



Tomato Multimodal Knowledge Graph Construction and Dynamic Updating

Ziming Gao^{1,2,3,4,5}, Huaji Zhu^{1,2,3,4,5,*} and Yusen Yang^{2,3,4,5}

¹ College of Computer Science and Technology, Inner Mongolia Minzu University, Tongliao 028000, China

² National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

³ Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

⁴ Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs, Beijing 100097, China

⁵ Beijing Research Institute of Agricultural Artificial Intelligence and Robotics, Beijing 100097, China

Abstract

Tomato disease knowledge is scattered across agronomic texts and leaf images from heterogeneous sources, and static knowledge graphs that write extraction outputs directly to a database cannot separate model candidates from currently valid facts. We construct a multimodal knowledge graph for tomato diseases together with an evidence-gated dynamic update workflow. On the text side, a Chinese-RoBERTa-wwm-ext encoder with GlobalPointer identifies entity spans, and a directed relation pointer network extracts relations. On the image side, Chinese-CLIP fine-tuned with supervised contrastive learning aligns disease images with disease-text prototypes. On the graph side, stable entity and fact identifiers, source evidence and stateful writes support traceable maintenance, and a disease-centred interface visualises the graph. Named-entity recognition (NER) and relation extraction (RE)

achieved F1 scores of 0.7326 and 0.7042. In a source-group-disjoint alignment test, image-to-text Recall@1 was 0.8987, text-to-image Recall@1 was 0.9500, and prototype classification reached an Accuracy of 0.9295 and a Macro-F1 of 0.9317. For dynamic updating, evidence gating achieved Precision/Recall/F1 of 0.9643/0.9000/0.9310, and the false-admission rate fell from 0.2308 (direct writing) to 0.0357. Checking relation structure, local semantics and source evidence before admission therefore removed most erroneous facts, while every admitted fact remained linked to its evidence and could be replayed when models or rules change.

Keywords: multimodal knowledge graph, named-entity recognition, relation extraction, image-text alignment, dynamic updating.

1 Introduction

Tomato disease prevention and control involve interrelated information on disease names, symptoms, pathogens, affected plant parts and control measures. These materials are distributed across agronomic texts, research documents and leaf images, and a



Submitted: 11 August 2026

Revised: 01 October 2026

Accepted: 20 October 2026

Published: 22 October 2026

Vol. 2, No. 4, 2026.

doi:10.62762/TMI.2026.572797

*Corresponding author:

✉ Huaji Zhu

zhuhj@nercita.org.cn

Citation

Gao, Z., Zhu, H., & Yang, Y. (2026). Tomato Multimodal Knowledge Graph Construction and Dynamic Updating. *ICCK Transactions on Machine Intelligence*, 2(4), 224–235.



© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

single source rarely presents the full context of a disease. Knowledge graphs organise heterogeneous data through entities and relations and support querying, verification and knowledge acquisition [1]. In agricultural applications, previous studies have extracted tomato leaf pest and disease text into a graph stored in Neo4j [2], constructed CropDP-KG for crop diseases and pests in China [3], and developed the multimodal vegetable knowledge graph VEG-MMKG [4]. These studies show that domain graphs are moving from static terminology organisation towards automatic extraction and cross-modal knowledge services. However, whether a relation can be used as a current fact still depends on its source evidence, provenance and validity state.

Existing agricultural knowledge graphs generally combine rule or ontology organisation, pretrained-model extraction, graph-database materialisation and integration with recognition or question-answering systems. Surveys of automatic graph construction distinguish knowledge acquisition, refinement and evolution as connected stages [5]. Wang et al. [2] used ALBERT-BiLSTM-CRF to extract tomato pest and disease knowledge and provide graph visualisation; Yan et al. [3] built a cross-crop pest and disease graph from a Chinese image-text database; and Lv et al. [4] applied pretrained extraction to a multimodal vegetable graph. Other studies have combined agricultural graphs with large language models or graph neural networks for plant disease detection [6], or fused images, text and sensor data for disease detection and question answering [7]. These studies improve coverage, retrieval and downstream applications, but they seldom specify how recognition outputs, candidate triples and verified facts are kept apart, or which rules move a record from one layer to the next.

Agricultural text extraction is a major entry point for graph construction. Zhang et al. [8] used BERT, sequence labelling and relation classification to generate triples for Chinese crop diseases. Liang et al. [9] improved pest and disease entity recognition with RoBERTa-wwm and adversarial training, while Zhang et al. [10] addressed nested entities, ambiguous boundaries and class imbalance in wheat pest and disease text with data augmentation and rule correction. Tang et al. [11] further combined gated fusion and attention. At the relation level, Wang et al. [12] constructed a graph from expert texts on soybean pests and diseases, and Veena et al. [13] extracted agricultural triples with semi-supervised

and syntactic-semantic cues. These studies improve candidate acquisition, but entity errors can propagate through a pipeline, and domain labels and annotation cost limit transfer across sources. OneRel [14] and GlobalPointer-based joint extraction with relation filtering [15] offer alternatives for overlapping spans and relations, while Chinese pretrained-model research [16] provides the basis for text representation. Whether a pipeline or a joint extractor is used, model scores alone cannot show that an object type, relation direction or evidence location satisfies the requirements for graph admission.

Leaf images provide observations of disease phenotypes beyond text. Reviews of plant disease recognition show that data availability, imaging quality and separation of healthy and diseased classes continue to affect model performance [17]. Static ontologies in agriculture struggle to accommodate the dynamic knowledge acquired from heterogeneous sources; ontology-based models that support dynamic updates and structured knowledge reuse provide a more maintainable foundation for domain graph construction [18]. A review of tomato disease and pest studies using real-field data notes that results on controlled-background datasets may overestimate application performance [19], and plant-disease dataset research also emphasises differences between collection and deployment scenes [20]. Feature-level fusion of tomato disease images with environmental parameters demonstrates the complementary value of multiple data sources for classification [21]. These studies generally output recognition classes or detection targets. Even when an image resembles a disease class, it cannot by itself prove a specific pathogen or control relation, nor can it replace the treatment conditions given in the text.

