

BCMGF: Feature Interaction with Glyph-Enhanced Representations for Chinese Agricultural Named Entity Recognition

Yongqi Zhang

Anhui Science and Technology University

Jinyu Zeng

Guizhou University

Chenguang Song

Anhui Science and Technology University

Xue Qi

Anhui Science and Technology University

Bingjing Jia

jiabj@ahstu.edu.cn

Anhui Science and Technology University

Article

Keywords: Named entity recognition, Global Pointer, Feature representation, Feature interaction attention, Adversarial training

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11011901/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

BCMGE: Feature Interaction with Glyph-Enhanced Representations for Chinese Agricultural Named Entity Recognition

Yongqi Zhang¹, Jinyu Zeng², Chenguang Song¹, Xue Qi¹,
Bingjing Jia^{1*}

^{1*}Anhui Science and Technology University, No.1501 Huangshan
Avenue, Bengbu, 233030, China.

²State Key Laboratory of Public Big Data, College of Computer
Science and Technology, Guizhou University, Guiyang, 550025, China.

*Corresponding author(s). E-mail(s): jiabj@ahstu.edu.cn;

Contributing authors: yjs2024148@ahstu.edu.cn;

yjs2023060@ahstu.edu.cn; songcg@ahstu.edu.cn; qix@ahstu.edu.cn;

Abstract

Agricultural information texts in Chinese face challenges such as a high density of rare domain-specific terms and ambiguous entity boundaries, which limit the accuracy of existing models in feature modeling and complex entity recognition for agricultural texts. To address these limitations, a novel agricultural named entity recognition method, termed BCMGE, is proposed, integrating feature representation enhancement with Global Pointer modeling. To enhance the representation capability for rare and visually similar characters, stroke-level features derived from CW2Vec are incorporated to compensate for the representation gap of agricultural domain vocabulary in general pre-trained models. Moreover, a Feature Interaction Attention Module (FIAM) for multi-level feature interaction, implemented using multi-head cross-attention and dynamic self-attention mechanisms, is employed to enable multi-round interaction and deep semantic fusion within character-level representations, thereby effectively capturing long-range dependencies and complex entity relationships. Finally, to improve the model’s robustness under noisy text conditions, adversarial training based on the Fast Gradient Method (FGM) is incorporated. Extensive experiments were conducted on an agricultural-domain dataset and three general-domain NER datasets. BCMGE achieves an F1 score of 85.1% on the AgriCHN dataset, yielding a 1.6 percentage-point improvement over the strongest baseline, HiNER.

On the general datasets MSRA, Resume and Weibo, F1 scores were 95.21%, 96.5%, and 71.91% respectively, validating the model's generalization capability. Compared with other models, this approach achieves more accurate and robust entity recognition in complex agricultural text scenarios, providing a potential NLP component for agricultural knowledge extraction and downstream smart agriculture applications.

Keywords: Named entity recognition, Global Pointer, Feature representation, Feature interaction attention, Adversarial training

1 Introduction

Named Entity Recognition (NER) is a fundamental task in natural language processing (NLP), aiming to identify entities with specific meanings from unstructured text, such as person names, geographical locations, and organization names, while also discovering previously unseen entities according to known entity composition patterns[1]. NER plays a pivotal role in a wide range of downstream applications, including information extraction[2], knowledge graph construction[3], question answering[4], and machine translation[5]. In recent years, deep learning approaches, exemplified by LSTM-CRF architectures, have substantially improved NER performance[6]; however, Chinese NER continues to face challenges arising from complex semantics, ambiguous entity boundaries, and nested entities[7].

The difficulties of Chinese NER are primarily manifested in three aspects. First, Chinese does not use explicit word boundaries, which can lead to substantial segmentation ambiguity. For instance, in the phrase "河南新郑红枣种植基地" (Henan Xinzheng red jujube planting base), different segmentation methods can directly affect the preliminary delineation of distinct entities such as "新郑红枣" (Xinzheng red jujube) or "红枣种植基地" (red jujube planting base). Second, polysemy is particularly prevalent in Chinese vocabulary. For example, "苹果" (apple) may refer either to a fruit or to the name of a company, thereby increasing the difficulty of contextual semantic interpretation. Finally, nested entity structures are widespread, where one entity is embedded within another, requiring models to simultaneously identify entity boundaries and hierarchical relationships. These challenges are particularly pronounced in the agricultural domain. Agricultural texts contain a large number of rare characters and domain-specific terms, such as "褐飞虱" (brown planthopper) and "稻瘟病型病斑" (rice blast-type lesions), while involving diverse entity types, including diseases, pests, symptoms, anatomical parts, and hosts. Nested and polysemous phenomena are therefore more frequent. As illustrated in Fig. 1, "小麦白秆病" (wheat white-stem disease) represents a complete disease entity while containing the crop entity "小麦" (wheat). The model must therefore determine from the contextual information that the entire phrase should be recognized as a disease entity, while the nested crop mention "小麦" (wheat) should be identified independently according to the annotation scheme, which imposes more stringent requirements on entity type classification. Similarly, in "西藏高寒麦区" (high-altitude cold wheat-growing region in Xizang), "高寒"

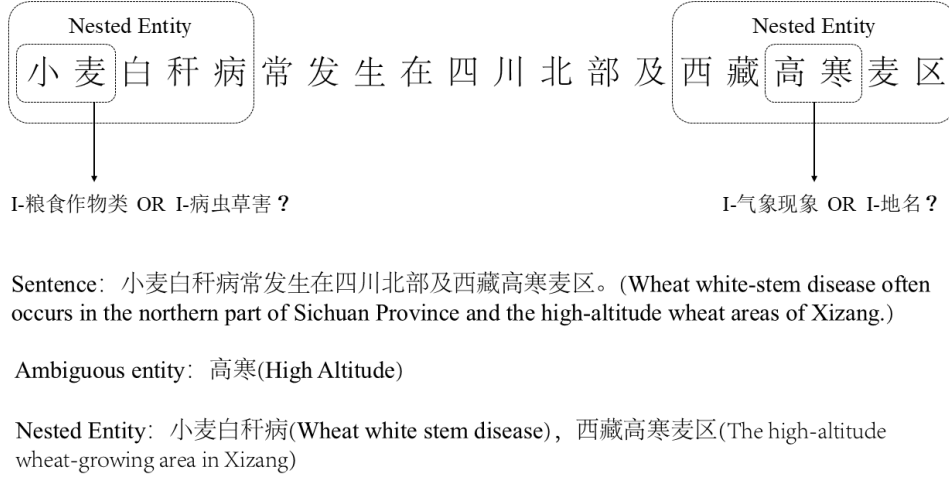


Figure 1 An example of a Chinese agricultural sentence containing ambiguous and nested entities. (Entity labels: I-粮食作物类 denotes grain crops, I-病虫害草害 denotes pests, diseases, or weeds, I-气象现象 denotes meteorological phenomena, and I-地名 denotes geographical locations.)

(cold climate) does not constitute an independent entity but forms a composite geographical entity together with "西藏" (Xizang) and "麦区" (wheat-growing region). If a model relies solely on local character- or word-level information, it may incorrectly classify "高寒" (cold climate) as a meteorological phenomenon, thereby disrupting the overall entity boundary. More broadly, the sentence simultaneously contains multiple nested entity structures. For instance, "西藏高寒麦区" (high-altitude cold wheat-growing region in Xizang) contains the geographical entity "西藏" (Xizang), as well as the descriptive modifier "高寒" (cold climate) and the domain-specific term "麦区" (wheat-growing region), resulting in entity boundaries characterized by hierarchical and dependent relationships. Such nested structures require models not only to identify entities but also to accurately model the inclusion relationships among them. This example demonstrates that the coexistence of nested entities, ambiguous entity boundaries, and polysemous words substantially increases the difficulty of Chinese NER in the agricultural domain. Character-level information alone is often insufficient to accurately distinguish entity types and their hierarchical relationships. More comprehensive integration of contextual semantics and domain knowledge is therefore required to achieve precise recognition of complex entity structures.

Although existing studies have explored NER tasks in the agricultural domain, agricultural NER as a whole continues to face three interrelated systemic challenges.

1. Semantic ambiguity and complex nested structures: Entities in agricultural texts frequently occur in nested forms. For instance, the disease entity "小麦白秆病" (wheat white-stem disease) contains the crop entity "小麦" (wheat). Moreover, identifying an entity often requires contextual information spanning multiple sentences or even

different fields to establish associations between symptom descriptions and disease names. Traditional sequence labeling models have difficulty modeling such complex hierarchical relationships and long-range semantic dependencies.

2. The representation gap between glyphs and semantics: Agricultural terminology contains a large number of low-frequency and visually similar Chinese characters, such as "芐", "芥", "螟", "蟥", "稂", and "耙". Because these characters occur relatively infrequently in the training corpora, general-purpose pre-trained models may fail to acquire sufficiently effective semantic representations for them, making entities with similar glyphs but different meanings highly susceptible to confusion.

3. Insufficient robustness caused by textual noise in non-standard inputs: Agricultural texts are commonly derived from optical character recognition (OCR), manual data entry, or colloquial records and may therefore contain typographical errors and non-standard expressions. Models trained on clean data may experience substantial performance degradation when exposed to such perturbations.

Among the aforementioned challenges, the presence of rare characters further exacerbates the difficulty of nested entity recognition, while noise additionally obscures glyph information and entity boundaries. Consequently, improvements to any single module are insufficient to address this composite challenge, highlighting the need for a systematic solution.

