Railway accident entity extraction method based on accident phase classification and mutual learning

Zhibo Cheng, Yanhua Wu and Zheqian Liu
China Academy of Railway Sciences Corporation Limited,
Institute of Computing Technologies, Beijing, China
Yong Shi
Safety Supervision and Management Bureau, China Railway Group Ltd,
Beijing, China, and
Ze Li
China Academy of Railway Sciences Corporation Limited,
Institute of Computing Technologies, Beijing, China

Received 25 August 2025
Revised 28 September 2025
Accepted 30 September 2025

Abstract

Purpose – This study aims to enhance the accuracy of key entity extraction from railway accident report texts and address challenges such as complex domain-specific semantics, data sparsity and strong inter-sentence semantic dependencies. A robust entity extraction method tailored for accident texts is proposed.

Design/methodology/approach – This method is implemented through a dual-branch multi-task mutual learning model named R-MLP, which jointly performs entity recognition and accident phase classification. The model leverages a shared BERT encoder to extract contextual features and incorporates a sentence span indexing module to align feature granularity. A cross-task mutual learning mechanism is also introduced to strengthen semantic representation.

Findings – R-MLP effectively mitigates the impact of semantic complexity and data sparsity in domain entities and enhances the model’s ability to capture inter-sentence semantic dependencies. Experimental results show that R-MLP achieves a maximum F1-score of 0.736 in extracting six types of key railway accident entities, significantly outperforming baseline models such as RoBERTa and MacBERT.

Originality/value – This demonstrates the proposed method’s superior generalization and accuracy in domain-specific entity extraction tasks, confirming its effectiveness and practical value.

Keywords Accident report texts, Entity extraction, Accident phase classification, Multi-task model, Mutual learning mechanism

Paper type Research article

1. Introduction

With the advancement of information technology in the field of railway safety regulation, the Railway Safety Supervision and Management Information System has been implemented across the entire network. It integrates a large volume of multimodal railway accident data, including structured data such as accident severity, line name, and line grade, as well as unstructured texts like accident overview, casualties, direct economic losses, and corrective measures. These multi-source heterogeneous safety data provide a comprehensive description of accident details, offering effective support for accident report text analysis and the

---

Funding: This research was funded by the Technology Research and Development Plan Program of China State Railway Group Co., Ltd. (No. Q2024T001), and the Foundation of China Academy of Railway Sciences Co., Ltd. (No: 2024YJ259).

Railway Sciences
Vol. 4 No. 6, 2025
pp. 815-832
Emerald Publishing Limited
e-ISSN: 2755-0915
p-ISSN: 2755-0907
DOI 10.1108/RS-08-2025-0030

quantitative analysis of accident losses (Shi *et al.*, 2024). Existing railway accident texts are highly specialized in content organization, writing style, and terminology. To enhance analysis and information extraction capabilities for railway accident texts, it is necessary to develop a knowledge system in the railway safety supervision field. Entity extraction, as a key foundational task, faces challenges such as unclear entity boundaries, insufficient corpus acquisition, and a lack of high-quality samples. Additionally, influenced by the complexity of the railway system and the sporadic nature of accidents, accident-related entities exhibit significant sparsity in the corpus (Chen, 2023), demanding higher capabilities in entity feature analysis and semantic modeling.

In recent years, both domestically and internationally scholars have conducted extensive research on text entity extraction methods (Xiao & Chen, 2024). Traditional dictionary and rule-based pattern matching approaches, which rely on clear rule templates, struggle with comprehensive entity coverage. Statistical machine learning methods depend on high-quality, large-scale annotated corpora and expert-designed feature templates, limiting model generalization and transferability in railway-specific applications. Deep learning models, which do not require expert-designed entity features, have gradually become dominant in entity extraction, enhancing model performance. Significant progress has been made in general domains as well as in specific fields such as law (Lu & Li, 2024), medicine (Yang *et al.*, 2016), military (Yin, Zhao, Zhao, Yao, & Huang, 2020), and finance (Xu, Zhu, Luo, & Dong, 2021). While entities like time, location, railway bureau, and equipment names in railway accident texts have been recognized (Du, Jin, Dai, Xue, & Wu, 2023), there remains a lack of analysis and extraction of complex causative factors, initial events, intermediate events, or final events (Shi *et al.*, 2024). With the widespread application of large language models (e.g. GPT and ERNIE) in natural language processing, their robust semantic modeling capabilities offer new possibilities for entity extraction tasks. However, pre-trained large-scale models are typically trained on general-domain corpora and thus lack the capability to comprehend domain-specific terminology (Liang, Zhang, & Yan, 2024; Li *et al.*, 2025), resulting in insufficient generalization and weak domain adaptability in railway-specific safety supervision text extraction. Given that existing general-purpose entity extraction methods fall short of meeting the requirements of safety text data analysis and mining, there is an urgent need for further in-depth exploration in areas such as extraction model architecture design, semantic modeling capability, and domain adaptability.

Compared with general-domain texts, railway accident texts exhibit a more pronounced stage-oriented narrative structure, typically recorded in accordance with the accident's evolution, including the occurrence process, specific causes, response measures, and rectification actions. Incorporating accident phase information into the entity extraction process can thus provide richer contextual and semantic relationship cues, thereby improving the accuracy of entity type recognition and extraction boundary determination. Based on these characteristics, this study proposes a railway accident text entity extraction model that integrates accident phase classification and mutual learning. The model jointly performs Entity Recognition (ER) and Accident Phase Classification (APC) using a shared Bidirectional Encoder Representations from Transformers (BERT) encoder, and introduces a Mutual Learning Mechanism (MLM) between the two task branches to enable bidirectional information exchange. This design fully exploits the potential associations of entities across different accident phases and enhances the accuracy and robustness of domain-specific entity extraction in scenarios with sparse samples.

2. Railway accident phase division and text entity definition

The railway sector has accumulated a large body of textual materials—such as determination documents, investigation reports, and handling reports—for various types of accidents, which systematically record the dynamic evolution of accidents over time (Shi, 2023). According to the literature (Shi *et al.*, 2024), railway accidents can be divided, based on their development process, into different stages of an event chain: the initial event, the intermediate event, and the final event. The initial event refers to an incidental functional failure within the railway operation system; the intermediate event denotes a subsequent occurrence in the event chain that bridges the initial and final events; and the final event corresponds to the scenario that directly results in loss or damage, such as collisions, derailments, fires, or explosions.