Vision-language models establish shared representations for images and natural-language descriptions. Chinese-CLIP supports cross-modal retrieval through Chinese image-text contrastive pretraining [22]. Few-shot vision-language classification of crop diseases uses image descriptions and prompts to construct class representations [23], cross-modal fusion models combine leaf images with lesion-text features [24], and multimodal large language models have been used for crop pest and disease image question answering [25]. These approaches expand class association and semantic interaction, but their objectives are generally classification, retrieval or answer generation. The resulting similarity scores and textual outputs are

not field-diagnosis probabilities and do not carry independent source evidence for graph relations. We therefore use image-text alignment to organise candidate image observations and disease-text prototypes, rather than writing class-level similarity directly into the graph as a definite disease-control fact.

Reliable graph maintenance also requires contextual validity, temporal information and provenance. Knowledge graph surveys distinguish knowledge acquisition, completion and temporal knowledge graphs as separate research tasks [26]. Hofer et al. [27] note that one-off graph construction is relatively mature, whereas the coordination of incremental updating, metadata management and quality assurance remains to be systematised. Temporal knowledge graph completion uses temporal validity to improve missing-link prediction [28], but predicting a potential link is not equivalent to verifying a fact supported by newly ingested material. Provenance-aware knowledge representation has discussed several data models that attach provenance to relation statements [29], and provenance-aware knowledge graphs have been constructed from agricultural documents by combining named-entity extraction with semantic annotation models, enabling traceable querying of domain knowledge [30]. Dynamic updating and provenance are therefore not new in themselves. The question we address is a practical one: how to decide whether repeated records come from the same source, how to keep evidence locations and model versions, and how to prevent new evidence, conflicts and state changes from being hidden by simple append-only writes.

In short, domain extraction improves candidate coverage, vision-language models strengthen image-text class association, and work on graph construction, temporal modelling and provenance [27–30] provides frameworks for organising and evolving knowledge. What is still missing is a checkable interface between these components before a candidate becomes a current fact. We formulate tomato disease knowledge construction as a workflow linking candidate generation, evidence verification, state management and graph materialisation. Relation signatures, ontology constraints, object atomicity, entity grounding and local semantics are checked first; evidence is then aggregated by source group and fact states are maintained; finally, only facts satisfying the admission conditions are written to Neo4j, while

candidates and evidence chains are retained. The main contributions of this work are as follows.

- We construct a Chinese tomato-disease information-extraction module covering nine entity types and seven relation types, using Chinese-RoBERTa-wwm-ext with a bidirectional LSTM, GlobalPointer, span-length bias and R-Drop, and compare it with baseline and ablated settings under a fixed protocol.
- We establish a source-group-disjoint ten-class image-text alignment benchmark based on Chinese-CLIP and supervised contrastive learning, and report bidirectional retrieval and class-prototype classification.
- We propose an evidence-gated dynamic update method that combines independent-source noisy-OR fusion with a fact state machine, and compare it with direct writing and static-threshold strategies on the same candidate stream.
- We materialise accepted facts in Neo4j using stable identifiers and provide source tracing and visualisation around disease nodes through the Fact-Evidence-Document-ImageObservation chain.

The remainder of this paper is organised as follows. Section 2 introduces the materials and methods. Section 3 reports the experiments, dynamic-update comparisons, ablations and visualisation. Section 4 discusses the results and future work, and Section 5 concludes the paper.

2 Materials and Methods

To keep interfaces stable across modules, candidate records use the common fields `subject`, `predicate`, `object`, `entity_score`, `relation_score`, `evidence`, `source_id`, `observed_at` and `fact_id`. Text character ranges allow a candidate to be traced back to its source, whereas `image_id` and `source_group` identify image samples. Both record types can carry model versions, run times and review states. With this shared schema, outputs from both channels can be audited, ranked, merged and written as intermediate records.

The basic data flow is as follows. Raw text or images enter a normalisation layer and are processed by the text extractor and the image-text aligner to generate scored candidates. Candidates are checked for entity type, relation signature, evidence location and source identifier before entering the dynamic updater. The updater changes Fact states according to temporal decay, independent-source fusion and gating results.

