comput. Sci.* **2024**, *10*, e2311. <https://doi.org/10.7717/peerj-cs.2311>
24. Wang, L.; Wu, F.; Liu, X.; Li, J. Relationship Extraction Between Entities with Long Distance Dependencies and Noise Based on Semantic and Syntactic Features. *Sci. Rep.* **2025**, *15*, 15750. <https://doi.org/10.1038/s41598-025-00915-5>
25. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, 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; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; Volume 1, pp. 4171 - 4186. https://doi.org/10.18653/v1/N19-1423*
26. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* 2019, arXiv:1907.11692. <https://doi.org/10.48550/arXiv.1907.11692>
27. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*; Addis Ababa, Ethiopia, 2020.
28. He, P.; Liu, X.; Gao, J.; Chen, W. DeBERTa: Decoding-Enhanced BERT with Disentangled Attention. In *Proceedings of the 9th International Conference on Learning Representations (ICLR)*; Vienna, Austria, 2021.
29. Clark, K.; Luong, M.T.; Le, Q.V.; Manning, C.D. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*; Addis Ababa, Ethiopia, 2020.
30. Alt, C.; Hübner, M.; Hennig, L. Fine-Tuning Pre-Trained Transformer Language Models to Distantly Supervised Relation Extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational*

- Linguistics*; Association for Computational Linguistics: Florence, Italy, 2019; pp. 1388 - 1398. <https://doi.org/10.18653/v1/P19-1134>
31. Wang, X.; Gao, T.; Zhu, Z.; Zhang, Z.; Liu, Z.; Li, J.; Tang, J. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. *Trans. Assoc. Comput. Linguist.* 2021, 9, 176 - 194. https://doi.org/10.1162/tacl_a_00360
 32. Yamada, I.; Asai, A.; Shindo, H.; Takeda, H.; Matsumoto, Y. LUKE: Deep Contextualized Entity Representations with Entity-Aware Self-Attention. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Online, 2020; pp. 6442 - 6454. <https://doi.org/10.18653/v1/2020.emnlp-main.523>
 33. Joshi, M.; Chen, D.; Liu, Y.; Weld, D.S.; Zettlemoyer, L.; Levy, O. SpanBERT: Improving Pre-training by Representing and Predicting Spans. *Trans. Assoc. Comput. Linguist.* 2020, 8, 64 - 77. https://doi.org/10.1162/tacl_a_00300
 34. Lin, Y.; Xiao, H.; Liu, J. Knowledge-Enhanced Relation Extraction Dataset. *arXiv* 2022, arXiv:2210.11231. <https://doi.org/10.48550/arXiv.2210.11231>
 35. Beltagy, I.; Peters, M.E.; Cohan, A. Longformer: The Long-Document Transformer. *arXiv* 2020, arXiv:2004.05150. <https://doi.org/10.48550/arXiv.2004.05150>
 36. Swarup, A.; Bhandarkar, A.; Dizon-Paradis, O.P.; Pan, S.J.; Asadizanjani, N. Maximizing Relation Extraction Potential: A Data-Centric Study to Unveil Challenges and Opportunities. *IEEE Access* 2024, 12, 167655 - 167682. <https://doi.org/10.1109/ACCESS.2024.3494737>
 37. Sahu, S.K.; Thomas, D.; Chiu, B.; Sengupta, N. Relation Extraction with Self-determined Graph Convolutional Network. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*; ACM: New York, NY, USA, 2020; pp. 3075 - 3082. <https://doi.org/10.1145/3340531.3412072>
 38. Carbonell, M.; Riba, P.; Villegas, M.; Fornés, A.; Lladós, J. Named Entity Recognition and Relation Extraction with Graph Neural Networks in Semi Structured Documents. In *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)*; IEEE: Milan, Italy, 2021; pp. 9622 - 9627. <https://doi.org/10.1109/ICPR48806.2021.9412669>
 39. Blum, M.; Nolano, G.; Ell, B.; Cimiano, P. Investigating the Impact of Different Graph Representations for Relation Extraction with Graph Neural Networks. In *Proceedings of the Workshop on Deep Learning and Linked Data (DLnLD) @ LREC-COLING 2024*; ELRA and ICCL: Torino, Italy, 2024; pp. 1 - 13. <https://aclanthology.org/2024.dlnld-1.1>
 40. Yan, Z.; Yang, S.; Liu, W.; Tu, K. Joint Entity and Relation Extraction with Span Pruning and Hypergraph Neural Networks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Singapore, 2023; pp. 7512 - 7526. <https://doi.org/10.18653/v1/2023.emnlp-main.467>