To address the aforementioned systemic challenges, this study proposes a fine-grained framework designed to improve feature interaction and robustness, termed BCMGF. The framework adopts a coordinated four-stage design consisting of "joint glyph-semantic representation, deep feature interaction, global span decoding, and noise-resistant training", thereby systematically addressing the core challenges of agricultural NER. To mitigate the representation gap between glyphs and semantics, CW2Vec stroke-level features are introduced on top of BERT-based semantic encoding to integrate structural glyph information with contextual semantics, thereby enhancing the model's ability to distinguish rare and visually similar characters. To address semantic ambiguity and complex nested structures, a Feature Interaction Attention Module (FIAM) is designed to perform multiple rounds of interactive refinement among character representations through multi-head cross-attention and dynamic self-attention, thereby strengthening the capture of long-range semantic associations and entity boundary features. Meanwhile, a span-based Global Pointer is employed as the decoder to address the limitations of traditional sequence labeling approaches in handling nested entities. To alleviate the adverse effects of textual noise, the Fast Gradient Method (FGM) adversarial training strategy is introduced to impose perturbations in the representation space, encouraging the model to learn smooth and stable decision boundaries and consequently maintain high robustness in noisy agricultural texts. Overall, the BCMGF framework is designed around the characteristics of agricultural texts being "long, heterogeneous, and challenging", forming an integrated information extraction system in which semantic-glyph fusion provides high-quality input representations, FIAM performs feature refinement, Global Pointer conducts structured decoding, and FGM ensures recognition stability.

In summary, the main contributions of this study are as follows:

(1) A systematic BCMGF framework is proposed to address the compound challenges of agricultural NER. This study organically integrates glyph enhancement, deep interactive attention, global span decoding, and adversarial training into a closed loop of representation, interaction, decoding, and noise resistance, thereby overcoming the limitations of conventional methods in recognizing nested entities, rare characters, and noisy scenarios.

(2) To bridge the glyph–semantic gap, a stroke-level feature deep fusion mechanism is designed. Through the synergistic representation of CW2Vec and BERT, the model achieves enhanced representation capability for visually similar characters and low-frequency agricultural terms.

(3) To capture long-range dependencies and nested structures, a Feature Interaction Attention Module (FIAM) is constructed. By integrating multi-head cross-attention and dynamic self-attention, the proposed module enables multiple rounds of interactive feature refinement among character representations, effectively capturing hierarchical relationships within nested entities while enhancing the modeling of long-range dependencies and fine-grained semantic relationships.

(4) Extensive experiments on both an agricultural NER dataset and general-domain datasets demonstrate the effectiveness of the proposed framework. The results show that BCMGF achieves favorable performance in both domain-specific agricultural scenarios and general NER settings, demonstrating its effectiveness across both domain-specific and general-domain NER benchmarks, and offering technical support for entity recognition in agricultural pest and disease texts.

2 Related Work

In recent years, named entity recognition (NER) has continued to attract attention and has gradually expanded from general domains to specialized scenarios such as agriculture and biomedicine. The successful applications of deep learning in smart grids[8], agricultural disease detection[9], medical diagnosis[10], IoT security[11], remote sensing classification[12], and smart cities[13] also demonstrate that the deep integration of domain-specific features and deep models can effectively improve complex data modeling performance. Early studies were mostly based on rule-based systems, extracting entities through domain dictionaries, grammatical templates, and pattern matching. Quimbaya et al.[14] combined dictionary matching with semantic rules for electronic health texts, and Chatterjee et al.[15] constructed the RENT model for agricultural corpora. However, such methods suffer from high manual costs and poor cross-domain portability. To reduce reliance on rules, statistical machine learning gradually became mainstream. Szarvas et al.[16] employed AdaBoostM1 and C4.5 decision trees to improve recognition capability in complex scenarios, and Li et al.[17] incorporated agricultural entity features into a CRF model. Nevertheless, manually designed feature templates still struggle to cope with irregular text structures, variable semantics, and ambiguous boundaries. With the development of deep learning, BiLSTM-CRF[18] became a classic sequence labeling architecture. Shibuya and Hovy[19] combined sub-optimal sequence learning with CRF decoding to handle nested entities, Zhang et al.[20] proposed Lattice LSTM to fuse word lattices with character interactions, Zhao

et al.[21] constructed Stroke2Vec by introducing stroke sequences to alleviate confusion among visually similar characters, and Li et al.[22] verified the modeling capability of deep models for long biomedical texts through CNN-BiLSTM-CRF.

In Chinese NER, character-level feature enhancement is particularly important for the recognition of low-frequency and visually similar characters. Existing works have explored granularities such as radicals, pinyin, glyph images, and stroke sequences. Among them, stroke sequences directly encode writing trajectories and do not rely on word frequency statistics, thereby providing stable structural discriminative signals for rare terms in agricultural texts. Therefore, this study adopts stroke sequence-based CW2Vec as the character enhancement module. In view of the strong specialization, large entity spans, and numerous nested structures in agricultural pest and disease texts, traditional BIO and BMES sequence labeling easily leads to error propagation and boundary prediction bias. Researchers have further proposed various modeling strategies: Zhang et al.[23] utilized hidden Markov models combined with rules to recognize nested entities, Xu[24] designed annotation rules for agricultural and forestry pests and diseases to clarify boundaries, Li et al.[25] proposed W2NER to uniformly model entity recognition as word-word relation classification, Zhang and He[26] fused stroke, radical, and glyph features to enhance character representations, and Wu et al.[27] integrated hierarchical Chinese character embeddings to improve the recognition of complex terms and symptom descriptions.

Overall, agricultural NER is evolving from rule-based and statistical methods toward deep methods that integrate semantic modeling, structural modeling, and domain knowledge, and is moving from sequence labeling to span classification and from character feature fusion to multi-granularity information integration. However, given the high heterogeneity, rich terminology, and cross-context semantic instability of agricultural texts, highly robust models that simultaneously possess global modeling, deep feature interaction, and adversarial training capabilities are still urgently needed.

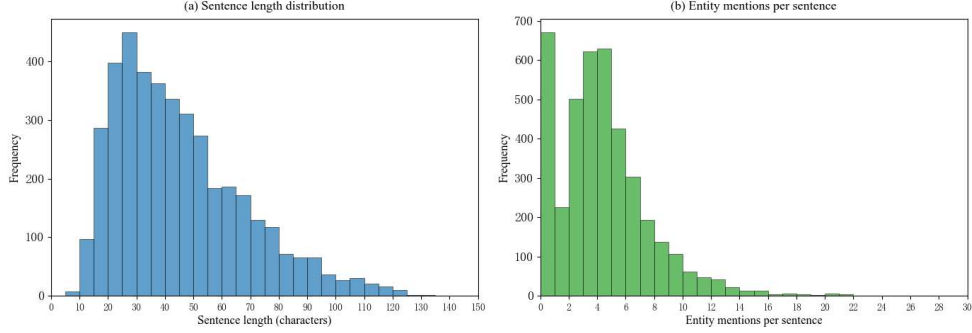
3 Materials and Methods

3.1 Dataset

This study uses AgriCHN[28] as the benchmark dataset for agricultural NER. AgriCHN is a Chinese agricultural NER dataset covering the three interdisciplinary fields of agriculture, hydrology, and meteorology. It was constructed through three rounds of iterative annotation by 20 annotators and 5 agricultural experts, encompassing three stages: double-blind annotation, expert arbitration, and model-assisted review. The dataset contains a total of 4,040 sentences and 15,799 entity annotations, covering 27 fine-grained entity types. The data are derived from 550 authoritative agricultural articles, 728 entity-free paragraphs, and 504 sentences from publicly available datasets. Table 1 summarizes the details of the data partition, and Figure 2(a)(b) illustrate the overall distributions of sentence length and entity density, respectively.

Table 1 Statistical table of the AgriCHN dataset.

Dataset	Number of Sentences	Number of Entities	Avg. Sentence Length	Avg. Entities per Sentence
Train set	3,230	12,859	45.8	3.98
Dev set	403	1,486	46.0	3.69
Test set	407	1,454	44.8	3.57

**Figure 2** Overview of the AgriCHN dataset: (a) sentence length distribution; (b) distribution of entity mentions per sentence.

AgriCHN covers 27 entity types. Figure 3 illustrates the distribution of each type in the training set, where a pronounced long-tail phenomenon is evident. In this study, entity types with fewer than 100 occurrences in the training set are defined as rare entities, including water bodies, livestock and poultry, hydrological phenomena, aquatic products, and diseases, with occurrence counts of 98, 95, 94, 30, and 25, respectively. Instances belonging to these rare entity types account for only 2.66% of all entity annotations. Although these entities represent key concepts in agricultural production and research, their low frequency makes it difficult for pretrained language models to acquire sufficient semantic representations. Therefore, this paper introduces CW2Vec stroke-level features to compensate for this deficiency.

AgriCHN contains numerous ambiguous boundaries and nested structures. In the sentence "甘蔗梢腐病，蔗田高温高湿或久旱遇雨或灌水过多易发病" (Sugarcane top rot disease: in sugarcane fields, high temperature and high humidity, or drought followed by rain, or excessive irrigation, can easily trigger the disease), "甘蔗梢腐病" (sugarcane top rot disease) is a disease entity, "高温高湿" (high temperature and high humidity) is a meteorological phenomenon entity, and as a holistic meteorological description, its internal components "高温" (high temperature) and "高湿" (high humidity) individually also belong to the meteorological phenomenon category and are prone to being misidentified as independent entities. Furthermore, "旱" (drought) and "雨" (rain) are each independent meteorological phenomenon entities, whereas the continuous expression "久旱遇雨" (drought followed by rain) constitutes a complete

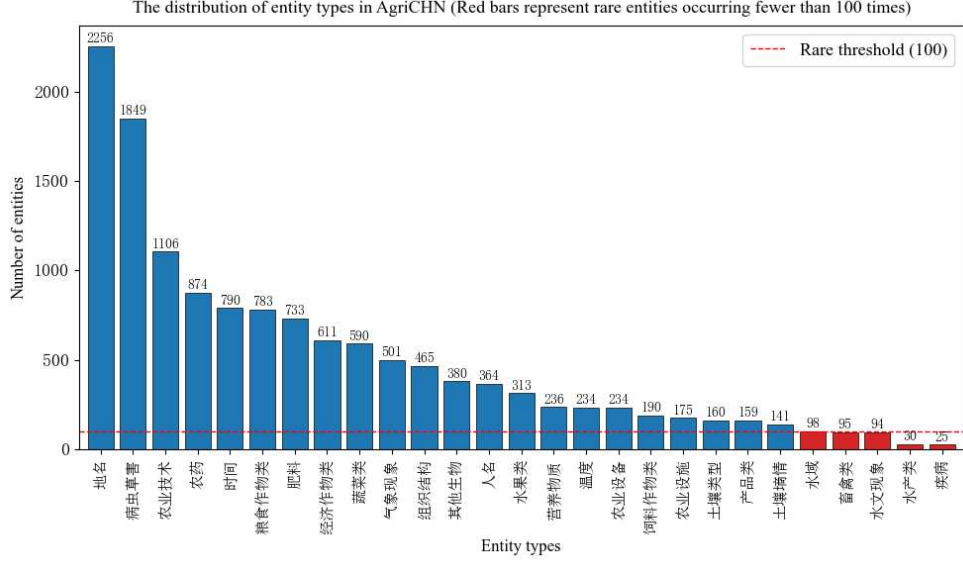


Figure 3 Distribution of 27 entity types in the AgriCHN training set (red bars indicate rare entities)

meteorological event. Therefore, the model must strike an accurate balance between recognizing fine-grained entities and understanding composite semantics.