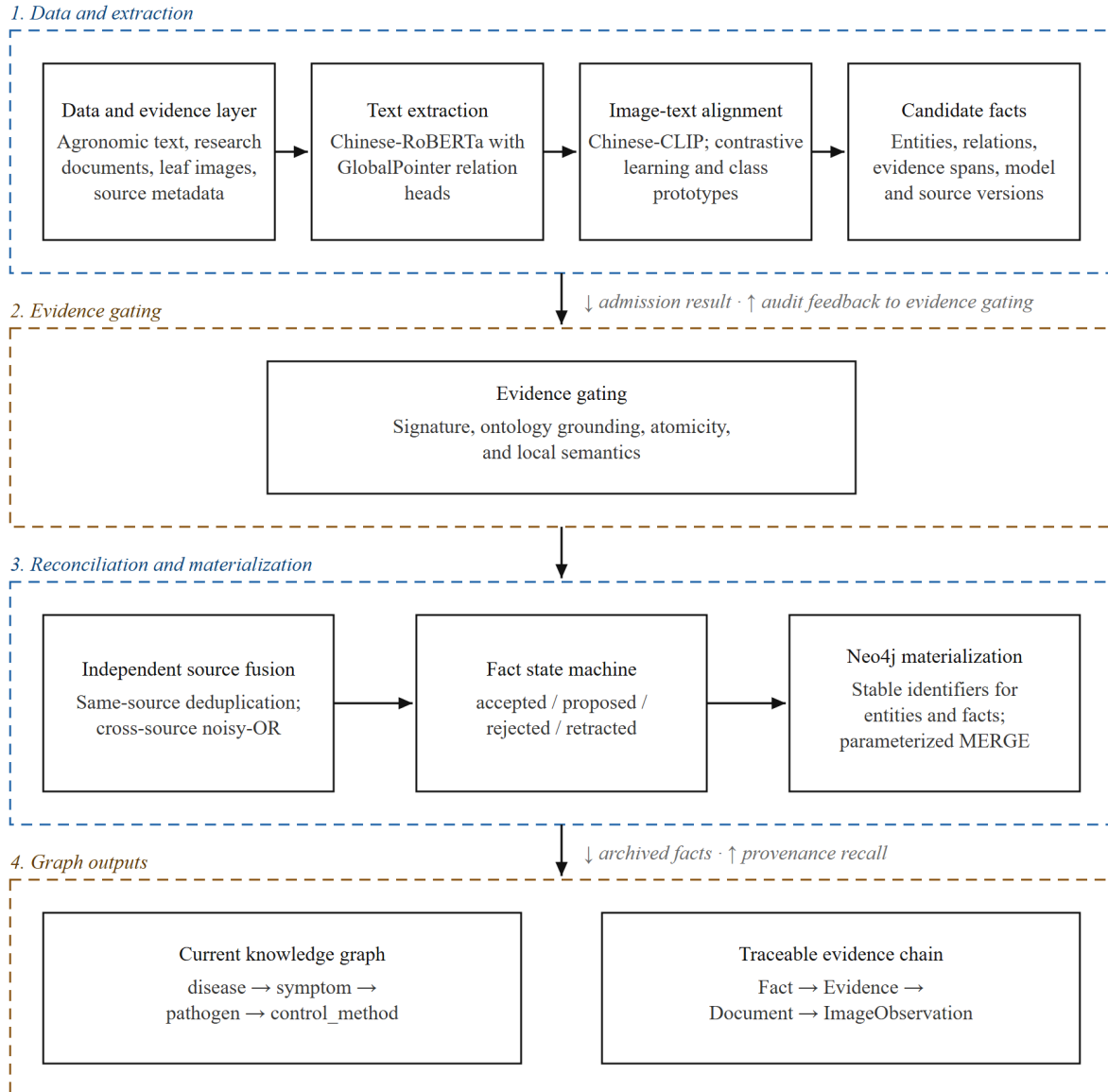


Figure 1. Overall framework for the evidence-gated dynamic update of a tomato multimodal knowledge graph.

A Neo4j writer then idempotently materialises nodes, evidence and current relations. Every stage retains its inputs, outputs and run batches, so the same data can be recomputed after rule updates without being collected again.

This section presents the overall framework, data sources, named-entity recognition, relation extraction, image-text semantic alignment, evidence-gated dynamic updating, knowledge graph materialisation and visualisation, and evaluation metrics.

2.1 Overall framework

Figure 1 shows the framework, which comprises data and extraction, evidence gating, reconciliation and materialisation, and graph outputs. Agronomic

texts, research documents, leaf images and source metadata form the data and evidence layer. Chinese-RoBERTa-wwm-ext [16] with GlobalPointer discovers text candidates, while Chinese-CLIP [22] contrastive learning and class prototypes organise image-text associations. Both channels are normalised as candidate facts with evidence spans and model versions.

Candidate facts pass through signature checks, ontology grounding, object-atomicity checks and local semantic verification. The reconciliation layer deduplicates same-source candidates, aggregates independent-source evidence with noisy-OR and manages the accepted, proposed, rejected and retracted states. Parameterised MERGE statements then

write stable identifiers to Neo4j. The output layer exposes a disease-symptom-pathogen-control-method graph and a traceable chain of Fact, Evidence, Document and ImageObservation nodes.

2.2 Data sources

Text data. Textual materials were collected from agricultural technology websites, specialised books on pest and disease control and technical standards. After collection, deduplication and manual curation, structured entries covered nine major tomato diseases, including tomato early blight and tomato late blight, together with healthy tomato. Each record contains the disease name, aliases, affected plant part, symptoms, epidemic rules, causes, control methods and recommended pesticides. After cleaning and segmentation, source texts were divided into training and validation sets. The RE corpus was built from the same sources after removing ambiguous or missing spans, coreferential disease names, duplicate triples and duplicate texts.

Image data. Images were obtained mainly from the public PlantVillage disease-image repository [31]; classes not covered by PlantVillage, such as tomato powdery mildew, were supplemented by field photographs collected by the authors in Tongliao, Inner Mongolia, China (2025). Images were organised into ten class directories and passed quality review for clarity, label accuracy and duplication.

The image-text collection comprised 2,132 paired samples distributed across ten categories, and the text corpus yielded 3,962 annotated entities across the ontology. Neither distribution was dominated by a single category; disease and control-method entities were the most frequent entity types. Figure 2 summarises this coverage, while Table 1 shows the paired format used to align a visual observation with a concise disease description.

The common ontology contains the entity types disease, alias, part, symptom, rule, method, medic, pathogen and reason. The relation types are alias, affected_part, symptom, epidemic_rule, cause, control_method and pesticide_control. Each relation specifies its subject and object types and its evidence requirements. For example, pesticide_control maps disease to medic (pesticide), whereas cause may map disease to pathogen or reason.

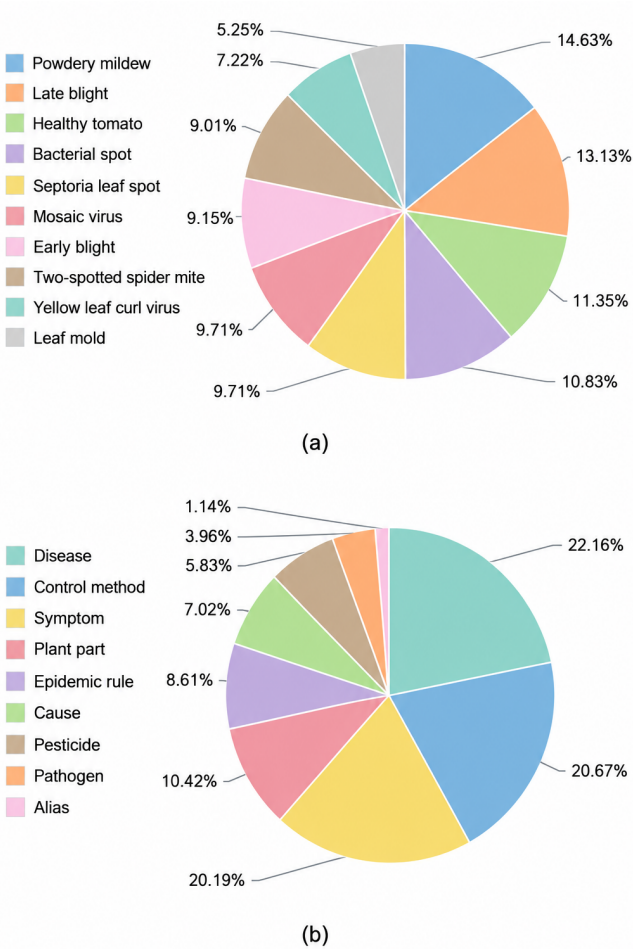


Figure 2. Dataset composition. (a) Percentage of image-text samples by disease category; (b) percentage of extracted entities by entity type.