Given that accident texts also contain information on the causes of accidents and post-incident handling measures, this study introduces two additional stages: the potential stage and the recovery stage. Accordingly, based on the accident overview texts, railway accidents are summarized into the following stages: potential causation stage, initial event stage, intermediate event stage, final event stage, and recovery/response stage. The initial and intermediate stages are described using patterns such as “human–operation–object,” “object–action–phenomenon,” and “location–existence–phenomenon.” The descriptions and examples of different accident stages are presented in Table 1.

Across different stages, accident texts provide comprehensive documentation of the overall incident, encompassing critical information such as the location of occurrence, causative factors, operational scope, responsible parties, response measures, and resulting impacts. Owing to the stochastic nature of accidents, entities associated with accident causation frequently exhibit pronounced sparsity. Based on the aforementioned informational dimensions, the key entity types within accident texts, along with representative examples, are defined as presented in Table 2.

Table 1. Accident phases and examples in railway accident texts

Accident phases	Description	Example
Potential Causation Stage	Accident has not yet occurred, but potential causative factors exist	Equipment aging, environmental changes, etc.
Initial Event Stage	Accident is triggered by an initial event	Human operational error, foreign object intrusion, etc.
Intermediate Event Stage	Relevant personnel or systems intervene to control accident progression	Emergency stop, train interception, etc.
Final Event Stage	Accident causes consequences	Collision, derailment, fire, explosion, etc.
Recovery/Response Stage	Post-incident site restoration, with impacts gradually eliminated	Rescue operations, manual handling

Source(s): Authors’ own work

Table 2. Labels, descriptions, and examples of key entities in railway accident texts

Entity type	Entity label	Example
Basic Information	Information	Time, location, personnel, organization, etc.
Causative Factors	Causality	Continuous heavy rainfall, strong winds, earthquakes, etc.
Initial Event	Initial Events	Catenary sagging into clearance limits, signalman misrouting turnout, etc.
Operation Type	Type	Train operation, shunting operation, etc.
Intermediate Event	Intermediate Events	Failure to stop, failure to intercept, etc.
Control Measures	Protective Measure	Driver-initiated stop, line closure, ground interception, etc.
Accident Consequences	Consequence	12-min delay to the train, disruption of three passenger trains, etc.
Disposal Measures	Treatment Measures	Door securing, rescue and rerailing, power supply restoration, etc.

Source(s): Authors’ own work

In accident texts, there is a pronounced correlation between accident phases and entity types. In the potential causation stage, entities include the location of occurrence and the causative factors leading to the initial event. In the initial event stage, initial event entities are described in relation to attributes such as human operational errors, equipment failures, or environmental changes. In the intermediate event stage, one or more failed control measures, such as personnel response or equipment alarms, constitute intermediate event entities. In the final event stage, the focus shifts to the impacts triggered by accident consequences, describing the set of entities corresponding to different outcomes. In the recovery/response stage, the set of entities related to disposal measures is described.

Incorporating accident phases into entity extraction enables a clear delineation of the stages within the event chain to which entities belong, thereby reinforcing the intrinsic logical relationships of entities within the accident evolution process. It also provides constraint information for entity extraction. For example, crane equipment damage may be classified as a causative factor in the potential causation stage, but as an intermediate event in the intermediate stage. The associations between accident phases and key entities in railway accident texts are presented in Table 3.

To address the challenges of semantic complexity, entity sample sparsity, and strong inter-sentence dependencies in railway accident texts, this study proposes a multi-task mutual learning–based entity extraction model that jointly performs entity recognition and phase classification, leveraging phase information to enhance entity boundary detection and semantic discrimination. The framework adopts dual-branch architecture with a shared BERT encoder, employs sentence span indexing to achieve feature granularity alignment, and incorporates a mutual learning mechanism to promote semantic fusion between the two tasks. This design enables the extraction and structured output of key entities in the text, thereby improving the intelligent analysis and mining capabilities for railway accident texts. The overall framework of the proposed railway accident text entity extraction approach is shown in Figure 1.

3. Railway accident text entity extraction model based on accident phase classification and mutual learning

Railway accident texts have domain-specific characteristics, with entities typically demonstrating complex structures, fine-grained semantics, and sparse distributions. To improve the accuracy of key entity recognition in railway safety accident texts, this study proposes a text entity extraction model called R-MLP (Railway Entity Extraction with Mutual Learning and Accident Phase Classification), which models entity recognition and phase classification as independent tasks, while enhancing extraction performance through inter-task information exchange. A pre-trained language model, BERT, is employed as a shared encoder to capture the initial semantic features of the text, which are then fed into the entity recognition branch and the accident phase classification branch, respectively. In the entity recognition branch, a Conditional Random Field (CRF) (Shi, 2023; Pooja &

Table 3. Association between accident entities and accident phases in railway accident texts

Accident stage	Common entities
Potential Causation Stage	Basic Information, Causative Factors
Initial Event Stage	Initial Event, Operation Type
Intermediate Event Stage	Basic Information, Intermediate Event, Control Measures
Final Event Stage	Accident Consequences
Recovery/Response Stage	Basic Information, Disposal Measures

Source(s): Authors’ own work

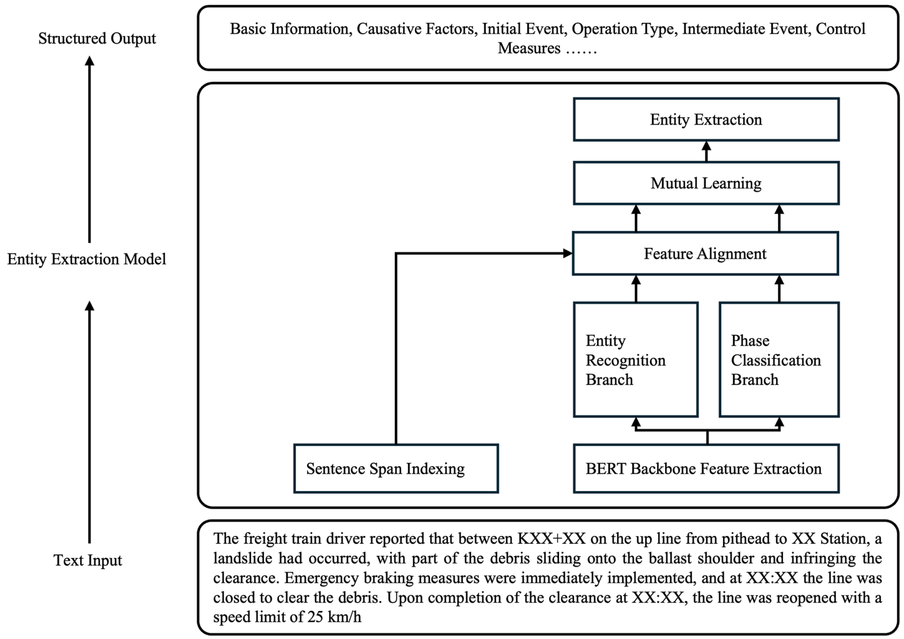

Figure 1. Framework for entity extraction from railway accident texts. **Source(s):** Authors' own work

