Domain Adaptation in Nested Named Entity Recognition From Scientific Articles in Agriculture

Doan Thai Binh Phan
University of Information Technology
Ho Chi Minh City, Vietnam
Vietnam National University
Ho Chi Minh City, Vietnam
20520043@gm.uit.edu.vn

Phuoc Vinh Linh Le
University of Information Technology
Ho Chi Minh City, Vietnam
Vietnam National University
Ho Chi Minh City, Vietnam
20521531@gm.uit.edu.vn

Ngoc Hoang Luong
University of Information Technology
Ho Chi Minh City, Vietnam
Vietnam National University
Ho Chi Minh City, Vietnam
hoangln@uit.edu.vn

Tahar Kechadi
University College Dublin
Dublin, Ireland
tahar.kechadi@ucd.ie

Hung Q. Ngo
Technological University Dublin
Dublin, Ireland
hung.ngo@tudublin.ie

ABSTRACT

In the realm of digital agriculture, the ability to make timely, profitable, and actionable decisions depends on agronomists using agricultural data and related cultivated data, including text sources such as news articles, farm notes, and agricultural scientific reports. Named entity recognition (NER) and agricultural entity recognition (AGER) facilitate semantic understanding, enabling precise identification, categorization of farming components, and knowledge discovery. However, current approaches to agricultural entity recognition encounter limitations due to limited resources. Moreover, the necessity to identify nested named entities emerges from the complexities inherent in the agricultural domain. Relevant information often traverses multiple interconnected elements rather than residing as isolated entities. For instance, comprehending a target farming practice might necessitate pinpointing the crop, the associated nutrients, or diseases—each constituting a nested entity within a broader context. Consequently, agricultural entity recognition from unstructured text gives high importance to information retrieval and knowledge construction within this domain. This study constructs the SAGRI dataset, incorporating a novel tagset for AGER that encompasses prevalent agricultural and scientific concepts, methodically established through annotation. This tagset enables the extraction of domain-independent concepts from scientific article abstracts. This study also introduces a cutting-edge deep learning baseline with an advanced Triaffine attention mechanism for robust entity extraction. Additionally, it presents a pioneering few-shot learning strategy that optimizes cross-domain categorization, mainly when dealing with scarce training data. Notably, this strategy achieves high F1 scores compared to the baseline, underscoring its potential to curtail required training data considerably.

SOICT 2023, December 07–08, 2023, Ho Chi Minh, Vietnam

ACM ISBN 979-8-4007-0891-6/23/12.
https://doi.org/10.1145/3628797.3628958

CCS CONCEPTS

• **Computing methodologies → Natural language processing**; • **Applied computing → Agriculture**.

KEYWORDS

Nested named entity recognition, digital agriculture, scientific entity, information retrieval

ACM Reference Format:
Doan Thai Binh Phan, Phuoc Vinh Linh Le, Ngoc Hoang Luong, Tahar Kechadi, and Hung Q. Ngo. 2023. Domain Adaptation in Nested Named Entity Recognition From Scientific Articles in Agriculture. In *The 12th International Symposium on Information and Communication Technology (SOICT 2023), December 07–08, 2023, Ho Chi Minh, Vietnam.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3628797.3628958

1 INTRODUCTION

Named Entity Recognition (NER) is a crucial Natural Language Processing (NLP) task that involves identifying and categorizing various entities such as individuals, organizations, locations, etc., within unstructured text. Nested Named Entity Recognition (Nested NER) takes this a step further by recognizing and classifying entities that are hierarchically nested within one another, presenting a more complex and nuanced understanding of textual information (Yuan et al., 2020 [29] and Alex et al., 2007 [1]). In recent years, Nested NER has garnered substantial interest, especially within specialized domains like agriculture ([2] and [14]), where the precise identification of hierarchical entities can provide valuable insights for information extraction and knowledge representation.