Table 1. Representative image-text pairs and disease descriptions.

Entity type	Text description	Image
Disease	Early blight is identified by the colour, shape, size and concentric-ring pattern of the leaf lesions, together with recent temperature and humidity changes.	
Disease	Powdery mildew lesions enlarge and merge, or cover the whole leaf, so that the leaf surface looks as if it had been dusted with flour.	

2.3 Named-entity recognition

Chinese-RoBERTa-wwm-ext [16] is followed by a bidirectional LSTM, which adds local sequence information. Given a token sequence x of length L , the encoder output is $H = [h_1, \dots, h_L]$:

$$H = \text{Encoder}(x), \quad h_i \in \mathbb{R}^{d_e}, \quad 1 \leq i \leq L. \quad (1)$$

GlobalPointer [32] models entity spans directly. For compound symptoms, pesticide names and multi-character disease names, it scores start and end positions in a shared two-dimensional space rather than relying on concatenated BIO labels. For entity type c , h_i is projected to a query q_i^c and a key k_j^c of dimension d_h , and the span score is

$$S_{ij}^c = \frac{(q_i^c)^\top k_j^c}{\sqrt{d_h}} + b_{c, \min(j-i+1, M)}, \quad 1 \leq i \leq j \leq L, \quad (2)$$

where $b_{c,\ell}$ is a learnable span-length bias that provides a prior over typical entity lengths, and M is the length beyond which all spans share one bias bucket.

During training, each entity type uses a start-end matrix and a multilabel classification loss over valid tokens. The span loss over the set V_b of valid upper-triangular spans in sample b is

$$\mathcal{L}_{\text{span}} = \text{Avg}_{b,c} \left[\log \left(1 + \sum_{\substack{(i,j) \in V_b \\ y_{ij}^c=0}} e^{S_{ij}^c} \right) + \log \left(1 + \sum_{\substack{(i,j) \in V_b \\ y_{ij}^c=1}} e^{-S_{ij}^c} \right) \right]. \quad (3)$$

R-Drop [33] runs two stochastic forward passes with different dropout masks and constrains their outputs to be consistent. Following our implementation, the consistency term is the mean squared error between the sigmoid span probabilities, with weight 0.5, so the NER objective is

$$\mathcal{L}_{\text{NER}} = \frac{\mathcal{L}_{\text{span}}^{(1)} + \mathcal{L}_{\text{span}}^{(2)}}{2} + \frac{1}{2} \text{MSE}_{V_b} (\sigma(S^{(1)}), \sigma(S^{(2)})). \quad (4)$$

At inference, a span is decoded as an entity when its logit exceeds zero. The same boundary is used for all types, and no threshold is tuned on the evaluation data. The boundary auxiliary loss is enabled only in the boundary-aware configuration.

2.4 Relation extraction

Relation candidates are generated with a valid-token mask and mapped back to original character positions. Multiple entity pairs can produce multiple candidates, which are aggregated by fact identifier without overwriting different objects. A directed relation pointer head produces a subject query q_i^r and an object key k_j^r for relation r :

$$S_{ij}^r = \frac{(q_i^r)^\top k_j^r}{\sqrt{d_h}}, \quad 1 \leq i, j \leq L. \quad (5)$$

With P_b denoting the set of ordered non-padding token pairs, the relation loss is

$$\mathcal{L}_{\text{RE}} = \text{Avg}_{b,r} \left[\log \left(1 + \sum_{\substack{(i,j) \in P_b \\ y_{ij}^r=0}} e^{S_{ij}^r} \right) + \log \left(1 + \sum_{\substack{(i,j) \in P_b \\ y_{ij}^r=1}} e^{-S_{ij}^r} \right) \right]. \quad (6)$$

The extraction stage targets high recall; final admission is determined by evidence gating.

2.5 Image-text semantic alignment

Chinese-CLIP [22] encodes images and text as ℓ_2 -normalised representations and is fine-tuned with a supervised contrastive objective [34]. Let B be the batch size, z_i and t_p the normalised image and text representations, $P(i)$ the indices of texts in the batch that share the label of image i , and $\tau = 0.07$ the temperature. The image-to-text loss is

$$\mathcal{L}_{i \rightarrow t} = -\frac{1}{B} \sum_{i=1}^B \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot t_p / \tau)}{\sum_{j=1}^B \exp(z_i \cdot t_j / \tau)}, \quad (7)$$

and $\mathcal{L}_{t \rightarrow i}$ is defined symmetrically. The alignment loss is

$$\mathcal{L}_{\text{align}} = \frac{\mathcal{L}_{i \rightarrow t} + \mathcal{L}_{t \rightarrow i}}{2}. \quad (8)$$

For the normalised text prototype t_c of disease class c , the predicted class of image i is

$$\hat{c}_i = \arg \max_c z_i \cdot t_c. \quad (9)$$

Image hashes, source groups, predicted classes, similarities and model versions are retained as reviewable observations.

2.6 Evidence-gated dynamic updating

The updater converts candidate facts into ordered state changes. Each fact receives a stable `fact_id`, each source a `source_id`, and evidence character ranges connect candidates, documents and facts. Structural errors trigger rejection, whereas incomplete semantics or weak evidence lower the score through penalties. Accepted facts are projected to current business relations, and intermediate candidates remain in the review queue.

For a fact $f = (s, p, o)$ from independent source d , let q_s , q_o and q_p be the evidence strengths of the subject, object and predicate. The base evidence is their geometric mean:

$$q_d = (q_s q_o q_p)^{1/3}. \quad (10)$$

For time-sensitive relations, the time difference Δt between the source and the update induces exponential decay with a relation-specific constant T_p :

$$\tilde{q}_d = q_d \exp\left(-\frac{\Delta t}{T_p}\right). \quad (11)$$

Under soft gating, the structural, atomicity, grounding and local-semantic penalties of source d are combined into a gate factor $g_d \in [0, 1]$, where $g_d = 1$ means no penalty. Evidence from the set $D(f)$ of independent sources is combined by noisy-OR:

$$Q(f) = 1 - \prod_{d \in D(f)} (1 - g_d \tilde{q}_d). \quad (12)$$

If $Q(f)$ reaches the admission threshold θ , the state becomes accepted; otherwise the fact remains proposed or is rejected, and the decision reason is stored. Model scores are used as evidence strengths and are not interpreted as calibrated probabilities.