3.2 Model Framework

The overall architecture of BCMGF is illustrated in Figure 4, and its design strictly corresponds to the three major challenges identified above. The text representation layer, combined with stroke information enhancement, jointly addresses the glyph-semantic gap challenge by generating enhanced representations that integrate contextual semantics and glyph structural information. The Feature Interaction Attention Module (FIAM) focuses on the challenges of semantic ambiguity and long-range dependencies by refining features through multiple rounds of cross-attention and self-attention. The span-based Global pointer addresses the challenges of entity nesting and overlap at the decoding level. Adversarial training with FGM targets the challenge of textual noise, thereby improving model robustness. These four components are closely interconnected, forming an information extraction system oriented toward the "long, complex, and difficult" characteristics of agricultural texts.

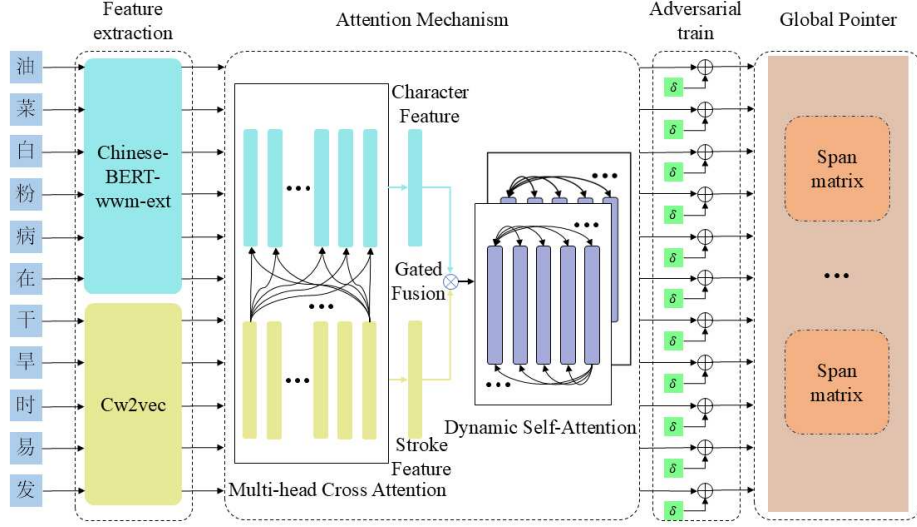


Figure 4 Architecture of the proposed BCMGF framework

3.3 Text Representation Layer

The original character sequence is defined as $S = \{c_1, c_2, \dots, c_n\}$, where n denotes the sequence length, and each character c_i represents a Chinese character, punctuation mark, or other basic symbol. To obtain contextual semantic representations of the sequence, Chinese-BERT-wwm-ext is adopted as the main encoder. This model is a pre-trained Chinese language model based on the Transformer architecture, built upon the Whole Word Masking strategy, which better preserves word-level semantic structures in Chinese.

Given the input character sequence S , the contextualized vector representation H learned through the Chinese-BERT-wwm-ext module can be expressed as Equation (1):

$$H = [h_1, h_2, \dots, h_n] \quad (1)$$

where $h_i \in \mathbb{R}^d$ denotes the contextual representation vector of the i -th character, and d is the hidden dimension of Chinese-BERT-wwm-ext, typically $d = 768$. The sequence representation $H \in \mathbb{R}^{n \times d}$ captures deep contextual semantic information obtained through layer-by-layer modeling of the multi-layer Transformer encoder, effectively modeling complex and long-distance semantic dependencies in agricultural domain texts.

3.4 Stroke Information Enhancement

To further address the glyph–semantic gap challenge, this study introduces CW2Vec stroke embeddings. Rare characters in agricultural terminology may be segmented by

BERT into single characters or subword fragments, leading to ambiguous semantic representations. Stroke sequences, as a stable glyph feature, can effectively distinguish visually similar characters with different semantics. Therefore, to compensate for the deficiency of semantic vectors, stroke embedding representations are introduced, where the function $\text{Glyph}(\cdot)$ maps each character to a stroke vector:

$$z_i = \text{Glyph}(c_i) \quad (2)$$

where $z_i \in \mathbb{R}^{d_c}$, scalar d_c denotes the dimension of the stroke vector (e.g., 128 or 256). The stroke vector sequence $Z = [z_1, \dots, z_n] \in \mathbb{R}^{n \times d_c}$ can provide discriminative information in scenarios where characters have similar strokes but different semantics. The semantic vector and the stroke vector are concatenated along the feature dimension to obtain a composite vector:

$$u_i = [h_i; z_i] \quad (3)$$

where $u_i \in \mathbb{R}^{d+d_c}$, the symbol $[\cdot; \cdot]$ denotes the vector concatenation operation along the last dimension. To ensure that the vector dimension fed into the subsequent attention module is consistent with that of the encoder, a projection matrix $W_p \in \mathbb{R}^{d \times (d+d_c)}$ and a bias $b_p \in \mathbb{R}^d$ are introduced to complete the mapping:

$$\tilde{h}_i = W_p u_i + b_p \quad (4)$$

Thus, the projected representation sequence $\tilde{H} = [\tilde{h}_1, \dots, \tilde{h}_n] \in \mathbb{R}^{n \times d}$ is obtained. In the following text, $\tilde{h}_i$ is regarded as the "enhanced feature" of the character c_i , which encompasses both contextual semantics and stroke information.

The design motivation for this section is as follows: semantic vectors excel at expressing contextual meaning, while stroke vectors excel at distinguishing visually similar and rare characters. The complementarity of the two can significantly reduce ambiguity and misjudgment in agricultural terminology, providing a more robust discriminative signal, especially under the influence of OCR noise and dialectal spelling variations.

3.5 Feature Interaction Attention Module

Although the base representation $\tilde{H}$ integrates contextual semantics and stroke-level glyph information, agricultural texts impose two structural demands that single-layer attention cannot adequately fulfill. First, as shown in Table 1, the average sentence length in AgriCHN reaches 45.8 characters, with entity mentions frequently spanning over 10 characters (e.g., "西藏高海拔小麦种植区" [Xizang high-altitude wheat-growing area] spans 11 characters). Such extended spans require the model to capture long-range dependencies between distant entity constituents. Second, agricultural sentences often exhibit multi-grained composability: a single utterance simultaneously conveys disease names, symptom descriptors, environmental conditions, and host crops, with these semantic roles distributed across non-adjacent positions, as in "黄化伴随粉状物出现在稻田低洼处, 疑似稻瘟病" (yellowing accompanied by powdery substances in low-lying paddy fields, suspected rice blast), where the disease mention "稻瘟病" (rice blast) at the end must establish dependency with symptom cues "黄化" (yellowing) and "粉状物" (powdery substances) over a 15-character distance.

Under such conditions, single-layer attention suffers from two limitations. On one hand, softmax normalization distributes probability mass across numerous weakly relevant tokens, diluting the signal from truly informative distant cues. On the other, complex multi-dimensional associations, such as those linking symptoms, environments, and diseases, cannot be adequately resolved by single-layer attention alone and demand iterative refinement. This empirical insufficiency is corroborated by our ablation results (Table 6): replacing FIAM with a single-layer self-attention yields an F1 of 0.835, substantially below the full model’s 0.851, with the performance gap primarily attributed to increased boundary errors, which account for 27.4% of total errors (Table 7). These observations motivate the design of a multi-round interactive refinement mechanism that progressively sharpens boundary discrimination and aggregates long-range semantic cues before decoding.

First, the projected text feature representation is duplicated into two sets of initial representations:

$$V^{(l)} = \tilde{H}, \quad Q^{(l)} = \tilde{H} \quad (5)$$

Here, $V^{(l)}$ and $Q^{(l)}$ represent the two sets of feature representations at the l -th layer, where $l = 1, \dots, L$. Although both branches are initialized from the same enhanced representation, they are subsequently projected using independent parameter matrices and updated iteratively through distinct query-key-value transformations, thereby enabling the two branches to learn complementary contextual views. The FIAM consists of two stages: first, a cross-feature interactive attention mechanism is introduced, as shown in Figure 5(a), with the objective of explicitly modeling the fine-grained correlations between the two sets of feature representations. In this stage, one set of feature representations serves as the query, and the other set serves as the key and value, enabling bidirectional information complementation between different feature representations through the attention mechanism. Given the input features $V^{(l)}$ and $Q^{(l)}$ of the l -th layer, the cross-feature interactive attention can be formally expressed as:

$$\alpha(V^{(l)}, Q^{(l)}, Q^{(l)}) = \text{softmax} \left(\frac{V^{(l)}(Q^{(l)})^T}{\sqrt{d}} \right) Q^{(l)} \quad (6)$$

where d is the feature dimension. To integrate information from multiple perspectives, we adopt a multi-head mechanism. Let z be the number of heads, and the computation for the i -th head is:

$$\text{head}_i = \alpha \left(V^{(l-1)} W_i^V, Q^{(l-1)} W_i^Q, Q^{(l-1)} W_i^{Q'} \right) \quad (7)$$

Here, $W_i^V, W_i^Q, W_i^{Q'} \in \mathbb{R}^{d \times d/z}$ are the trainable weight matrices for query, key, and value projections of the i -th attention head, respectively, designed to project the query, key, and value features into a subspace specific to each head. After concatenating the outputs of all attention heads along the feature dimension, information is integrated through a linear projection:

$$\tilde{V}^{(l)} = [\text{head}_1; \dots; \text{head}_z] W^O \quad (8)$$

where $W^O \in \mathbb{R}^{d \times d}$ is a trainable projection matrix used to map the concatenated features back to the original dimension. Symmetrically, using $Q^{(l-1)}$ as the query and

$V^{(l-1)}$ as the key and value, the same multi-head attention computation is performed to obtain the updated $\tilde{Q}^{(l)}$:

$$\tilde{Q}^{(l)} = [\text{head}'_1; \dots; \text{head}'_z] W^{O'} \quad (9)$$

where $W^{O'} \in \mathbb{R}^{d \times d}$ is another trainable projection matrix. Through this symmetric cross-feature interaction, the two sets of features can acquire each other's contextual information, forming complementary enhanced representations.