NER tasks classify entities into three distinctive types (Figure 1), each shedding light on the varying structural intricacies within different domains. The first type (S_1), flat entities, stands as the most prevalent and straightforward category [15] [8]. These entities, frequently encountered in the agricultural realm, are characterized by their uninterrupted textual spans that directly correspond to entity names. On the other hand, disjointed entities (S_2), prevalent in scientific domains, present a more complex challenge [24]. These entities are characterized by non-continuous word sequences that collectively form a complete entity, necessitating advanced recognition techniques to decipher their significance accurately. Lastly,

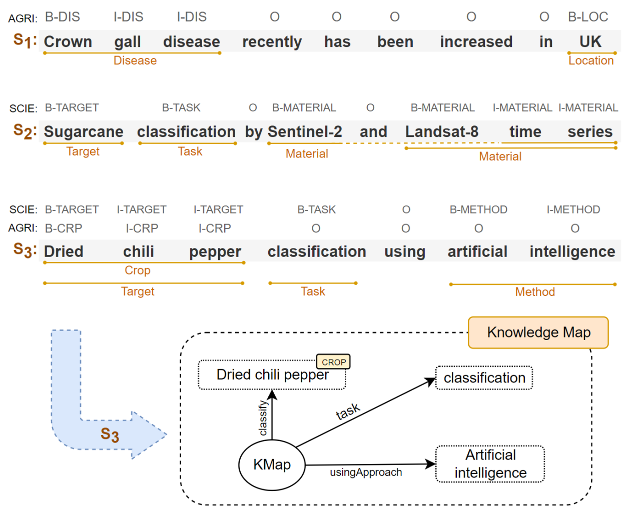

Figure 1: Examples of three types of NER subtasks.

nested entities (S_3) introduce an added layer of complexity, often found in the crossroads of agriculture and scientific literature [30]. (S_3) structure is used to extract knowledge to build knowledge maps as proposed in the OAK model [9, 13]. In this configuration, one entity encapsulates another, generating a hierarchical structure that requires nuanced processing. This triad of entity types encapsulates the multifaceted nature of NER, providing insights into the diverse linguistic patterns and contextual nuances across domains.

In addition, the scope of NER has expanded beyond traditional categories to encompass *scientific entities*, which refer to entities in the scientific domain, such as genes, proteins, and diseases within the scientific literature. Nested NER with scientific entities is a challenging task, as scientific entities can be nested within each other in complex ways. However, it is a valuable task, as it can be used to extract information from scientific literature and to build knowledge graphs of the scientific domain [13].

There are several approaches, which have been proposed to tackle the task of Nested NER and handle the challenges posed by nested structures in last few years. These approaches are mainly based on the deep learning approach, including layered-based, hypergraph-based, span-based (or region-based) and other approaches. Layered-based approaches incorporate multiple layers to capture nested entities of different levels. By utilizing hierarchical structures, these models effectively identify and classify nested NER. The Pyramid model has shown remarkable performance on NNE, ACE, and GENIA datasets [22], [23]. Hypergraph-based approaches leverage hypergraph structures to represent dependencies between tokens and entities. By modeling relationships between nested entities and tokens using graph-based techniques, these methods effectively recognize nested named entities on ACE-2004, ACE-2005 and GENIA dataset [21]. In addition, span-based approaches focus on identifying and classifying spans that contain nested named entities [27], [17]. These methods define specific spans of interest within the text and employ techniques such as rule-based matching or machine learning to detect and classify nested entities. [30] proposed a method that combines various factors and the triaffine mechanism for improved Nested NER. Their approach achieved superior performance by integrating lexical, syntactic, and semantic information with BERT embeddings and multi-head self-attention.

A pioneering few-shot learning strategy was also utilized to improve domain classification. The strategy was shown to be effective in improving domain classification accuracy, particularly in scenarios with limited training data [4]. The few-shot learning strategy works by first creating a small set of labelled instances from each domain. These labelled instances are then used to train a classifier model that can be used to classify unlabeled instances. The few-shot learning strategy was evaluated on the SAGRI dataset, which contains an imbalanced number of labelled instances. The few-shot learning strategy was shown to be effective in improving domain classification accuracy, even with a small number of labelled instances. This suggests that the few-shot learning strategy has the potential to significantly reduce the amount of training data required for Nested NER in the agriculture domain.

This paper *firstly* presents a novel dataset tailored specifically for Nested NER in the agriculture domain, serving as a valuable resource for researchers and practitioners. *Secondly*, it proposes a model that achieves competitive performance on this dataset. These contributions aim to advance the development of more effective Nested NER models customized for domain-specific applications within the agriculture community. *Thirdly*, this work presents an innovative few-shot learning approach that improves classification between domains by optimizing the categorization of instances between domains, particularly in situations with scarce training data. This approach exhibits considerable consensus between non-expert annotators after discussions with field experts. It also attains high F1 results with the baseline model and highlights the potential of training large NLP models effectively with limited resources.