2.7 Knowledge graph materialisation and visualisation

Reviewed NER and RE outputs are converted into triples and stored in Neo4j using stable identifiers and merge semantics. Unique constraints on entity identifiers, relation keys and image SHA-256 identifiers make repeated imports idempotent. A force-directed web interface renders disease-centred local subgraphs and supports entity search, query display, node details, zooming and panning, relation provenance and image review status.

2.8 Evaluation metrics

Exact-match Precision (P), Recall (R) and F1 are used for NER and RE:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F1 = \frac{2PR}{P + R}. \quad (13)$$

Bidirectional Recall@ k is reported for image-text alignment:

$$\text{Recall@}k = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \mathbb{I}(\exists j \in \text{Top}_k(i) : \text{label}(j) = \text{label}(i)), \quad (14)$$

where Q is the query set and $\text{Top}_k(i)$ the k highest-ranked candidates of query i . Dynamic updating is evaluated by Accuracy, Precision, Recall, F1 and the false-admission rate, defined as the number of incorrectly admitted candidates divided by the number of all admitted candidates.

3 Experiments

3.1 Training settings and implementation

Training and evaluation used fixed random seeds, fixed data partitions and a locked protocol. The Chinese RoBERTa encoder used a maximum input length of 128, a batch size of 8, a learning rate of 2×10^{-5} and 18 epochs. NER was trained with the multilabel span loss in Eq. (3) over valid spans with reverse spans masked; the epoch-18 checkpoint and the zero-logit decoding boundary were used. RE reused the same tokenisation and entity-boundary mapping. Dynamic updating aggregated states by `fact_id` and independent evidence by `source_id`. Every write record retained the model version, run identifier, evidence ranges and decision reasons.

3.2 Overall performance

Table 2 summarises the locked-protocol results. Text extraction primarily generates candidates, image-text alignment is evaluated on a source-group-disjoint set, and dynamic updating uses frozen thresholds.

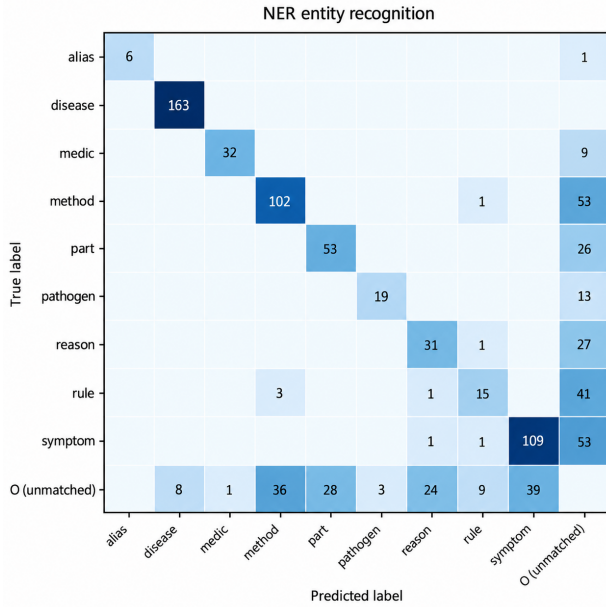
3.3 Named-entity recognition and ablation

The complete NER model achieved $F1 = 0.7326$, with $Precision = 0.7726$ and $Recall = 0.6965$ (Table 3). The complete model exceeded the F1 scores obtained after removing the BiLSTM (0.7226) or the span-length bias (0.7282), and both ablations exceeded the baseline (0.7145). Removing the BiLSTM slightly increased Recall (0.7240) but reduced Precision, so the gain of

Table 2. Overall performance of the modules.

Module	Evaluation scale	P	R	F1	Other metrics
NER	761 gold entities	0.7726	0.6965	0.7326	-
RE	277 gold relations	0.6215	0.8123	0.7042	-
Image-text alignment	227 test images, 10 classes	-	-	-	I→T R@1/R@5 = 0.8987/0.9868; T→I R@1/R@5 = 0.9500/0.9500; Acc = 0.9295; Macro-F1 = 0.9317
Dynamic updating	scorable candidates	0.9643	0.9000	0.9310	Acc = 0.8974; false-admission rate = 0.0357

the complete model comes mainly from fewer false positives. Figure 3 shows the entity-type confusion matrix; errors were concentrated in the O (unmatched) column, with limited inter-class confusion.

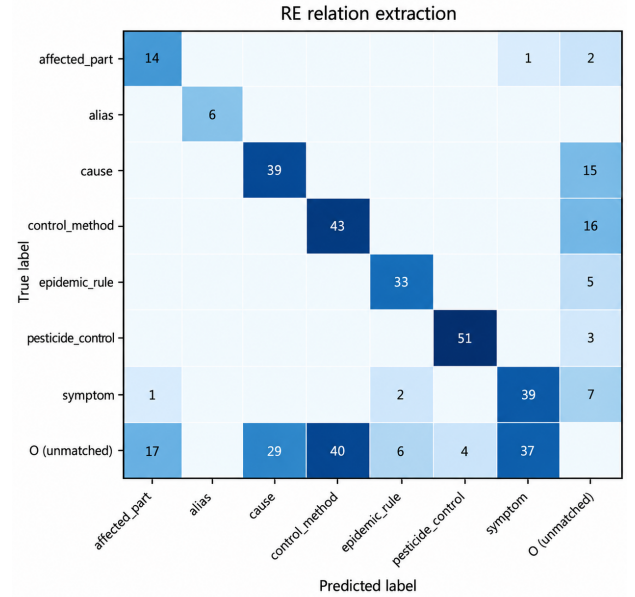
**Figure 3.** Confusion matrix of named-entity recognition by entity type.**Table 3.** NER ablation results.

Setting	P	R	F1
Baseline	0.7154	0.7135	0.7145
Complete model	0.7726	0.6965	0.7326
Without BiLSTM	0.7212	0.7240	0.7226
Without span-length bias	0.7422	0.7148	0.7282

3.4 Relation-extraction results

The relation extractor predicted 362 relations against 277 gold-standard relations, with 225 exact matches (Precision = 0.6215, Recall = 0.8123). This candidate coverage gives the evidence gate enough input for type, direction and evidence-span checks. Representative outputs are normalised as typed

subject-predicate-object triples while retaining the source spans that support each relation (Table 4). The examples cover complementary relation signatures and are not intended to enumerate the full graph. Figure 4 shows the relation confusion matrix; false positives mainly occurred in the O row.