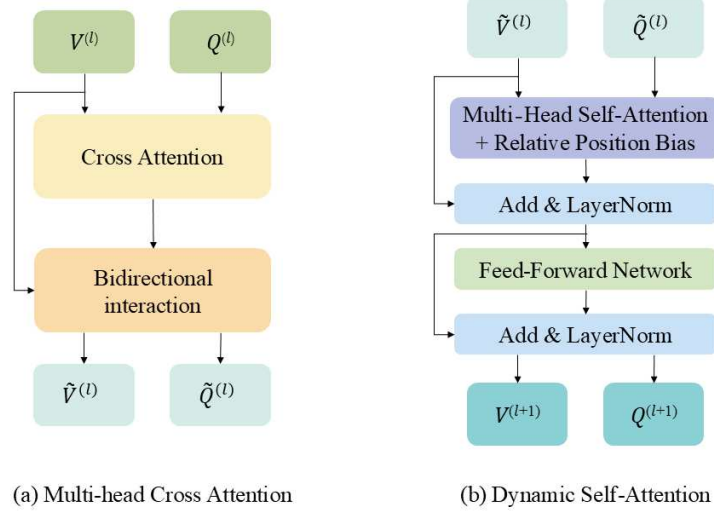


Figure 5 The multi-head cross-attention in the Feature Interaction Attention Module (FIAM) is employed to capture bidirectional fine-grained correlations between distinct feature representations, while the dynamic self-attention achieves gated feature fusion and self-attention refinement through relative position bias.

After cross-feature interaction, the model proceeds to the dynamic self-attention refinement stage, as shown in Figure 5(b). This stage first generates a feature-wise gating matrix from the current representations, thereby enabling adaptive amplification or suppression of each feature dimension. Specifically, the gate for V is defined as:

$$h_v^{(l)} = \tanh \left(\tilde{V}^{(l)} \mathbf{W}_v + \mathbf{b}_v \right) \quad (10)$$

$$h_q^{(l)} = \tanh \left(\tilde{Q}^{(l)} \mathbf{W}_q + \mathbf{b}_q \right) \quad (11)$$

$$g^{(l)} = \sigma \left(\left[h_v^{(l)}; h_q^{(l)} \right] \mathbf{W}_g \right) \quad (12)$$

Here, $\mathbf{W}_v, \mathbf{W}_q \in \mathbb{R}^{d \times d}$, $\mathbf{b}_v, \mathbf{b}_q \in \mathbb{R}^d$, $\mathbf{W}_g \in \mathbb{R}^{2d \times d}$ are trainable parameters, σ is the sigmoid function, and $[\cdot; \cdot]$ denotes concatenation along the feature dimension. The

gating vector $\mathbf{g}^{(l)} \in \mathbb{R}^{n \times d}$ is used to perform a weighted fusion of the two sets of features, resulting in the fused feature representation:

$$\mathbf{F}_v^{(l)} = \mathbf{g}^{(l)} \odot \tilde{\mathbf{V}}^{(l)} + (1 - \mathbf{g}^{(l)}) \odot \tilde{\mathbf{Q}}^{(l)} \quad (13)$$

where $\odot$ denotes element-wise multiplication. A symmetric operation is performed on $\mathbf{Q}$ to obtain $\mathbf{F}_q^{(l)}$.

Subsequently, we apply a self-attention mechanism to the fused features and introduce relative position encoding to enhance the model's perception of the sequence structure. Taking V as an example, its self-attention computation is:

$$\text{head}_i = \text{softmax} \left(\frac{\mathbf{Q}_i \mathbf{K}_i^T + \mathbf{P}_i}{\sqrt{d/z}} \right) \mathbf{V}_i \quad (14)$$

where:

$$\mathbf{Q}_i = \mathbf{F}_v^{(l)} \mathbf{W}_i^Q \quad (15)$$

$$\mathbf{K}_i = \mathbf{F}_v^{(l)} \mathbf{W}_i^K \quad (16)$$

$$\mathbf{V}_i = \mathbf{F}_v^{(l)} \mathbf{W}_i^V \quad (17)$$

are the query, key, and value projections for the i -th head, $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V \in \mathbb{R}^{d \times d/z}$ are trainable parameter matrices. $\mathbf{P}_i \in \mathbb{R}^{n \times n}$ is the relative position bias matrix, whose elements are determined by the relative distance between the query position and the key position. The multi-head outputs are concatenated, passed through a linear projection, and then through a feed-forward network with residual connections to obtain the final update:

$$\mathbf{O}_v^{(l)} = [\text{head}'_1; \dots; \text{head}'_z] \mathbf{W}_v^o \quad (18)$$

$$\mathbf{R}_v^{(l)} = \text{LayerNorm} \left(\mathbf{O}_v^{(l)} + \mathbf{F}_v^{(l)} \right) \quad (19)$$

$$\mathbf{V}^{(l+1)} = \text{LayerNorm} \left(\text{FFN}(\mathbf{R}_v^{(l)}) + \mathbf{R}_v^{(l)} \right) \quad (20)$$

Among them:

$$\text{FFN}(x) = \max(0, x \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2 \quad (21)$$

The same operation is performed on $\mathbf{Q}$ to obtain $\mathbf{Q}^{(l+1)}$. By continuously stacking such update units in an alternating fashion, the FIAM enables the sequence representations, through multiple rounds of cross-feature interaction and self-modality refinement, to progressively focus on features most relevant to entity boundary and type determination. Finally, after processing through L layers of FIAM, the original information is fused via a residual connection to obtain the enhanced representation:

$$\mathbf{H}^* = \mathbf{V}^{(L)} + \mathbf{Q}^{(L)} + \tilde{\mathbf{H}} \quad (22)$$

This residual fusion retains the original projected information while incorporating the enhanced information from multiple rounds of interaction. $\mathbf{H}^* \in \mathbb{R}^{n \times d}$ will serve as the input for subsequent adversarial training and Global Pointer entity recognition.

This mechanism is particularly important for agricultural texts because the iterative refinement process directly addresses the challenges of semantic ambiguity and complex nested structures, distinguishing BCMGF from approaches that rely solely on single-layer attention, because entity determination often depends on semantic cues distributed across positions. For instance, in the sentence "Yellowing accompanied by powdery substances was found in low-lying areas of the paddy field, suspected to be rice blast", the semantic associations among "yellowing", "powdery substances", and "rice blast" are difficult for single-round attention mechanisms to fully capture.

3.6 Span-based Global Pointer for Entity Recognition

In Chinese agricultural texts, entities often exhibit nesting issues. For example, the phrase "Southern China rice sheath blight" contains both a location name and a disease name. Traditional sequence labeling-based recognition methods are constrained by the linear label assumption and struggle to represent entity nesting and overlap inherent in nested structures. To this end, the model adopts a span-based global approach: for each entity type, an $n \times n$ start-end position scoring matrix is constructed, and all possible start-end pairs are directly classified, thereby jointly scoring all candidate spans in a global manner, thereby naturally accommodating nested and overlapping entities.

In this module, the input representation may be either the standard representation H^* or the adversarially enhanced representation H_{adv}^* generated during adversarial training. For clarity, the following formulas use the generic notation H° to denote either case, where $H^\circ \in \mathbb{R}^{n \times d}$, and the representation vector at each position is denoted as $\hat{h}_i$. The sequence representation $H^\circ = [\hat{h}_1, \dots, \hat{h}_n]$ is transformed through two linear mappings into a query vector sequence $\mathbf{Q}$ and a key vector sequence $\mathbf{K}$, respectively. Assuming there are C entity types, a separate set of projection parameters is learned for each type so that different types can be distinguished in the projected space. Formally, for entity type c , the projection results are expressed as:

$$\mathbf{Q}^{(c)} = [q_1^{(c)}, \dots, q_n^{(c)}], \quad \mathbf{K}^{(c)} = [k_1^{(c)}, \dots, k_n^{(c)}] \quad (23)$$

where $q_i^{(c)}, k_i^{(c)} \in \mathbb{R}^{d_h}$, and the scalar d_h denotes the inner dimension of the Global Pointer, which enhances representation capacity in the joint space of entity type and position. The index $c \in \{1, \dots, C\}$ represents the entity type, and each type corresponds to an independent set of query and key projection parameters. For each entity type c and each position pair (i, j) (where $1 \leq i \leq j \leq n$), the span score is computed as:

$$s_{(i,j)}^{(c)} = \frac{\langle q_i^{(c)}, k_j^{(c)} \rangle}{\sqrt{d_h}} \quad (24)$$

where $\langle \cdot, \cdot \rangle$ denotes the vector inner product, and the scaling factor $1/\sqrt{d_h}$ is used to stabilize the numerical range during training. To enhance sensitivity to relative positional information, Rotary Position Embedding (RoPE) can be incorporated after linear projection and before the dot-product computation. By applying sine and

cosine rotational phase transformations to even and odd dimensions respectively, the resulting score matrix simultaneously encodes relative positional information.