 41. Mandya, A.; Bollegala, D.; Coenen, F. Graph Convolution over Multiple Dependency Sub-graphs for Relation Extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*; International Committee on Computational Linguistics: Barcelona, Spain (Online), 2020; pp. 6424 - 6435. <https://doi.org/10.18653/v1/2020.coling-main.565>
 42. Nan, G.; Guo, Z.; Sekulić, I.; Lu, W. Reasoning with Latent Structure Refinement for Document-Level Relation Extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*; Association for Computational Linguistics: Online, 2020; pp. 1546 - 1557. <https://doi.org/10.18653/v1/2020.acl-main.141>
 43. Zhang, Y.; Yang, Z.; Yang, Y.; Wang, J. Location-Enhanced Syntactic Knowledge for Biomedical Relation Extraction. *J. Biomed. Inform.* 2024, 156, 104676. <https://doi.org/10.1016/j.jbi.2024.104676>
 44. Wang, X.; Bai, X. Entity-Aware Self-Attention and Contextualized GCN for Enhanced Relation Extraction in Long Sentences. *arXiv* 2024, arXiv:2409.13755. <https://doi.org/10.48550/arXiv.2409.13755>
 45. Zhang, Y.; Qi, P.; Manning, C.D. Graph Convolution over Pruned Dependency Trees Improves Relation Extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium, 2018; pp. 2205 - 2215. <https://doi.org/10.18653/v1/D18-1244>

46. Guo, Z.; Zhang, Y.; Lu, W. Attention Guided Graph Convolutional Networks for Relation Extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*; Association for Computational Linguistics: Florence, Italy, 2019; pp. 241 - 251. <https://doi.org/10.18653/v1/P19-1024>
47. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371 - 3408.
48. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Pecht, M. Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4681 - 4690. <https://doi.org/10.1109/TII.2019.2943898>
49. Sun, Y.; Zhang, J.; Zhang, Y. New Deep Learning Network (Deep Residual Shrinkage Network) Is Applied for Lithology Identification to Search for the Reservoir of CO₂ Geological Storage. *Energy Fuels* **2024**, *38*, 2200 - 2211. <https://doi.org/10.1021/acs.energyfuels.3c04548>
50. Chen, Z.; Li, Z.; Wu, J.; Li, Y.; Li, W. Deep Residual Shrinkage Relation Network for Anomaly Detection of Rotating Machines. *J. Manuf. Syst.* **2022**, *65*, 579 - 590. <https://doi.org/10.1016/j.jmsy.2022.10.014>
51. Zheng, C.; Feng, J.; Fu, Z.; Cai, Y.; Li, Q.; Wang, T. Multimodal Relation Extraction with Efficient Graph Alignment. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21)*; ACM: New York, NY, USA, 2021; pp. 5298 - 5306. <https://doi.org/10.1145/3474085.3476968>
52. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 5998 - 6008.
53. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*; San Diego, CA, USA, 2015.
54. Luong, M.T.; Pham, H.; Manning, C.D. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Lisbon, Portugal, 2015; pp. 1412 - 1421. <https://doi.org/10.18653/v1/D15-1166>
55. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*; PMLR: Lille, France, 2015; Volume 37, pp. 2048 - 2057.
56. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735 - 1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
57. Cheng, J.; Dong, L.; Lapata, M. Long Short-Term Memory-Networks for Machine Reading. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Austin, TX, USA, 2016; pp. 551 - 561. <https://doi.org/10.18653/v1/D16-1053>
58. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder - Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 1724 - 1734. <https://doi.org/10.3115/v1/D14-1179>
59. Zhou, W.; Huang, K.; Ma, T.; Huang, H. Document-Level Relation Extraction with Adaptive Thresholding and Localized Context Pooling. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Palo Alto, CA, USA, 2021; Volume 35, pp. 14612 - 14620. <https://www.google.com/search?q=https://doi.org/10.1609/aaai.v35i16.17718>
60. Sun, Q.; Zhang, K.; Huang, K.; Xu, T.; Li, X.; Liu, Y. Document-Level Relation Extraction with Two-Stage Dynamic Graph Attention Networks. *Knowl.-Based Syst.* **2023**, *268*, 110428. <https://doi.org/10.1016/j.knsys.2023.110428>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.