Jagadeesh, 2024) is used to model label dependencies, thereby improving the consistency and accuracy of boundary recognition. In the phase classification branch, semantic information is aggregated to achieve effective accident phase identification. Considering that the same entity may carry different semantics across phases, a feature alignment and mutual learning mechanism (Wang, Li, & Ge, 2023) is introduced between the two tasks to facilitate information sharing and interaction. This mechanism provides contextual constraints for entity recognition, enhances the model's semantic parsing capability, and improves robustness and generalization under conditions of sample sparsity. Ultimately, the model achieves accurate extraction of key accident entities (Entity Extraction, EE). The framework of the proposed R-MLP model is illustrated in Figure 2.

3.1 Sentence span indexing module

In dual-branch multi-task learning models, the entity recognition (ER) task is typically annotated at the character (token) level, whereas the accident phase classification (APC) task is performed at the sentence level. This difference in granularity prevents feature alignment between the two branches during the mutual learning process. To address this issue, this study designs a sentence span indexing module, as shown in Equation (1). The primary function of this module is to segment the input text T_n according to natural language clause boundaries indicated by punctuation marks (e.g., commas, periods), and to extract the start and end indices of the original tokens for each clause. These index mappings serve as the key basis for achieving feature alignment, and the process can be expressed as follows:

$$SPAN_m = SpanSplit(T_n) \tag{1}$$

In the equation, $T_n = \{c_1, c_2, c_3 \ldots, c_n\}$ is a text sequence of length n, and $SPAN_m = \{span_1, span_2, span_3 \ldots, span_m\}$ represents the set of boundary indices for sentence

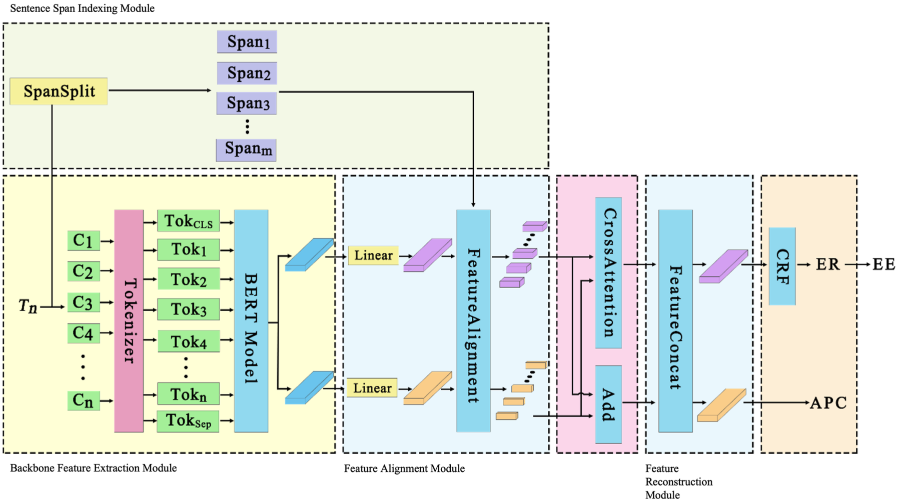

Figure 2. R-MLP model framework. **Source(s):** Authors' own work

segmentation. Each $span_{i=}[s_i, e_i]$ specifies the start token index s_i and the end token index e_i of the *i*-th sentence. The sentence span indexing results are illustrated in Table 4, where each pair of numbers in parentheses indicates the start and end character positions of a clause in the original text. These boundary indices provide essential alignment information between token-level and sentence-level features in the subsequent dual-branch mutual learning process.

3.2 Backbone feature extraction module

Although the objectives of the entity recognition task and the phase classification task differ, both operate on the same text sequence and thus exhibit a high degree of overlap in underlying linguistic features. Sharing backbone features not only enables the information-interaction model to improve its understanding of one task while optimizing the other, but also enhances overall modeling efficiency and effectiveness. As shown in Equation (2), the backbone feature extraction module primarily encodes the input text T_n to obtain the initial semantic features $F_{blackbone}$. First, the Tokenizer module segments and encodes the text T_n, automatically appending special tokens such as [CLS] and [SEP] to meet the structural input requirements of the BERT model. Then, the BERT model employs a multi-layer Transformer encoder to perform representation learning on the input sequence, effectively modeling bidirectional dependencies and semantic relationships within the context (He, Zhao, & Tang, 2025). This process generates high-dimensional semantic feature vectors for each character, enriched with contextual information, and can be expressed as:

Table 4. Results of clause index extraction

Text	Sentence span Indexing ($SPAN_m$)
Due to a landslide halting at K122 + 863, the locomotive's left-side pilot (cowcatcher) struck the debris during the stop. At 03:31, the driver requested rescue assistance	(0,15) (17,32) (34,44)

Source(s): Authors' own work

$$F_{balckbone} = Bert(Tokenizer(T_n)) \tag{2}$$

As shown in Figure 3, the initial semantic features $F_{blackbone}$ are simultaneously fed into both the ER branch and the APC branch. By employing a shared backbone feature extraction module, the framework not only improves parameter utilization efficiency but also provides a unified and high-quality representation space for subsequent tasks, thereby facilitating effective multi-task collaborative learning.