**Figure 4.** Confusion matrix of relation extraction by relation type.

3.5 Image-text semantic alignment

On the source-group-disjoint test set, image-to-text Recall@1 and Recall@5 were 0.8987 and 0.9868, text-to-image Recall@1 and Recall@5 were both 0.9500, and prototype classification achieved Accuracy = 0.9295 and Macro-F1 = 0.9317. Figure 5 shows the ten-class confusion matrix. The main confusion was between tomato powdery mildew and tomato leaf mould: six powdery-mildew images were assigned to leaf mould.

Table 4. Representative extracted triplets.

Statement	Head entity	Relation	Tail entity
Leaf mould is commonly called black mould, is caused by <i>Passalora fulva</i> , and overwinters in plant debris.	Leaf mould	alias	black mould
Bacterial speck is caused by <i>Pseudomonas syringae</i> pv. <i>tomato</i> .	Bacterial speck	cause	<i>Pseudomonas syringae</i> pv. <i>tomato</i>
Mosaic virus disease causes mottled or necrotic fruit and stunted, slow-growing plants.	Mosaic virus disease	symptom	mottled or necrotic fruit
Powdery mildew can be controlled by fumigating with chlorothalonil smoke or by spraying flusilazole.	Powdery mildew	pesticide_control	chlorothalonil smoke

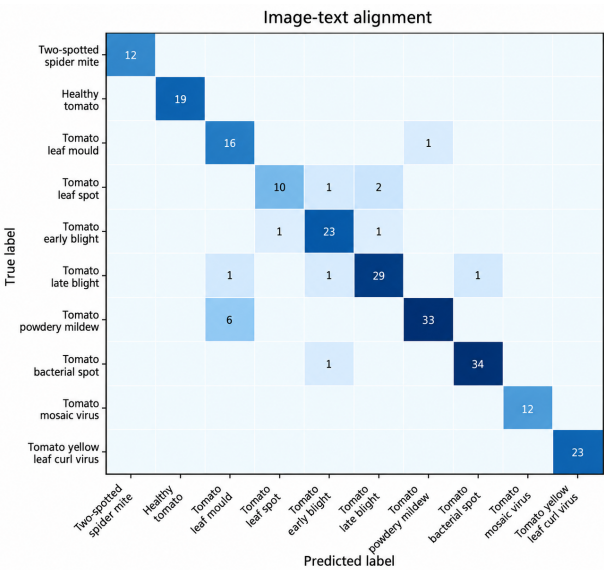


Figure 5. Confusion matrix of image-text alignment and prototype classification.

3.6 Dynamic-update comparison and ablation

Table 5 compares admission strategies on the same candidate stream. The complete evidence-gating method achieved F1 = 0.9310, Accuracy = 0.8974 and Recall = 0.9000. The false-admission rate was 0.2308 for direct writing, 0.2162 for the static threshold and 0.0357 for complete gating. The static threshold removed few wrong candidates because erroneous facts often carry high model scores; structural and evidence checks were needed to reject them. As shown in Table 6, fully hard gating reduced Recall from 0.9000 to 0.8000 and F1 from 0.9310 to 0.8727 while slightly raising the false-admission rate from 0.0357 to 0.0400.

Table 5. Comparison of dynamic-admission strategies.

Strategy	P	R	F1	Acc	FAR
Direct writing	0.7692	1.0000	0.8696	0.7692	0.2308
Static threshold	0.7838	0.9667	0.8657	0.7692	0.2162
Complete evidence gating	0.9643	0.9000	0.9310	0.8974	0.0357

FAR: false-admission rate.

Table 6. Soft versus fully hard evidence gating.

Strategy	P	R	F1	FAR
Complete model (soft gating)	0.9643	0.9000	0.9310	0.0357
Fully hard gating	0.9600	0.8000	0.8727	0.0400

3.7 Knowledge graph visualisation

Accepted facts were bulk-imported into Neo4j and rendered with the interface described in Section 2.7. Figure 6 shows a disease-centred subgraph in which each edge can be traced back to its evidence and source document.

4 Discussion

Candidate discovery and fact admission. The results support separating candidate discovery from fact admission. NER and RE provide coverage, while evidence gating checks structure, semantics and provenance. The high RE Recall (0.8123) and lower Precision (0.6215) are what one expects from a candidate generator. On the same candidate stream, complete gating raised Precision from 0.7692 to 0.9643 and reduced the false-admission rate from 0.2308 to 0.0357 relative to direct writing.

Error patterns in text extraction. NER performance varied by class: disease reached F1 = 0.9760, whereas rule reached only 0.3448 and reason 0.5345. Missed detections dominated rule errors (45 of 60 gold rule entities were missed), because epidemic rules are dispersed across sentences and have ambiguous boundaries. RE errors were dominated by false positives in control_method, symptom and cause, which are the targets of the ontology and local semantic checks.

Limitations and future work. The graph currently integrates text and leaf images but does not include disease-progress videos, field-environment