All type-specific scores are arranged into a three-dimensional tensor $S \in \mathbb{R}^{C \times n \times n}$. For each type c , the upper triangular region ($i \leq j$) of the matrix $S^{(c)} = [s_{(i,j)}^{(c)}]$ represents valid entity candidates. To prevent the model from attending to padding positions or invalid spans, a masking matrix is applied to S . Finally, a multi-label cross-entropy loss is used to train the model by comparing S with the manually annotated span matrix $Y \in \{0, 1\}^{C \times n \times n}$.

Compared with token-level sequence labeling methods, the span-based Global Pointer naturally handles nested and overlapping entities. Moreover, it performs normalized boundary prediction from a global perspective. In agricultural texts, this approach demonstrates clear advantages in recognizing long and nested entities. Figure 6 presents the corresponding matrix representation of entities within a sentence.

油	0	0	0	0	1	0	0	0	0	0	0
菜		0	0	0	0	0	0	0	0	0	0
白			0	0	0	0	0	0	0	0	0
粉				0	0	0	0	0	0	0	0
病					0	0	0	0	0	0	0
在						0	0	0	0	0	0
干							0	1	0	0	0
旱								0	0	0	0
时									0	0	0
易										0	0
发											0
	油	菜	白	粉	病	在	干	旱	时	易	发

Figure 6 An example of a Chinese agricultural sentence annotated by the span-based Global Pointer decoder. "油菜白粉病在干旱时易发" (Rapeseed powdery mildew is prone to occur during drought.)

3.7 Adversarial Training

To enhance the model's reliability under real-world agricultural texts (including noise from OCR, manual entry errors, dialectal expressions, aliases, etc.), this study introduces Fast Gradient Method (FGM) adversarial training during the training

process. FGM injects small perturbations along the gradient direction in the representation space, enabling the model to maintain discriminative stability when confronted with slight yet misleading input variations.

Specifically, let $L(\cdot)$ denote the supervised loss function based on the output of the Global Pointer. The standard representation H^* is obtained from the normal forward pass without perturbation using Formula 22, and the loss $L(H^*)$ is computed. The gradient of H^* is calculated as $g = \nabla_{H^*} L(H^*)$. FGM generates a normalized perturbation vector:

$$\delta = \epsilon \cdot \frac{g}{\|g\|_2} \quad (25)$$

where $\epsilon > 0$ is a hyperparameter controlling the perturbation magnitude, and $\|\cdot\|_2$ denotes the $L2$ norm of the vector. The perturbation is added to the enhanced representation to obtain the adversarial sample representation:

$$H_{\text{adv}}^* = H^* + \delta \quad (26)$$

Adversarial training employs a strategy involving two forward passes and loss combination, with the objective function formulated as:

$$L_{\text{total}} = L(H^*) + L(H_{\text{adv}}^*) \quad (27)$$

This approach encourages the model to minimize loss not only on normal samples but also to maintain discriminative robustness along the most destructive directions in the "vicinity" of normal samples. It should be emphasized that the FGM method adopted in this study incurs low computational cost and yields considerable robustness improvement for the span scoring structure of Global Pointer.

The proposed framework follows a four-stage design: First, stroke embeddings compensate for semantic blind spots. Second, multi-block interactive attention strengthens long-range and cross-feature semantic aggregation. Third, adversarial training is incorporated to enhance the stability of boundary discrimination. Finally, Global Pointer span modeling effectively resolves the recognition of nested and overlapping entities, thereby addressing key challenges in agricultural NER. For the task of agricultural pest and disease named entity recognition, this design not only improves the recognition of rare and visually similar characters but also maintains robustness under noise and distribution shift, making it suitable for practical application scenarios such as disease monitoring reports, agricultural inspection records, and historical archive analysis.

4 Results

4.1 Experimental Environment and Evaluation Metrics

The experimental model is built using the PyTorch deep learning framework, with AdamW adopted as the optimizer. The specific experimental environment for Chinese named entity recognition on the agricultural domain dataset is presented in Table 2, and the hyperparameter settings of the model are shown in Table 3.

Table 2 Experiment environment

Environment	Configuration
OS	Ubuntu 22.04
CPU	12 vCPU Xeon(R) Platinum 8352V
GPU	vGPU-32GB
RAM	90GB
PyTorch	2.3.0
Python	3.12

Table 3 Main parameter settings of the BCMGF model

Parameter	Value	Parameter	Value
Maximum sequence length	256	Batch size	32
Learning rate	3×10^{-5}	Weight decay	0.05
Training epochs	10	FIAM blocks	5
FIAM attention heads	8	Attention dropout	0.1

In terms of evaluation metrics, this study adopts the standard metrics in the field of named entity recognition: precision, recall, and F1-score. Precision measures the proportion of predicted entity spans that exactly match the ground truth entities, recall measures the proportion of ground truth entities that are correctly identified by the model, and the F1-score is the harmonic mean of precision and recall, providing a comprehensive assessment of the model’s overall performance. For scenarios involving nested entities, a strict span-based evaluation strategy is employed, where a predicted entity is considered correct only if both its type and its start and end positions exactly match the annotation.

4.2 Comparative Experiments and Analysis

To comprehensively validate the effectiveness of the proposed method, we compare BCMGF with several representative baseline models on the AgriCHN test set, including mainstream Chinese named entity recognition methods from the paper "AgriCHN: A Comprehensive Cross-domain Resource for Chinese Agricultural Named Entity Recognition", sequence labeling-based BiLSTM-CRF[29] and RoBERTa[30], the MTNER model[28], the recent HiNER model[31] designed for nested entities, the large language model-based Qwen2.5-7B[32], and the zero-shot multi-agent collaborative CMAS method[33]. The experimental results are presented in Table 4.

The traditional sequence labeling methods BiLSTM-CRF and SWA[34] achieve F1 scores of 0.764 and 0.770, respectively, achieving the lowest F1 scores among the evaluated models. Their linear labeling paradigm cannot model hierarchical relations among nested entities, resulting in low recall, with missed detection being the primary failure mode. After introducing BERT-series pretrained models, performance improves significantly, with BERT, Chinese-BERT-wwm, and RoBERTa achieving F1 scores of 0.801, 0.805, and 0.811, respectively. However, all three exhibit a pattern of "high

Table 4 Results of comparative experiment

Model	Precision	Recall	F1-score
BiLSTM-CRF	0.768	0.761	0.764
SWA	0.770	0.771	0.770
CMAS	0.783	0.805	0.794
Qwen2.5-7B	0.802	0.795	0.799
BERT	0.842	0.764	0.801
Chinese-BERT-wwm	0.845	0.769	0.805
RoBERTa	0.839	0.785	0.811
MTNER	0.850	0.802	0.825
HiNER	0.841	0.828	0.835
BCMGMF	0.858	0.844	0.851

precision, low recall", with recall ranging only from 0.764 to 0.785, indicating that the sequence labeling decoder still constrains the model's coverage of nested and long-span entities.

MTNER improves the F1 score to 0.825 through a multi-task learning mechanism. HiNER, based on a character-character relation classification framework, achieves a recall of 0.828 and an F1 of 0.835, demonstrating better recognition coverage for nested entities. However, HiNER's precision (0.841) is slightly lower than that of MTNER (0.850), reflecting the gap between local relation modeling and global boundary determination. The large language model-based Qwen2.5-7B and CMAS achieve F1 scores of 0.799 and 0.794, respectively, which are overall lower than those of domain-finetuned models, indicating that general-purpose LLMs, without sufficient domain adaptation, struggle to accurately capture professional terminology and nested structures in agricultural texts.

The proposed BCMGMF achieves the best precision (0.858), recall (0.844), and F1 score (0.851). Compared with the strongest baseline HiNER, the F1 score improves by 1.6 percentage points, with recall and precision increasing by 1.6 and 1.7 percentage points, respectively, achieving simultaneous improvements in both metrics. Compared with MTNER, the F1 score improves by 2.6 percentage points, and recall increases by 4.2 percentage points, further validating the advantage of global span decoding in alleviating entity missed detection. The performance advantage of BCMGMF stems from the collaboration of multiple modules: CW2Vec provides glyph discriminative information to enhance the representation of low-frequency agricultural terms; FIAM strengthens long-range dependency modeling through multiple rounds of cross-attention and dynamic self-attention; and Global Pointer natively supports nested entity recognition at the decoding level, avoiding label conflicts in sequence labeling.

In summary, comparative experiments demonstrate that BCMGMF significantly outperforms traditional sequence labeling, pretrained sequence labeling, and recent relation classification- and LLM-based baselines on the agricultural NER task, validating the effectiveness of the collaborative design framework of "glyph enhancement-deep interaction-global decoding-noise-resistant training".

4.3 Evaluation Results by Category

To further examine the fine-grained recognition capability of BCMGF across different entity types, Table 5 presents the precision, recall, and F1 scores of the model on all 27 entity types in the AgriCHN test set, sorted in descending order by the number of entities predicted by the model (Support). Overall, BCMGF achieves a micro-averaged precision of 0.858, a micro-averaged recall of 0.844, and a micro-averaged F1 of 0.851.

Table 5 Per-category recognition performance of BCMGF on the AgriCHN test set

Entity Type	Precision	Recall	F1-score	Support
地名	0.866	0.825	0.845	212
病虫害	0.920	0.920	0.920	200
农业技术	0.734	0.755	0.745	139
肥料	0.948	0.893	0.920	103
时间	0.844	0.800	0.822	95
经济作物类	0.851	0.889	0.870	90
农药	0.897	0.929	0.912	84
粮食作物类	0.929	0.890	0.909	73
气象现象	0.800	0.836	0.818	67
蔬菜类	0.871	0.984	0.924	62
其他生物	0.875	0.686	0.769	51
水果类	0.932	0.854	0.891	48
组织结构	0.696	0.744	0.719	43
人名	0.929	0.929	0.929	42
营养物质	0.800	0.800	0.800	35
农业设备	0.680	0.810	0.739	21
饲料作物类	0.950	0.950	0.950	20
土壤类型	0.824	0.778	0.800	18
产品类	0.571	0.500	0.533	16
土壤墒情	1.000	1.000	1.000	13
农业设施	0.667	0.364	0.471	11
水产类	1.000	0.250	0.400	4
水文现象	1.000	0.500	0.667	4
温度	0.800	1.000	0.889	4
疾病	1.000	1.000	1.000	3
水域	0.500	0.500	0.500	2
畜禽类	1.000	0.500	0.667	2
Micro-average	0.858	0.844	0.851	

From an overall perspective, the model attains high F1 scores on the majority of high-frequency categories. For example, "病虫害" (diseases, pests, and weeds) achieves an F1 of 0.92, "肥料" (fertilizer) an F1 of 0.92, "蔬菜类" (vegetables) an F1 of 0.924, and "粮食作物类" (grain crops) an F1 of 0.909, indicating that the model has strong recognition capability for core agricultural terminology. These categories typically exhibit relatively standardized expression forms and clear boundaries, with sufficient training data to facilitate feature learning. However, among them, the recall of "地名" (location names) is relatively low (0.825), partly because location names in

agricultural texts frequently co-occur with directional words and administrative levels in nested forms, making boundary delineation difficult.