3.3 Feature alignment module

The feature alignment module, as a core component of the dual-branch mutual learning framework, is designed to ensure semantic consistency between the token-level modeling of the ER task and the sentence-level modeling of the APC task. The module first segments the input sequence into sentence-level units based on sentence boundary indices ($SPAN_m$). For the ER task, this segmentation provides explicit entity boundary constraints, preventing incorrect recognition of entities that span multiple sentences. For the APC task, the module generates stable sentence representations by aggregating (via mean pooling) the token features within each sentence, treating each clause as an independent classification unit to reduce semantic ambiguity and phase overlap. More importantly, through a structured index mapping mechanism, the module strengthens the correspondence between token-level and sentence-level features despite the difference in task granularity, thereby achieving explicit cross-task representation space alignment. The specific process is as follows:

The current dual branches share the same feature dimensionality, the features of the branches can be defined as:

$$H = [h_1, h_2, h_3 ... h_z] \in \mathbb{R}^{z \times d} \tag{3}$$

where z denotes the total number of tokens and the feature dimension. Based on the sentence boundary indices $SPAN_m$, the token feature set for the same sentence can be extracted as:

$$s_i = H[s_i : e_i + 1] \tag{4}$$

where $s_i \in \mathbb{R}^{l_i \times d}$ denotes the token-level feature block after segmentation, and $l_i = e_i + 1 - s_i$ represents the length of the i-th sentence. Ultimately, the token-level feature block set S^{er} for the ER branch and the token-level feature block set S^{apc} for the APC branch are obtained, which can be expressed as:

$$S^{er} = \left\{ s_1^{er}, s_2^{er}, s_3^{er} ... s_m^{er} \right\} \tag{5}$$

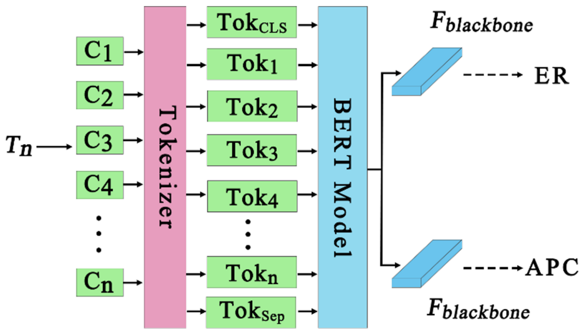

Figure 3. Backbone feature extraction module. **Source(s):** Authors' own work

$$S^{apc} = \left\{ s_1^{apc}, s_2^{apc}, s_3^{apc} \ldots s_m^{apc} \right\} \tag{6}$$

For the phase classification branch, each s_i^{apc} is further aggregated using mean pooling to obtain the sentence-level vector g_i:

$$g_i^{apc} = MeanPooling\left(s_i^{apc}\right) \tag{7}$$

where $g_i \in \mathbb{R}^{1\times d}$ denotes the sentence-level feature vector, and the final sentence-level feature vector set is $G^{apc} \in \mathbb{R}^{m\times d}$:

$$G^{apc} = \left\{ g_1^{apc}, g_2^{apc}, g_3^{apc} \ldots g_m^{apc} \right\} \tag{8}$$

As shown in Figure 4, the feature alignment module uses $\ominus$ to denote the feature segmentation symbol. Through a structured segmentation approach, the ER branch can satisfy the strict boundary requirements of entity recognition, while the APC branch can extract sentence-level feature representations to meet the overall semantic understanding needs of the phase classification task. At the same time, the boundary index–based method preserves the alignment relationship between the token-level features in the ER branch and the sentence-level features in the APC branch.

3.4 Asymmetric mutual learning module

The entity recognition task is a fine-grained sequence labeling task that classifies each token, whereas the phase classification task is a coarse-grained sentence classification task that assigns labels to entire sentences. Owing to the differences in task granularity and information requirements, this study adopts an asymmetric mutual learning module design.

In the ER branch, to enhance the model's ability to learn contextual information, this study designs a cross-attention–based feature fusion module (citation 14), where the feature g_i^{apc} from the APC branch serves as the Query, and the feature s_i^{er} from the entity recognition branch serves as both the Key and the Value. Through the attention mechanism, the feature s_i^{er} is adaptively weighted and adjusted, enabling each token in s_i^{er} to learn the overall semantic context of its corresponding sentence, thereby improving entity recognition accuracy in complex contexts. Since $g_i^{apc} \in \mathbb{R}^{1\times d}$ and $s_i^{er} \in \mathbb{R}^{l_i\times d}$, the feature g_i^{apc} is first replicated along the token dimension l_i of s_i^{er}, which can be expressed as:

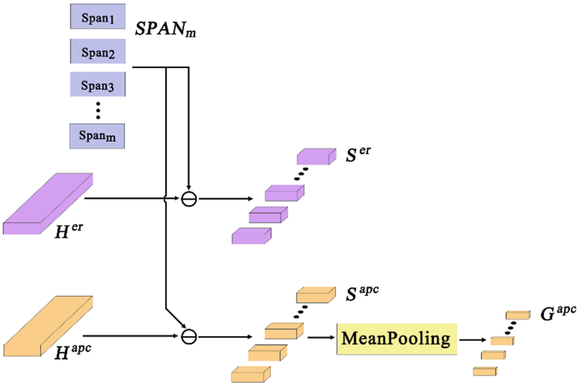

Figure 4. Feature alignment module. **Source(s):** Authors' own work

$$q_i = repeat(g_i^{apc}, l_i) \tag{9}$$

where the resulting query matrix $q_i \in \mathbb{R}^{l_i \times d}$ has the same dimensionality as s_i^{er}, and together with s_i^{er} serving as the key k_i and value v_i, is fed into the cross-attention module. The cross-attention process can be expressed as:

$$Attention(q_i, k_i, v_i) = softmax\left(\frac{q_i k_i^{\mathrm{T}}}{\sqrt{d}}\right) v_i \tag{10}$$

To mitigate gradient vanishing and accelerate model convergence, the attention-weighted features are added to the original s_i features via a residual connection, forming the fused features s_i^{fused}. This process can be expressed as:

$$s_i^{fused} = s_i^{er} + Attention(q_i, k_i, v_i) \tag{11}$$

As shown in Figure 5, the cross-attention–based feature fusion module uses $\otimes$ to denote the repeat operation, in which the sentence-level feature g_i^{apc} is replicated along the token dimension l_i of s_i^{er}. This fusion strategy preserves the fine-grained token-level semantic representation in the ER branch while incorporating sentence-level information, thereby enhancing the entity recognition branch's ability to capture the semantics of the containing sentence. In addition, the residual connection design facilitates optimized information flow and alleviates the vanishing gradient problem in deep networks.