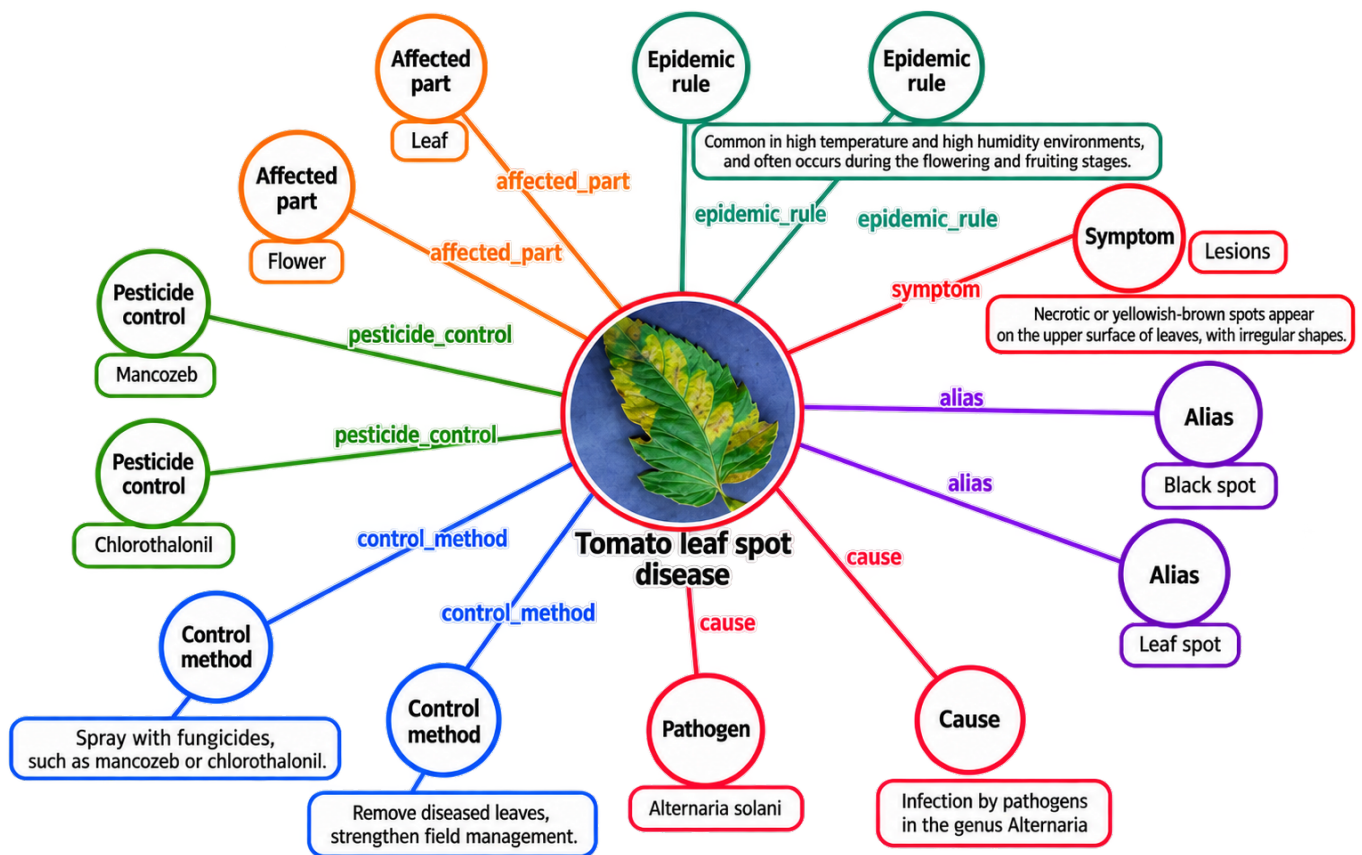


Figure 6. Disease-centred visualisation of the tomato disease knowledge graph.

monitoring or time series. Without environmental and temporal dimensions, it cannot express when a disease is likely to occur under particular temperature and humidity conditions. Source conflicts, source reliability and changes in pesticide registration still require manual review. The image set is also drawn largely from controlled-background photographs, so field performance may be lower [19]. Future work will expand the corpus and source diversity, add video and environmental data, evaluate graph-side structural changes and error persistence, and integrate the graph with retrieval-augmented generation while retaining the evidence and review status for each answer.

5 Conclusions

We built a workflow for a tomato multimodal knowledge graph that combines named-entity recognition, relation extraction, image-text alignment and evidence-gated dynamic updating. Under a fixed protocol, the extractors produced candidates with F1 scores of 0.7326 (NER) and 0.7042 (RE), image-text prototype classification reached an Accuracy of 0.9295, and evidence gating lowered the false-admission

rate to 0.0357 at an F1 of 0.9310. Keeping model predictions, evidence records and graph states separate made it possible to reject most erroneous facts and to audit every incremental update. For production use, source-reliability levels should be configurable and pesticide and treatment relations should receive stricter review. When models are upgraded, earlier candidates and decision records should be kept so that batch replay can compare state changes for the same fact_id.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the National Major Agricultural Science and Technology Project under Grant 0304.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable. This study did not involve human participants or animals; all images are photographs of plant leaves.