For categories with moderate sample sizes, such as "经济作物类" (economic crops) with an F1 of 0.87, "农业技术" (agricultural technology) with an F1 of 0.745, and "组织结构" (organizational structure) with an F1 of 0.719, the recognition performance varies considerably across categories. The categories "农业技术" and "组织结构" suffer from low precision due to diverse terminological expressions and ambiguous boundaries. Meanwhile, "其他生物" (other organisms) shows a low recall of 0.686, reflecting that this category covers an overly broad range with diverse entity forms, making it difficult for the model to cover all potential expressions.

The performance of long-tail categories (Support < 20) fluctuates sharply. For instance, although categories such as "土壤墒情" (soil moisture), "疾病" (disease), and "温度" (temperature) achieve F1 values equal to or close to 1, their sample sizes are extremely small (no more than 13), limiting the reliability of category-level performance estimates. In contrast, "农业设施" (agricultural facilities) with an F1 of 0.471, "水产类" (aquatic products) with an F1 of 0.4, and "水域" (water bodies) with an F1 of 0.5 perform poorly, primarily due to the limited number of training instances and the fact that these entities often appear implicitly in the text, making boundaries difficult to define. It is noteworthy that although "产品类" (products) has 16 samples, its F1 is only 0.533, with both precision and recall being unsatisfactory. This is attributable to the large intra-category variation in entity forms and semantic overlap with other categories.

The above results indicate that BCMGF is highly competitive in handling common agricultural entities, yet there remains room for improvement in few-shot categories and categories with highly ambiguous boundaries. Future work may improve the recognition of long-tail categories through data augmentation, the incorporation of external knowledge, or the design of targeted loss functions.

4.4 Ablation Study Analysis

To comprehensively evaluate the independent contributions and synergistic effects of each key component in the BCMGF model, we conduct systematic ablation experiment on the basis of the Global Pointer (GP) framework by progressively removing or replacing core modules, including the stroke enhancement module (CW2Vec), the Feature Interaction Attention Module (FIAM), and the adversarial training mechanism (FGM). The experimental results are presented in Table 6.

Overall, the complete BCMGF model achieves the best precision (0.858), recall (0.844), and F1 score (0.851), validating the effectiveness of the multi-module collaborative design. Further analysis of the impact of each module reveals that different components contribute to model performance with distinct emphases. The introduction of the CW2Vec module improves the F1 score from 0.841 to 0.851, with its contribution mainly reflected in the improvement of precision. This indicates that by incorporating stroke-level structural information, the model can effectively distinguish visually similar and rare characters in agricultural texts, reducing type misclassification caused by glyph similarity and thereby lowering the false positive rate while maintaining relatively high recall. This result is consistent with the aforementioned

Table 6 Results of the ablation study

Model	Precision	Recall	F1-score
BCMGMF	0.858	0.844	0.851
w/o CW2Vec	0.844	0.839	0.841
w/o FIAM	0.836	0.825	0.830
w/o FIAM (replaced with Self-Attention)	0.830	0.841	0.835
w/o Global Pointer (replaced with CRF)	0.791	0.801	0.796
w/o FGM	0.849	0.834	0.842
w/o CW2Vec & FIAM	0.821	0.822	0.822

motivation for selecting stroke features and supports the effectiveness of stroke-level information in this setting.

The removal of the FIAM interaction module causes the F1 score to drop significantly from 0.851 to 0.830, with precision and recall decreasing by 0.022 and 0.019, respectively. Both metrics are substantially affected, with a slightly greater loss in precision. Through the alternating stacking of multi-block cross-attention and dynamic self-attention, FIAM realizes multi-round deep interaction of character-level features, effectively capturing long-range semantic dependencies and boundary cues among agricultural entities. Therefore, it is crucial for both boundary determination and type discrimination. Furthermore, when FIAM is replaced with a vanilla self-attention layer, the F1 score increases to 0.835. Although this is higher than directly removing FIAM, it remains far below the complete model, these results indicate that the multi-block cross-attention and dynamic gating mechanisms in FIAM outperform a single self-attention layer and better capture complex semantic interactions the complex correlations between semantic and stroke features.

Comparative experiments at the decoder level show that when CRF is used to replace Global Pointer, the F1 score drops sharply to 0.796, with simultaneous declines in precision and recall. This suggests that Global Pointer, supported by span-level classification and rotary position encoding, offers particular advantages in agricultural entity scenarios characterized by ambiguous boundaries and diverse types.

After removing FGM, the F1 score decreases by 0.009. The performance loss is relatively small but cannot be ignored. This indicates that adversarial training, by injecting gradient-direction perturbations into the representation space, enhances the model’s robustness to input noise and slightly expands its generalization boundary, contributing to stable performance in real-world agricultural text scenarios.

Finally, simultaneously removing CW2Vec and FIAM yields an F1 score of only 0.822, lower than the result of removing either module alone. The glyph discriminative information provided by CW2Vec and the deep interaction mechanism of FIAM cooperate with each other: the former enriches the feature input, while the latter further refines the fused representations. Together, they enhance the model’s ability to model complex agricultural entities.

In summary, the ablation study not only verifies the independent effectiveness of each module in BCMGMF but also clarifies their distinct emphases: Global Pointer serves as the core for handling nested entities, FIAM is the key to improving feature

representation and boundary precision, CW2Vec provides necessary information supplementation for low-frequency and visually similar characters, and FGM enhances the model’s stability under noisy conditions. The collaborative design of multiple modules enables BCMGF to achieve superior performance in agricultural named entity recognition tasks.

4.5 Error Analysis

To further understand the error patterns of BCMGF in agricultural named entity recognition, we conduct a categorized statistical analysis of all incorrectly recognized entities on the test set. According to the nature of the errors, they are classified into three categories: recognition omissions (entity omission), boundary errors, and type confusion. Table 7 presents the counts and proportions of these three error types.

Table 7 Distribution of recognition error types of BCMGF

Error Type	Count	Proportion (%)
Recognition omission	143	60.3
Boundary error	65	27.4
Type confusion	29	12.2

Recognition omissions mainly occur in scenarios where semantic expressions are implicit or entity forms are irregular. For example, the single-character meteorological entity "霜" (frost) in "荷, 喜温, 遇霜茎叶凋萎" (myoga ginger prefers warmth; when frost occurs, stems and leaves wither) is completely missed; the pesticide entity "马拉松" (malathion) in "可用50%马拉松800倍液" (can be treated with 50% malathion diluted 800 times) is missed by the model; and the agricultural equipment entity "TBS围栏" (TBS fence) in "设置1个60米TBS围栏" (set up one 60-meter TBS fence) is likewise omitted. These entities are characterized by low frequency, abbreviations, or mixed domain-specific names, making it difficult for the model to learn stable discriminative features. This reflects that the coverage capability of BCMGF for non-standard expressions remains insufficient.

Boundary errors mainly manifest as partial overlap between predicted and true entities, but with boundaries extended or contracted on either side. For instance, in "年均温21-23°C以上" (average annual temperature above 21–23°C), the model incorrectly incorporates the modifier "年均温" (average annual temperature) into the temperature entity "21-23°C以上" (above 21–23°C). In "大量元素水溶肥" (macroelement water-soluble fertilizer), the fertilizer entity "元素水溶肥" (element water-soluble fertilizer) is predicted as "大量元素水溶肥" (macroelement water-soluble fertilizer), with the boundary over-including the modifier "大量" (macro). Such errors indicate that when associated modifiers appear adjacent to entities, the attention mechanism tends to incorporate them into the entity span, causing boundary drift.

Type confusion occurs between entities with similar semantics or ambiguous category boundaries. For example, "腰果" (cashew) is predicted as a fruit category

although it should be classified as an economic crop; "瓢虫" (ladybug) and "食蚜蝇" (hoverfly) are predicted as pests, diseases, and weeds while actually belonging to other organisms (natural enemy insects). These confusions reflect semantic overlaps among fine-grained categories, especially in distinguishing crop types and biological categories, where the model relying solely on contextual and glyph features still struggles to achieve precise classification.

Overall, BCMGF exhibits strong recognition performance on common entity categories but still shows notable limitations when handling single-character entities, compound entities with modifiers, alphanumeric mixed expressions, and semantically similar categories. Future work may incorporate syntactic boundary information, design boundary regression auxiliary tasks, or employ data augmentation and category balancing strategies to enhance the model’s robustness to long-tail categories and complex expressions.

4.6 Generalization Across NER Benchmarks

To evaluate the generalization capability of the BCMGF model, we conduct evaluations on three general Chinese NER datasets: Resume, MSRA, and Weibo.

The Resume dataset is derived from executive resumes on Sina Finance, containing eight entity types such as nationality, job title, address, and educational background, with a relatively small scale. The MSRA dataset has a large volume and standardized annotations, including three entity types: person names, organization names, and location names. The Weibo dataset is derived from social media and contains highly colloquial and non-standard text, involving four entity types: person, organization, address, and political entity. To highlight the performance improvement of the proposed model, the following models are selected as baselines.