In the APC branch, to incorporate fine-grained local semantic information from the ER branch, this study employs a lightweight compression-fusion module. The token-level features s_i^{er} from the ER branch are aggregated via mean pooling to obtain the compressed feature s_i', which can be expressed as:

$$s_i' = MeanPooling\left(s_i^{er}\right) \tag{12}$$

where $s_i' \in \mathbb{R}^{1 \times d}$ has the same dimensionality as g_i^{apc}, and is directly added to g_i^{apc} to achieve semantic fusion:

$$g_i^{fused} = s_i^{cls} + g_i^{apc} \tag{13}$$

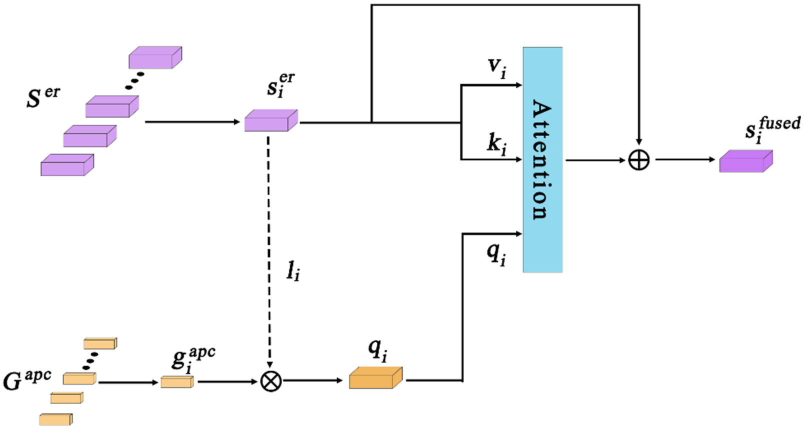

Figure 5. Feature fusion module based on cross-attention. **Source(s):** Authors' own work

As shown in Figure 6, compared with the complex cross-attention mechanism adopted in the ER branch, the APC branch employs a more lightweight strategy for feature fusion. By simply applying feature compression followed by an addition operation, it can incorporate supplementary semantic information from the ER branch. This asymmetric fusion design ensures effective information interaction while significantly reducing computational overhead, thereby avoiding structural redundancy in the model and achieving a balance between efficiency and performance.

This study proposes an asymmetric mutual learning strategy that fully accounts for the fundamental differences between the entity recognition task and the phase classification task in terms of granularity level, semantic focus, and modeling objectives. In the ER branch, a cross-attention–based learning mechanism is introduced, whereby phase-related contextual information is provided to help the model more accurately locate and distinguish key entities. In the APC branch, a compression-fusion learning mechanism is employed to effectively incorporate local entity information, thereby enhancing the precision of sentence-level semantic representations. This asymmetric learning strategy not only enables effective interaction between features of different granularities but also balances computational efficiency with structural simplicity.

3.5 Feature reconstruction module

During the mutual learning process, the model performs segmentation of the original continuous features and local feature fusion, resulting in the features being divided into multiple independent segments. The feature reconstruction module accurately concatenates these segmented features in their original order to restore a complete and coherent feature representation, thereby providing the subsequent entity recognition and phase classification tasks with complete input representations. In the ER branch, the segmented features are reassembled to restore the complete token feature sequence $\widehat{S}$, and the process is expressed as:

$$\widehat{S} = Concat\left(s_1^{fused}, s_2^{fused}, s_3^{fused} \ldots s_m^{fused}\right) \in \mathbb{R}^{n\times d} \tag{14}$$

Similarly, in the APC branch, the segmented features are reassembled to restore the complete sentence feature sequence $\widehat{G}$.

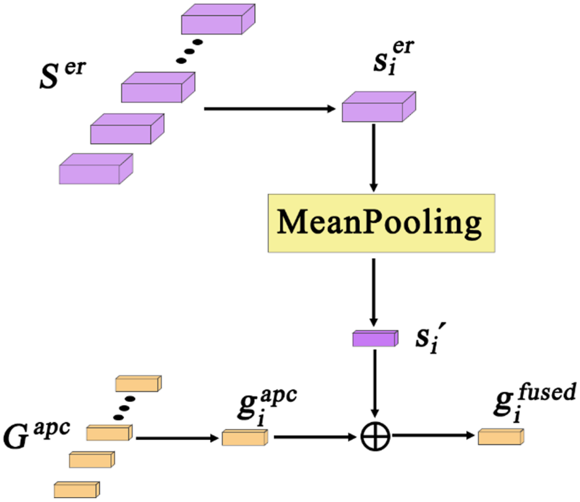

Figure 6. Compressed fusion model. **Source(s):** Authors' own work

3.6 Dual-branch prediction and joint loss function design

In the entity recognition task, the model performs fine-grained label prediction at the token level, i.e., it identifies and classifies each position in the reconstructed feature sequence $\widehat{S}$. A fully connected layer maps the feature dimension d of the fused representation to the number of entity categories C_{er}, yielding the label distribution $logits_{er}$ for each token. Given the strong structural constraints of the entity recognition task, this study incorporates a Conditional Random Field (CRF) as the label prediction layer in the decoding stage to enhance the model's ability to discriminate entity boundaries and categories (Zhang *et al.*, 2024), producing the optimal label sequence y_{er} for the entire sentence. This process can be expressed as:

$$y_{er} = CRF.decode(logits_{er}) \tag{15}$$

In addition, during training, the loss function of the CRF models the entire label sequence by maximizing the log-likelihood of the ground truth label sequence. This process can be expressed as:

$$loss_{er} = -\log p(label_{er}|logits_{er}) \tag{16}$$

where $label_{er}$ denotes the ground truth entity labels, and $loss_{er}$ represents the prediction loss value for the entity recognition task.

In the phase classification task, label prediction is performed on the sentence-level feature vectors $\widehat{G}$. Similarly, a fully connected layer maps the fused feature dimension d to the number of phase categories C_{apc}, yielding the class label distribution $logits_{apc}$ for each sentence. Since the phase classification task does not involve label structures with contextual dependencies, the cross-entropy loss function (Shen, Wang, & Zhao, 2025) is adopted for loss computation. This process can be expressed as:

$$loss_{apc} = CrossEntropy\left(label_{apc}, logits_{apc}\right) \tag{17}$$

where $label_{er}$ denotes the ground truth phase labels, and $loss_{apc}$ represents the prediction loss value for the phase classification task.