References

- [1] Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4), 1-37. [\[CrossRef\]](#)
- [2] Wang, K., Miao, Y., Wang, X., Li, Y., Li, F., & Song, H. (2024). Research on the construction of a knowledge graph for tomato leaf pests and diseases based on the named entity recognition model. *Frontiers in Plant Science*, 15, 1482275. [\[CrossRef\]](#)
- [3] Yan, R., An, P., Meng, X., Li, Y., Li, D., Xu, F., & Dang, D. (2025). A knowledge graph for crop diseases and pests in China. *Scientific Data*, 12(1), 222. [\[CrossRef\]](#)
- [4] Lv, B., Wu, H., Chen, W., Chen, C., Miao, Y., & Zhao, C. (2024). VEG-MMKG: Multimodal knowledge graph construction for vegetables based on pre-trained model extraction. *Computers and Electronics in Agriculture*, 226, 109398. [\[CrossRef\]](#)
- [5] Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. (2023). A comprehensive survey on automatic knowledge graph construction. *ACM Computing Surveys*, 56(4), 1-62. [\[CrossRef\]](#)
- [6] Zhao, X., Chen, B., Ji, M., Wang, X., Yan, Y., Zhang, J., ... & Lv, C. (2024). Implementation of large language models and agricultural knowledge graphs for efficient plant disease detection. *Agriculture*, 14(8), 1359. [\[CrossRef\]](#)
- [7] Lu, Y., Lu, X., Zheng, L., Sun, M., Chen, S., Chen, B., ... & Lv, C. (2024). Application of multimodal transformer model in intelligent agricultural disease detection and question-answering systems. *Plants*, 13(7), 972. [\[CrossRef\]](#)
- [8] Zhang, W., Wang, C., Wu, H., Zhao, C., Teng, G., Huang, S., & Liu, Z. (2022). Research on the Chinese named-entity-relation-extraction method for crop diseases based on BERT. *Agronomy*, 12(9), 2130. [\[CrossRef\]](#)
- [9] Liang, J., Li, D., Lin, Y., Wu, S., & Huang, Z. (2023). Named entity recognition of Chinese crop diseases and pests based on RoBERTa-wwm with adversarial training. *Agronomy*, 13(3), 941. [\[CrossRef\]](#)
- [10] Zhang, D., Zheng, G., Liu, H., Ma, X., & Xi, L. (2023). AWdpCNER: Automated Wdp Chinese named entity recognition from wheat diseases and pests text. *Agriculture*, 13(6), 1220. [\[CrossRef\]](#)
- [11] Tang, W., Wen, X., & Hu, Z. (2024). Named entity recognition for crop diseases and pests based on gated fusion unit and manhattan attention. *Agriculture*, 14(9), 1565. [\[CrossRef\]](#)
- [12] Wang, P., Zhang, C., Wang, D., Zhang, S., Wang, J., Wang, X., & Huang, L. (2023). Relation extraction for knowledge graph generation in the agriculture domain: A case study on soybean pests and disease. *Applied engineering in agriculture*, 39(2), 215-224. [\[CrossRef\]](#)
- [13] Veena, G., Gupta, D., & Kanjirangat, V. (2023). Semi-supervised bootstrapped syntax-semantics-based approach for agriculture relation extraction for knowledge graph creation and reasoning. *IEEE Access*, 11, 138375-138398. [\[CrossRef\]](#)
- [14] Shang, Y. M., Huang, H., & Mao, X. (2022, June). Onerel: Joint entity and relation extraction with one module in one step. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 36, No. 10, pp. 11285-11293). [\[CrossRef\]](#)
- [15] Liu, B., Tao, J., Chen, W., Zhang, Y., Chen, M., He, L., & Tang, D. (2024). Integration of relation filtering and multi-task learning in globalpointer for entity and relation extraction. *Applied Sciences*, 14(15), 6832. [\[CrossRef\]](#)
- [16] Cui, Y., Che, W., Liu, T., Qin, B., Wang, S., & Hu, G. (2020, November). Revisiting pre-trained models for Chinese natural language processing. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 657-668). [\[CrossRef\]](#)
- [17] Shoaib, M., Shah, B., El-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., ... & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in plant science*, 14, 1158933. [\[CrossRef\]](#)
- [18] Ngo, Q. H., Kechadi, T., & Le-Khac, N. A. (2022). Knowledge representation in digital agriculture: A step towards standardised model. *Computers and Electronics in Agriculture*, 199, 107127. [\[CrossRef\]](#)
- [19] Jelali, M. (2024). Deep learning networks-based tomato disease and pest detection: A first review of research studies using real field datasets. *Frontiers in Plant Science*, 15, 1493322. [\[CrossRef\]](#)
- [20] Xu, M., Park, J. E., Lee, J., Yang, J., & Yoon, S. (2024). Plant disease recognition datasets in the age of deep learning: challenges and opportunities. *Frontiers in Plant Science*, 15, 1452551. [\[CrossRef\]](#)
- [21] Zhang, N., Wu, H., Zhu, H., Deng, Y., & Han, X. (2022). Tomato disease classification and identification method based on multimodal fusion deep learning. *Agriculture*, 12(12), 2014. [\[CrossRef\]](#)
- [22] Yang, A., Pan, J., Lin, J., Men, R., Zhang, Y., Zhou, J., & Zhou, C. (2022). Chinese clip: Contrastive vision-language pretraining in chinese. *arXiv preprint*

- arXiv:2211.01335*. [CrossRef]
- [23] Zhou, Y., Yan, H., Ding, K., Cai, T., & Zhang, Y. (2024). Few-shot image classification of crop diseases based on vision-language models. *Sensors*, 24(18), 6109. [CrossRef]
- [24] Liu, W., Wu, G., Wang, H., & Ren, F. (2025). Cross-modal data fusion via vision-language model for crop disease recognition. *Sensors*, 25(13), 4096. [CrossRef]
- [25] Wang, Y., Wang, F., Chen, W., Lv, B., Liu, M., Kong, X., ... & Pan, Z. (2025). A large language model for multimodal identification of crop diseases and pests. *Scientific Reports*, 15(1), 21959. [CrossRef]
- [26] Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2), 494-514. [CrossRef]
- [27] Hofer, M., Obraczka, D., Saeedi, A., Köpcke, H., & Rahm, E. (2024). Construction of knowledge graphs: Current state and challenges. *Information*, 15(8), 509. [CrossRef]
- [28] Cai, B., Xiang, Y., Gao, L., Zhang, H., Li, Y., & Li, J. (2023). Temporal knowledge graph completion: A survey. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* (pp. 6545-6553). [CrossRef]
- [29] Sikos, L. F., & Philp, D. (2020). Provenance-aware knowledge representation: A survey of data models and contextualized knowledge graphs. *Data Science and Engineering*, 5(3), 293-316. [CrossRef]
- [30] Ayadi, N. Y., Bernard, S., Bossy, R., Courtin, M., Happi, B. G. H., Larmande, P., ... & Faron, C. (2024). A unified approach to publish semantic annotations of agricultural documents as knowledge graphs. *Smart Agricultural Technology*, 8, 100484. [CrossRef]
- [31] Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*. [CrossRef]
- [32] Su, J., Murtadha, A., Pan, S., Hou, J., Sun, J., Huang, W., ... & Liu, Y. (2022). Global pointer: Novel efficient span-based approach for named entity recognition. *arXiv preprint arXiv:2208.03054*. [CrossRef]
- [33] Wu, L., Li, J., Wang, Y., Meng, Q., Qin, T., Chen, W., ... & Liu, T. Y. (2021). R-drop: Regularized dropout for neural networks. *Advances in neural information processing systems*, 34, 10890-10905.
- [34] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., ... & Krishnan, D. (2020). Supervised contrastive learning. *Advances in neural information processing systems*, 33, 18661-18673.



Ziming Gao is currently pursuing the master's degree in big data technology and engineering at the College of Computer Science and Technology, Inner Mongolia Minzu University, Tongliao, China. His research interest is knowledge graphs. (Email: Nercita_GJYH_02@126.com)



Huaji Zhu is a researcher whose work focuses on agricultural big data technology and digital village technology. He is currently the Deputy Director of the Big Data and Intelligence Department, Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, the Secretary-General of the Intelligent Agriculture Committee of the Chinese Association for Artificial Intelligence, and a chief scientist of a National Key R&D Program project. (Email: zhuhj@nercita.org.cn)



Yusen Yang received the M.Sc. degree in computer engineering from the College of Design and Engineering, National University of Singapore (NUS), Singapore. He is an Engineer with the National Engineering Research Center for Information Technology in Agriculture (NERCITA) and the Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences (BAAFS). His research interests include smart agriculture, multi-agent reinforcement learning (MARL), agricultural foundation models, and personalized water and fertilizer management decision support systems. (Email: yangys@nercita.org.cn)