(1) Lattice LSTM[20]: This model enhances character sequence labeling by incorporating lexicon information, effectively avoiding the impact of word segmentation errors on recognition performance.

(2) LGN[35]: This model transforms the Chinese named entity recognition task into a graph node classification problem by constructing a graph structure between characters and words to model the complex dependencies between them.

(3) CGN[36]: By constructing a collaborative graph structure between characters and words, this model integrates both character-level and word-level contextual information, effectively alleviating the recognition challenges posed by Chinese word segmentation ambiguity.

(4) FLAT[37]: Its core innovation lies in the proposed Flat-Lattice structure. By introducing a lossless representation of lexicon information and relative position encoding, this structure can effectively model local dependencies and long-range context in text, thereby significantly enhancing model performance on Chinese sequence labeling tasks.

(5) LEBERT[38]: This model incorporates lexical information into the lower layers of the BERT architecture to enhance the model’s understanding of languages with ambiguous word boundaries, such as Chinese.

(6) AT-CBGP[39]: It presents a Chinese NER architecture built upon Global Pointer with adversarial training. It integrates rotary relative position encoding into the

adversarial training mechanism, thereby boosting global entity awareness and model robustness.

Table 8 presents the results of each model on the three datasets. On MSRA, BCMGF achieves the best F1 score (95.21%), outperforming Lattice LSTM (93.18%); it also achieves a slight improvement over AT-CBGP (95.12%), which similarly employs adversarial training. On Resume, BCMGF attains the highest F1 score (96.5%), slightly higher than AT-CBGP (96.48%). On Weibo, all models exhibit a marked performance decline, yet BCMGF still obtains the best F1 score (71.91%), with improvements over AT-CBGP (71.19%) and CGN (63.09%), indicating that the model possesses a certain degree of robustness to non-standard text.

Table 8 Experimental results on MSRA, Resume and Weibo datasets

Model	MSRA			Resume			Weibo		
	P(%)	R(%)	F	P(%)	R(%)	F	N(%)	G(%)	E(%)
Lattice LSTM	93.57	92.79	93.18	94.81	94.11	94.46	53.04	62.25	58.79
LGN	94.19	92.73	93.46	95.28	95.46	95.37	55.34	64.98	60.21
Char LSTM	90.97	86.96	88.81	93.66	93.31	93.48	-	-	-
+Bichar+softword	92.97	90.80	91.87	94.53	94.29	94.41	-	-	-
CGN	94.01	92.93	93.47	-	-	94.12	56.45	68.32	63.09
LEBERT	-	-	95.70	-	-	96.08	-	-	70.75
FLAT	-	-	94.12	-	-	95.45	-	-	60.32
AT-CBGP	95.33	95.16	95.12	96.38	96.63	96.48	-	-	71.19
BCMGF	95.54	94.88	95.21	96.60	96.40	96.50	-	-	71.91

It is noteworthy that the F1 score of BCMGF on the Weibo dataset is 71.91%. The relatively low performance is mainly attributed to its strong colloquialism, non-standard expressions, and high-noise characteristics, such as spelling variations, emoji mixing, and hashtag interference, which differ from the noise patterns in agricultural texts. CW2Vec stroke features and FGM adversarial training alleviate the effects of character deformation and input perturbations to a certain extent; however, when noise intensity is too high or entity boundaries are extremely blurred, the model still struggles to accurately identify entities. In addition, the entity type distribution and annotation standards of the Weibo dataset differ considerably from those of agricultural data, limiting transferability without domain adaptation. Future work may further improve performance by introducing social media pretraining or designing denoising modules.

4.7 Statistical Significance Test

To examine whether the performance improvement of BCMGF over the baseline HiNER is statistically significant, we train both models under the same five random seeds (42, 123, 2023, 7, 999) and record the entity-level micro-averaged F1 scores on the same test set. Since the two models share the same seed sequence, they form a natural paired design; therefore, a paired-sample *t*-test (one-tailed) is adopted to test the difference. The null hypothesis is that the mean F1 scores of the two models

are equal, and the alternative hypothesis is that the mean F1 score of BCMGF is significantly higher than that of HiNER.

Table 9 Comparison of entity recognition results between BCMGF and HiNER under five random seeds

Seed	HiNER			BCMGF		
	P	R	F1	P	R	F1
42	0.8309	0.8398	0.8353	0.8506	0.8413	0.8459
123	0.8426	0.8250	0.8337	0.8642	0.8399	0.8519
2023	0.8328	0.8277	0.8302	0.8575	0.8440	0.8507
7	0.8413	0.8277	0.8345	0.8558	0.8406	0.8482
999	0.8376	0.8297	0.8337	0.8648	0.8447	0.8547
Mean $\pm$ SD	0.8370 $\pm$ 0.0051	0.8300 $\pm$ 0.0060	0.8335 $\pm$ 0.0018	0.8586 $\pm$ 0.0055	0.8421 $\pm$ 0.0021	0.8503 $\pm$ 0.0032

The test results, as presented in Table 9, show that the mean F1 of BCMGF is 0.8503 ± 0.0032 , while that of HiNER is 0.8335 ± 0.0018 , yielding a mean difference of 0.0168 with a standard deviation of 0.0045. The calculated t -statistic is $t = 8.33$ with $df = 4$, corresponding to a one-tailed p -value of $p < 0.001$. This result is far below the significance level of 0.05; therefore, we reject the null hypothesis and conclude that the performance improvement of BCMGF is statistically highly significant and that this performance gain is unlikely to be explained by random initialization under the evaluated data split.

These results indicate that the advantage of BCMGF on the agricultural NER task is statistically reliable, providing quantitative evidence for its effectiveness.

4.8 Case Study

To further investigate the effectiveness of the proposed BCMGF model, we visualize the cross-feature interactive attention weights within the Feature Interaction Attention Module (FIAM), which plays a critical role in modeling interactions between enhanced feature representations.

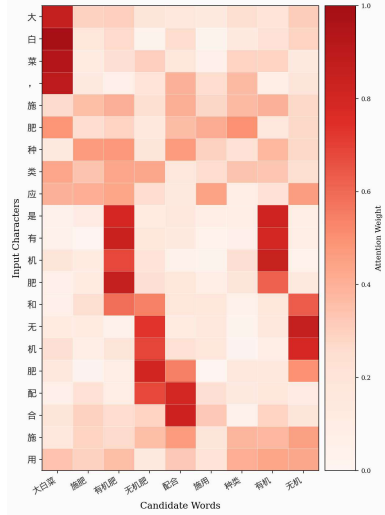


Figure 7 Visualization of cross-feature interactive attention weights in FIAM for an agricultural sentence involving fertilizer-related entities.

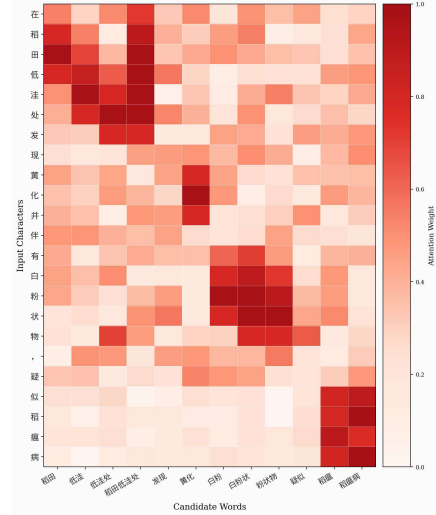


Figure 8 Visualization of cross-feature interactive attention weights in FIAM for a disease-related agricultural sentence.

As shown in Figure 7 and Figure 8, the attention distributions exhibit clear and interpretable patterns. Specifically, characters tend to assign higher attention weights to semantically relevant candidate words, demonstrating that BCMGF can capture fine-grained alignments across different granularities through FIAM.

In Figure 7, strong attention responses can be observed between key agricultural terms, such as "大白菜" (Chinese cabbage), "有机肥" (organic fertilizer), and "无机肥" (inorganic fertilizer), indicating that the model successfully aggregates semantic information from multiple granularities to enhance entity representation.

In Figure 8, the attention weights are more concentrated around disease-related expressions, including "白粉病" (powdery mildew) and "稻瘟病" (rice blast), showing that BCMGF, empowered by FIAM, further strengthens the discrimination of entity boundaries and types in complex contexts.

The attention maps provide qualitative evidence that the model assigns relatively high interaction weights to semantically related tokens, which is crucial for recognizing nested and domain-specific entities.

5 Discussion

BCMGF, by integrating CW2Vec stroke features, FIAM multi-block interactive attention, Global Pointer global span decoding, and FGM adversarial training, achieves performance superior to existing baselines on the AgriCHN dataset. However, certain limitations remain in few-shot categories, boundary localization of entities with modifiers, and cross-domain transfer on non-standard texts.

Future work could investigate data augmentation and category-balanced loss functions to alleviate long-tail effects. Boundary-aware auxiliary objectives and syntactic

dependency information could also be explored to improve entity boundary localization. In addition, parameter-efficient adaptation methods such as LoRA could be investigated for cross-domain transfer. At the application level, BCMGF has the potential to serve as an NLP component for agricultural information extraction and knowledge graph construction, automatically extracting entities such as crops, diseases, symptoms, pesticides, and geographical regions from pest and disease reports, agricultural technical documents, and historical archives, and constructing domain knowledge networks through entity linking and relation extraction. It can also be embedded into intelligent diagnostic systems for farmers to rapidly identify key entities in disease descriptions and provide structured input for matching prevention and control plans, while reducing manual review costs in agricultural public opinion monitoring and agricultural input market supervision. As the model is iteratively optimized on data from more agricultural subfields, the proposed model provides a potential NLP component for agricultural information extraction systems.