To fully leverage the complementary advantages of the dual-branch structure in multi-task learning, this study adopts a joint optimization strategy to collaboratively model and train the entity recognition and phase classification tasks simultaneously. By integrating the task losses from both branches, the model facilitates the sharing of learned knowledge and semantic enhancement, thereby effectively improving overall performance. This process can be expressed as:

$$L_{total} = \lambda loss_{er} + (1 - \lambda) loss_{apc} \tag{18}$$

where $\lambda \in [0, 1]$ is an adjustable weighting coefficient used to balance the relative contributions of the two tasks during training. Given that the phase classification task serves as an auxiliary task in this study, λ is set to 0.7.

4. Experimental validation and results analysis

4.1 Dataset description

This study collects a total of 3,824 accident summary text records from the past four years (2020–2023) from the Railway Safety Supervision and Management Information System. The dataset primarily contains information such as date and time, accident code, railway line, responsible bureau, and detailed summary descriptions. The raw dataset was cleaned by removing blank entries, eliminating invalid symbols and stop words, performing paragraph segmentation, basic error correction, and encoding conversion. Certain sensitive data were replaced with symbols. This process resulted in the formation of a preprocessed dataset for pre-training. Examples of accident summary descriptions are presented in Table 5.

Table 5. Example of pre-training dataset

Number	Detailed overview
1	On XX, XX, 20XX, at XX:XX, Train XX of Line 6 operated by XX Bureau was traveling on the up line between Yanjin North and Linjiangxi when it collided with fallen rocks at kilometer marker K211 + 985 and came to a stop. Upon inspection, it was found that the locomotive's obstacle removal device and the right-side sand spreader were damaged. After maintenance work was completed, the train resumed operation at XX:XX, and the section was reopened at XX:XX. Cause: Due to various factors such as rainfall and earthquakes, a landslide occurred on the local road embankment, causing rocks to pile up on the road surface. Some rocks crossed the road's crash barrier and rolled down the left-hand embankment of the railway onto the tracks. Upon discovering the anomaly, the on-site inspection personnel alerted the driver. However, the train's braking distance was insufficient, and a collision occurred during the braking process
2	On XX, XX, 20XX, at XX:XX, the XX freight train on the XX line operated by the XX Bureau was traveling between Mengdu and Sanyuanba when it collided with fallen rocks and came to a stop at K157 + 303, affecting four freight trains. Cause: Due to recent continuous rainfall, the soil on the natural slope at the top of the first-class embankment on the left side of the XX Line at K157 + 400 became saturated with water. Localized areas were eroded by rainwater, causing the soil around and beneath the boulders in the overlying layer to soften, leading to a localized landslide. Due to the close proximity, the train collided with the fallen rocks during braking

Source(s): Authors' own work

Since the field of railway safety supervision covers multiple professional subfields such as engineering, electrical engineering, power supply, rolling stock, and locomotives, it has strong industry characteristics. Currently, there are no publicly available entity annotation datasets covering this scenario. Therefore, this paper constructs a specialized dataset based on railway accident overview texts and annotates the texts using the BIO tagging method. An example of the stage annotation is shown in Table 6.

4.2 Evaluation metrics description

To evaluate the performance of the proposed model, three metrics: precision P, recall R and F_1 score, are employed. These can be expressed as:

$$\begin{aligned} P &= \frac{T_p}{T_p + F_p} \\ R &= \frac{T_p}{T_p + F_n} \\ F1 &= \frac{2\times P\times R}{P + R} \end{aligned} \tag{19}$$

where T_p (True Positive)denotes the number of samples correctly predicted as positive, F_p (False Positive)denotes the number of negative samples incorrectly predicted as positive, and F_N (False Negative)denotes the number of positive samples incorrectly predicted as negative.

4.3 Ablation study

To comprehensively evaluate the specific contributions of each key module to the model's entity extraction performance, this study builds a multi-task fusion model by incorporating the accident phase classification task, mutual learning, and feature alignment mechanisms on top of entity recognition. To further analyze the role of each component in improving overall performance, a series of ablation experiments are designed. By progressively adding phase classification, the mutual learning mechanism, and the feature alignment module, the impact

Table 6. Examples of entity annotation and phase annotation

Original text	Due to rockfall, the train collided with the stationary vehicle												
Entity Annotation	O	O	B-Initial Event	I- Initial Event	I- Initial Event	I- Initial Event	O	O	O	O	O	B-Intermediate Event	0- Intermediate Event
Stage Annotation	Initial Event Stage							Intermediate Event Stage					

Source(s): Authors’ own work

of each structural enhancement on entity extraction effectiveness is systematically examined, thereby verifying the effectiveness of the model design and the synergistic gains of its components.

The structural configurations of the ablation experiments are shown in Table 7. Model 1 is the baseline model (BERT + ER), which only includes the entity recognition task branch and does not incorporate any auxiliary information. Model 2 introduces the accident phase classification (APC) task on top of the baseline model, forming a dual-branch structure. The ER and APC tasks share the BERT encoder to achieve the sharing of basic semantic features. Furthermore, Model 3 incorporates a mutual learning mechanism (MLM) into the dual-task architecture to promote high-level semantic information interaction and collaborative optimization between the two tasks. Finally, Model 4 (R-MLP) introduces a feature alignment module (FAM) on the aforementioned structure, achieving unified alignment between different tasks at the semantic granularity level through sentence-level segmentation of the dual branches and construction of segment mappings, thereby enhancing feature fusion and context modeling capabilities.

The results of the model ablation experiments are shown in Table 8, illustrating the performance changes after gradually introducing structural modules based on the Baseline model. The initial model BERT + CRF achieved an F1 score of 0.564, relying solely on basic sequence modeling capabilities, resulting in limited entity extraction performance. After introducing the accident stage classification task to construct a dual-branch structure, the F1 score of Model 2 improved to 0.645, indicating that auxiliary tasks can enhance entity recognition, but there is still a lack of deep semantic synergy between tasks. Model 3 further incorporates a mutual learning mechanism to achieve high-dimensional semantic interaction, but due to unaligned features, the F1 score drops back to 0.586, reflecting that semantic fusion imbalance may introduce noise. Finally, R-MLP introduced a feature alignment module on this basis to achieve semantic granularity unification and task collaboration, significantly improving the F1 score to 0.736, validating the key role of the alignment mechanism in multi-task mutual learning. Additionally, with structural improvements, the model's accuracy and recall rates both improved, indicating a comprehensive enhancement in overall recognition performance. By combining a dual-branch structure with mutual learning mechanisms and

Table 7. Structural differences of models in ablation study

Model	Description	Stage classification task	Mutual learning mechanism	Feature alignment
Model 1	BERT + CRF(Baseline)	×	×	×
Model 2	Baseline + APC	✓	×	×
Model 3	Baseline + APC + MLM	✓	✓	×
R-MLP	Baseline + APC + MLM + FAM	✓	✓	✓

Source(s): Authors' own work

Table 8. Experimental results of model ablation studies

Model	P	R	F1
Model 1	0.613	0.522	0.564
Model 2	0.694	0.603	0.645
Model 3	0.593	0.58	0.586
R-MLP	0.772	0.703	0.736

Source(s): Authors' own work

feature alignment design, the model effectively leverages complementary information across multiple tasks, significantly enhancing entity extraction model performance and demonstrating the advantages of multi-task mutual learning in complex information extraction tasks.

4.4 Comparative experiment

To verify the effectiveness of the proposed model, R-MLP is compared with several BERT-based optimized models developed in recent years. The comparative experimental results of different models are shown in Table 9. The RoBERTa + CRF model achieves an F1-score of only 0.151, significantly lower than expected, because RoBERTa is well suited for intra-sentence modeling (Mei, Wu, Huang, & Ling, 2023) but struggles to handle inter-sentence semantic dependencies present in railway accident texts. The BERT + CRF and BERT + BiLSTM + CRF models show comparable performance, with F1-scores of 0.564 and 0.596, respectively; although BiLSTM improves contextual modeling capability, its effectiveness in capturing long-range dependencies is limited. The MacBERT + CRF model achieves an F1-score of 0.605, benefiting from its optimized masking strategy, which better approximates real language usage. The proposed R-MLP model achieves the highest F1-score of 0.736, substantially outperforming the other models. This improvement is attributed to its multi-task learning strategy, which jointly performs entity recognition and accident phase classification, and to the incorporation of a cross-task mutual learning mechanism, enabling effective utilization of correlated information from both tasks to enhance feature representation and improve the model's understanding of low-resource semantic units.

To further assess the recognition capability of various models across different entity categories, this experiment conducts a detailed comparative analysis on the identification performance of six types of railway accident–related entities. The F1-scores of each model for different entity categories are compared in Figure 7. The RoBERTa + CRF model performs worst overall, lacking adequate contextual modeling capability. BERT + CRF shows notable improvements for labels such as "Causative Factors" and "Disposal Measures," benefiting from CRF's ability to model label dependencies. BERT + BiLSTM + CRF achieves better performance on "Intermediate Events," reflecting its stronger context awareness. MacBERT + CRF further improves on "Causative Factors" and "Accident Consequences," owing to its optimized pre-training strategy. The proposed R-MLP outperforms all other models across all labels, with particularly significant gains in "Initial Events" and "Control Measures," demonstrating its strong adaptability to complex contexts and diverse expressions.

4.5 Case analysis

To investigate the entity extraction results and their differences, this study conducts a case analysis, with the extraction performance of different models presented in Table 10. The RoBERTa + CRF model shows a high misclassification rate and struggles to handle

Table 9. Experimental results of different models

Model	P	R	F1
RoBERTa + CRF	0.212	0.117	0.151
BERT + CRF	0.613	0.522	0.564
BERT + LSTM + CRF	0.610	0.583	0.596
MACBERT + CRF	0.695	0.536	0.605
R-MLP	0.772	0.703	0.736

Source(s): Authors' own work

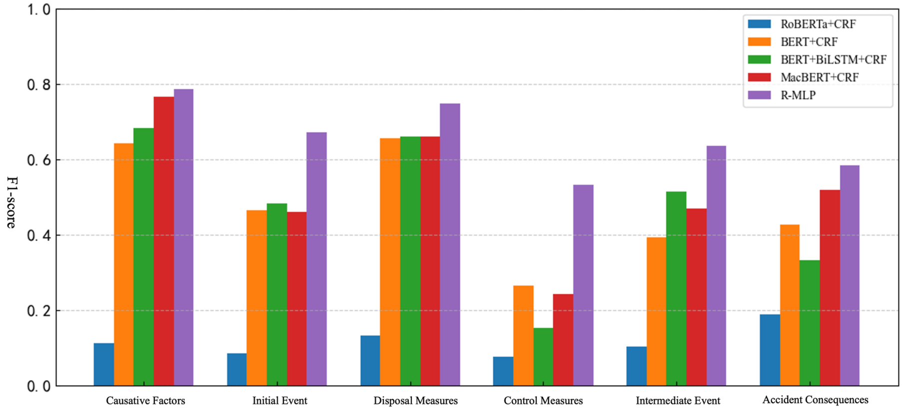

Figure 7. Comparison of F1 scores across different entity categories for each model. **Source(s):** Authors' own work

inter-sentence dependencies and domain-specific terminology. BERT + CRF and its BiLSTM variant achieve relatively complete extraction for fields such as “Control Measures” and “Disposal Measures,” but perform poorly on “Intermediate Events.” MacBERT + CRF performs well in “Causative Factors” and “Intermediate Events,” but its extraction of “Collapse” in the “Initial Event” category is overly generalized, lacking semantic refinement. In contrast, R-MLP achieves the best performance across all six entity types, accurately extracting complex intermediate events such as “rocks rolling onto the track” and producing outputs that closely align with human annotations in multiple key fields. Although its performance on “Intermediate Events” still has considerable room for improvement, R-MLP significantly outperforms other models. This advantage stems from its dual-branch architecture and mutual learning mechanism, which leverage the contextual information provided by the accident phase classification task to enhance the semantic perception capability of entity recognition, effectively addressing challenges such as sample scarcity, semantic complexity, and strong inter-sentence dependencies.

5. Conclusion

This study focuses on the entity extraction problem in railway accident texts, which present challenges such as strong domain specificity, complex semantic structures, and sparse entity distribution. To address these issues, a dual-branch mutual learning model, R-MLP, based on multi-task learning is proposed. The model employs a shared BERT encoder and jointly performs entity recognition and accident phase classification. A sentence span indexing module is introduced to resolve feature alignment issues arising from differences in task granularity, while a cross-task mutual learning mechanism is designed to achieve deep semantic fusion. Incorporating accident phase classification as an auxiliary task in entity recognition enriches the semantic context and helps the model capture stage-specific characteristics in the accident evolution process, thereby improving the accuracy and stability of entity recognition and extraction. Experimental results demonstrate that the proposed model significantly outperforms mainstream baselines in extracting multiple key entity categories, achieving a maximum F1-score of 0.736, and exhibiting stronger generalization and robustness. Ablation studies further validate the effectiveness of multi-task collaborative learning and cross-task mutual learning in enhancing entity extraction performance, providing a feasible and efficient solution for information extraction from domain texts with complex structures.