6 Conclusions

This paper proposes the BCMGF model to address the challenges of ambiguous entity boundaries, insufficient representation of rare characters, difficulties in modeling nested entities, and poor noise robustness in Chinese named entity recognition for the agricultural pest and disease domain. The model adopts Chinese-BERT as the encoder, introduces CW2Vec stroke features to enhance character representations, and realizes deep semantic interaction through multi-head cross-attention and dynamic self-attention. Global Pointer is employed for span-based entity decoding, natively supporting nested entity recognition, while FGM adversarial training is incorporated to improve noise robustness. Experiments on the AgriCHN dataset demonstrate that BCMGF outperforms existing baselines in precision, recall, and F1 score. Ablation studies verify the contributions of CW2Vec, multi-block interactive attention, and FGM to the overall performance. Results on general-domain datasets such as Resume, MSRA, and Weibo further demonstrate the model’s generalization capability. Overall, by systematically designing character representation, feature interaction, entity modeling, and training strategies, BCMGF provides an effective framework for Chinese named entity recognition in complex domains. Future work may explore larger-scale agricultural corpora and multimodal information fusion.

Data availability

The AgriCHN dataset used in this study is available from the original publication [28] and its repository. The MSRA, Resume, and Weibo datasets are standard public benchmarks for Chinese named entity recognition and can be obtained from their respective sources.

Acknowledgements

This research was funded by the Anhui Science and Technology University Science Foundation (Grant No. WDRC202103, XWYJ202301), the Anhui Provincial Department of Education Natural Science Major Project (Grant No. 2023AH040276).

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and experimental analysis were performed by Yongqi Zhang and Bingjing Jia. The first draft of the manuscript was written by Yongqi Zhang, who also conducted all experiments described in the paper. Bingjing Jia provided academic guidance and critical review of the manuscript. Jinyu Zeng offered textual suggestions on the manuscript. Chenguang Song and Xue Qi reviewed the manuscript to check formatting details and mathematical notation consistency. All authors read and approved the final manuscript.

Declarations

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

References

- [1] Jiao, K., Li, X., Zhu, R.: Overview of chinese domain named entity recognition. *Computer Engineering and Applications* **57**(16), 1–15 (2021)
- [2] Peng, B., Li, Y., Gong, X., *et al.*: Method for entity relation extraction based on heterogeneous graph neural networks and text semantic enhancement. *Computer Science* **51**(S1), 268–272 (2024)
- [3] Wu, S., Zhou, A., Xie, N., *et al.*: Construction of visualization domain-specific knowledge graph of crop diseases and pests based on deep learning. *Transactions of the Chinese Society of Agricultural Engineering* **36**(9), 177–185 (2020)
- [4] Khalid, N., Moulay, A.: Transformer models used for text-based question answering systems. *Applied Intelligence* **53**(9), 10602–10635 (2023)
- [5] Wang, X., Xu, Y., He, X., *et al.*: Reinforced negative sampling over knowledge graph for recommendation. In: *Proceedings of The Web Conference 2020*, pp. 99–109. ACM, New York, NY, USA (2020). <https://doi.org/10.1145/3366423.3380098>
- [6] Li, D., Luo, S., Zhang, X., *et al.*: Review on named entity recognition. *Journal of Frontiers of Computer Science and Technology* **16**(9), 1954–1968 (2022)

- [7] Liu, J., Xie, J.: Nested named entity recognition with global pointer network and contrastive learning. *Application Research of Computers*, 1–8 (2026) <https://doi.org/10.19734/j.issn.1001-3695.2025.04.0169>
- [8] Ali, A.M., Dawoud, O.K., Ghoneim, O.A., Abdel-Basset, M.: Machine learning models for smart grid stability prediction: a comparative analysis: A. ali et al. *Scientific Reports* **16**(1), 12694 (2026)
- [9] Alrashdi, I., Sharawi, M., Ali, A.M., Sallam, K.M., Arain, B., Abdel-Basset, M.: Utilizing deep learning models for early detection and classification of fruit diseases: towards sustainable agriculture and enhanced food quality. *Scientific Reports* **16**(1), 8167 (2026)
- [10] Sayed, E., Ali, A.M., Alrashdi, I., Sallam, K.M., Abdel-Basset, M., Ismail, M.M.: Sustainable diagnostic accuracy in healthcare 4.0: a comparative study of machine learning models with association rule-based feature reduction. *Cluster Computing* **29**(4), 224 (2026)
- [11] Hezam, I.M., Ismail, M.M., Ali, A.M., Sallam, K., Abdel-Basset, M.: Dp-dl-zt: a zero-trust-enhanced differential privacy framework with cnn-lstm for cyber threat detection in iot healthcare. *Neural Computing and Applications* **38**(8), 289 (2026)
- [12] Arain, B., Ali, A.M., Alrashdi, I., Sallam, K.M., Abdel-Basset, M.: Deep learning framework for land cover and land use classification: Five case studies with hyperspectral and rgb imagery. *Neural Computing and Applications* **37**(32), 26765–26822 (2025)
- [13] Sallam, K.M., Alrashdi, I., Ali, A.M., Radwan, I., Abdel-Basset, M.: A robust framework utilizing artificial intelligence to enhance services in smart city environments. *Cluster Computing* **28**(11), 741 (2025)
- [14] Quimbaya, A.P., M'unera, A.S., Rivera, R.A.G., *et al.*: Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Computer Science* **100**, 55–61 (2016)
- [15] Chatterjee, N., Kaushik, N.: Rent: Regular expression and nlp-based term extraction scheme for agricultural domain. In: *Proceedings of the International Conference on Data Engineering and Communication Technology (ICDECT 2016)*, pp. 511–522. Springer, Singapore (2016)
- [16] Szarvas, G., Farkas, R., Kocsor, A.: A multilingual named entity recognition system using boosting and c4.5 decision tree learning algorithms. In: *Proceedings of the International Conference on Discovery Science*, pp. 267–278. Springer, Berlin/Heidelberg, Germany (2006)
- [17] Li, X., Wei, X., Jia, L., *et al.*: Recognition of crops, diseases and pesticides named

- entities in chinese based on conditional random fields. Transactions of the Chinese Society for Agricultural Machinery **48**(Suppl.), 178–185 (2017)
- [18] Li, J., Sun, A., Han, J., *et al.*: A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering **34**(1), 50–70 (2020)
 - [19] Shibuya, T., Hovy, E.: Nested named entity recognition via second-best sequence learning and decoding. Transactions of the Association for Computational Linguistics **8**, 605–620 (2020)
 - [20] Zhang, Y., Yang, J.: Chinese ner using lattice lstm. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1554–1564. Association for Computational Linguistics, Stroudsburg, PA, USA (2018). <https://doi.org/10.18653/v1/P18-1144>
 - [21] Zhao, H., Yu, J., Lin, J.: Design and research of chinese character embedding model based on strokes. Journal of Chinese Information Processing **33**(4), 17–23 (2019)
 - [22] Li, L., Guo, Y.: Biomedical named entity recognition based on cnn-blstm-crf model. Journal of Chinese Information Processing **32**(8), 116–122 (2018)
 - [23] Zhang, J., Shen, D., Zhou, G.D., *et al.*: Enhancing hmm-based biomedical named entity recognition by studying special phenomena. Journal of Biomedical Informatics **37**(6), 411–422 (2004)
 - [24] Xu, S.: Research and implementation of question answering system based on the knowledge graph of wolfberry diseases and pests. Master’s thesis, Ningxia University, Yinchuan, China (2020)
 - [25] Li, J., Fei, H., Liu, J., *et al.*: Unified named entity recognition as word-word relation classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 10965–10973. AAAI Press, Palo Alto, CA, USA (2022)
 - [26] Zhang, Z., He, D.Z.: Chinese named entity recognition method based on multi-granularity embedding. In: Proceedings of the 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 2449–2453. IEEE, Piscataway, NJ, USA (2022)
 - [27] Wu, S., Song, X.N., Feng, Z.H.: MECT: Multi-metadata embedding based cross-transformer for Chinese named entity recognition. arXiv preprint arXiv:2107.05418 (2021)
 - [28] Zeng, L., Tong, Y., Guo, W., *et al.*: AgriCHN: A comprehensive cross-domain resource for Chinese agricultural named entity recognition. arXiv preprint arXiv:2506.17578 (2025)

- [29] Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
- [30] Liu, Y., Ott, M., Goyal, N., et al.: RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
- [31] Hou, S., Qian, Y., Chen, J., Zhao, J., Lv, H., Zhang, J., *et al.*: Hiner: Hierarchical feature fusion for chinese named entity recognition. *Neurocomputing* **611**, 128667 (2025)
- [32] Team, Q.: Qwen2.5 Technical Report. arXiv preprint arXiv:2412.15115 (2025)
- [33] Wang, Z., Zhao, Z., Lyu, Y., Chen, Z., Rijke, M., Ren, Z.: A cooperative multi-agent framework for zero-shot named entity recognition. In: *Proceedings of the ACM on Web Conference 2025*, pp. 4183–4195. ACM, New York, NY, USA (2025)
- [34] Yan, S., Chai, J., Wu, L.: Bidirectional gru with multi-head attention for chinese ner. In: *Proceedings of the 2020 IEEE 5th Information Technology and Electronics Engineering Conference (ITOEC)*, pp. 1160–1164. IEEE, Piscataway, NJ, USA (2020). <https://doi.org/10.1109/ITOEC49072.2020.9141551>
- [35] Gui, T., *et al.*: A lexicon-based graph neural network for chinese ner. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 1040–1050. Association for Computational Linguistics, Stroudsburg, PA, USA (2019)
- [36] Sui, D., Chen, Y., Liu, K., Zhao, J., Liu, S.: Leverage lexical knowledge for chinese named entity recognition via collaborative graph network. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 3830–3840. Association for Computational Linguistics, Stroudsburg, PA, USA (2019)
- [37] Li, X., Yan, H., Qiu, X., Huang, X.: Flat: Chinese ner using flat-lattice transformer. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6836–6842. Association for Computational Linguistics, Stroudsburg, PA, USA (2020). <https://doi.org/10.18653/v1/2020.acl-main.611>
- [38] Liu, W., Fu, X., Zhang, Y., Xiao, W.: Lexicon enhanced Chinese sequence labeling using BERT adapter. arXiv preprint arXiv:2105.07148 (2021)
- [39] Li, H., Cheng, M., Yang, Z., *et al.*: Named entity recognition for chinese based on global pointer and adversarial training. *Scientific Reports* **13**, 3242 (2023)