Table 10. Entity extraction performance of different models

Model	Causative factors	Initial event	Intermediate event	Control measures	Accident consequences	Disposal measures
Human Annotation	Rainfall, Earthquake	Slope Collapse	Rocks Rolling onto the Track	Warning to the Driver	Striking Fallen Rocks	Maintenance Handling
RoBERTa + CRF	Rainfall	Earthquake	/	Warning	Striking Fallen Rocks	Detecting Anomaly
BERT + CRF	Slope Collapse	Collapse	Crossing the Crash Barrier	Warning to the Driver	Stop	Maintenance Handling
BERT + LSTM + CRF	Earthquake	Highway Slope Collapse	Rocks Covering the Highway	Warning to the Driver	Striking Fallen Rocks and Stopping	Maintenance Handling
MACBERT + CRF	Rainfall, Earthquake	Collapse	Rolling onto the Track	Warning	Stop	Maintenance Handling
R-MLP	Rainfall, Earthquake	Slope Collapse	Rolling onto the Track	Warning to the Driver	Striking Fallen Rocks	Maintenance Handling

Note(s): Original text: Train XX came to a stop at the K211 + 985 mark on the up line due to a collision with fallen rocks. Upon inspection, it was found that the locomotive’s obstacle removal device and the right-side sand spreader were damaged. After maintenance work was completed, the train departed at 8:20 AM, and the section was reopened at 9:21 AM. Cause: Due to various factors such as rainfall and earthquakes, a landslide occurred on the local road embankment, causing rocks to pile up on the road surface. Some rocks crossed the road’s crash barrier and rolled down the left-hand embankment of the railway onto the tracks. Upon discovering the anomaly, the on-site inspection personnel alerted the driver. However, the train’s braking distance was insufficient, and a collision occurred during the braking process

Source(s): Authors’ own work

Zhibo Cheng is a researcher at China Academy of Railway Sciences Corporation Limited, specializing in research on railway big data analysis and safety data mining. She presided over or mainly participated in more than 20 scientific research projects, won 5 provincial and ministerial-level scientific and technological awards, and obtained 9 authorized national patents.

References

Chen, Y. (2023). Research on the construction and application of a railway accident knowledge graph. (Doctoral dissertation, Southwest Jiaotong University).

Du, W., Jin, Z., Dai, M., Xue, R., & Wu, S. (2023). Named entity recognition for railway freight one-price bargaining strategy based on the RoBERTa-BiLSTM-CRF model. *Railway Computer Applications*, *32*(5), 11–15. doi: 10.3969/j.issn.1005-8451.2023.05.03.

He, B., Zhao, R., & Tang, D. (2025). CABiLSTM-BERT: Aspect-based sentiment analysis model based on deep implicit feature extraction. *Knowledge-Based Systems*, *309*, 112782. doi: 10.1016/j.knosys.2024.112782.

Li, X., Li, G., Dai, M., Du, W., Zhao, Y., & Zhang, H. (2025). Application of large language models in intelligent operation and maintenance of railway infrastructure. *China Railway*, *2025*(1), 13–22.

Liang, J., Zhang, L., & Yan, S. (2024). Advances in research on named entity recognition based on large language models. *Computer Science and Exploration*, *18*(10), 2594–2615.

Lu, R., & Li, L. (2024). A named entity recognition model for legal documents. *Information Network Security*, *24*(11), 1783–1792.

Mei, X., Wu, X., Huang, Z., & Ling, J. (2023). Multi-scale semantic collaborative patent text classification model integrating RoBERTa. *Computer Engineering and Science*, *45*(5), 903–910.

Pooja, H., & Jagadeesh, M. P. (2024). A deep learning based approach for biomedical named entity recognition using multitasking transfer learning with BiLSTM, BERT and CRF. *SN Computer Science*, *5*(5), 482. doi: 10.1007/s42979-024-02835-z.

Shen, Z., Wang, Q., & Zhao, D. (2025). Multi-label classification CNN model for multi-tunnel resistance-change fault diagnosis in mine ventilation systems. *Journal of Safety and Environment*, *25*(5), 1822–1828.

Shi, L. (2023). Emergency scenario deduction and dynamic accident grading for railway emergencies. (Doctoral dissertation, Beijing Jiaotong University).

Shi, Y., Xi, N., Wang, G., Chen, J., Xu, H., & Zhao, Y. (2024). Classification and quantitative analysis method of railway traffic accident losses. *China Railway*, *2024*(4), 98–108.

Wang, W., Li, T., & Ge, H. (2023). Interacted grid tagging for sentiment word extraction. *Journal of Measurement Science and Instrumentation*, *14*(3), 369–378. doi: 10.62756/jmsi.1674-8042.2023042.

Xiao, L., & Chen, Z. (2024). A survey of data-driven Chinese entity extraction methods. *Computer Engineering and Applications*, *60*(16), 34–48.

Xu, Q., Zhu, P., Luo, Y., & Dong, Q. (2021). Research progress on Chinese named entity recognition in the financial domain. *Journal of East China Normal University*, *2021*(5), 1–13. doi: 10.3969/j.issn.1000-5641.2021.05.001.

Yang, J., Guan, Y., He, B., Qu, C., Yu, Q., Liu, Y., & Zhao, Y. (2016). Corpus construction for named entities and entity relations on Chinese electronic medical records. *Journal of Software*, *27*(11), 2725–2746. doi: 10.13328/j.cnki.jos.004880.

Yin, X., Zhao, H., Zhao, J., Yao, W., & Huang, Z. (2020). Named entity recognition in the military domain via multiple neural network collaboration. *Journal of Tsinghua University*, *60*(8), 648–655.

Zhang, N., Zhou, C., Wan, F., Liu, F., Wang, Y., & Xu, D. (2024). Tunnel construction safety domain named entity recognition based on BERT-BiLSTM-CRF. *China Safety Science Journal*, *34*(12), 56–63. doi: 10.16265/j.cnki.issn1003-3033.2024.12.0713.

Corresponding author
Zhibo Cheng can be contacted at: chengzhibo@rails.cn