Section 2 describes the dataset and the preprocessing steps applied to the data. Next, Section 3 presents the proposed model architecture and training procedure. Section 4 reports the experimental results and compares the approach with several baselines. Finally, Section 5 discusses some limitations and future directions and then concludes the paper.

2 DOMAIN SCIENTIFIC ENTITY CORPUS

This section introduces the Domain Scientific Entity Corpus, which is central to our study of Nested NER in the agriculture domain.

2.1 Entity Tagset

The CoNLL-2003 Shared Task is one of the fundamental NER tasks. This task includes training, development, and testing data frequently used to compare different NER systems. The dataset consists of six general named entity tags, including *location* (LOC), *organizations* (ORG), *person* (PER), *number* (NUM), *money* (MON), and *time* (TIM) [20]. Various shared datasets, such as GermEval 2014 NER Shared dataset, have been used to study NER in different domains [1]. Specific domains, like biomedical research, utilize tagsets tailored to their needs, such as BioNER or GenneNER [7].

In the AGER study [13], a tagset with 18 fine-grained entity tags has been defined to identify entities in agricultural documents. These tags span an array of categories including person, location,

[1]https://sites.google.com/site/germeval2014ner/home

organization, chemicals, crops, organisms, policies, climate, food, diseases, natural disasters, events, nutrients, counts, distances, quantities, money, temperature, and dates [11].

In the scientific domain, project SciNER [26] introduces an innovative and intricate approach to scientific named entity recognition that incorporates a multi-dimensional tagset. This tagset goes beyond the conventional boundaries of NER, encompassing six key dimensions: *Task*, *Method*, *Target*, *Dataset*, *Material*, and *Metric*. However, SciNER only focuses on six scientific entity types.

Inspired by the landscape of scientific entity recognition, this study presents a novel two-layered tagset customized for the field of agricultural sciences (as illustrated in Table 1). The first layer amalgamates general entity tags with agricultural-specific ones, while the second layer delves deeper into the scientific domain by employing six distinct entity tags.

However, this study introduces a refined perspective by narrowing its gaze to entities exclusively rooted in the agricultural domain. The tagset is structured around a pragmatic set of 18 distinct tags, artfully outlined in Table 1. This endeavor, inspired by the amalgamation of AGER's rich tagset and the intricacies of scientific named entity recognition, paves the way for an evolved understanding of agricultural entities within the scientific context.

Table 1: Details of SAGRI Entity Tagset and Corpus

Tag	Name - Description	Entities
LOC	Location names or addresses	4,603
ORG	Organization names, such as Microsoft	6,918
PER	Person names, such as Micheal, Peter	2,834
AMT	Amount of some things	21,578
TIME	Date, time, season	4,802
ANI	Animal - Name of animals	1,150
CRP	Crop - Fruits, vegetables, cereals, grains	11,455
FAM	Farm - Area of land, for growing crops	10,197
DIS	Disease - Affecting crop/livestock	5,113
MIO	Microorganism - Name of micro-organism	1,234
FOD	Food - Plant/animal products	3,099
NUT	Nutrients - Fats, minerals, vitamins	2,673
TEM	Temperature	1,154
CLI	Climate - Denotes the climatic conditions	743
CHE	Chemical - An agrochemical or chemical	3,196
FER	Fertiliser - Bio/chemical fertiliser	384
DST	Disasters affecting crop production	927
AGRI	Other concerned agriculture entities	5,537
TASK	Data mining task	6,769
METHOD	Algorithms using in knowledge	8,671
TARGET	Predicted labels of knowledge	1,953
DATASET	Dataset for training models	4,158
MATERIAL	Conditions of knowledge	813
METRIC	Evaluation metrics	4,084
Total		**114,045**

2.2 Corpus Characteristics

The resource of mined knowledge in digital agriculture is scientific papers published on relevant studies in agriculture. Raw resources are extracted from scientific papers published in two journals in the digital agriculture domain, including Computers and Electronics in Agriculture[2] (Elsevier) and Precision Agriculture[3] (Springer). The total number of articles is 3,381 articles and is filtered to select about 1,000 papers, which present computing results related to crops [12]. In this study, an annotated corpus called SAGRI was constructed from these scientific articles for the Nested NER task. From crawled articles, 4,200 abstracts were selected for annotation to build an annotated SAGRI dataset, incorporating scientific and agricultural entities.

Table 1 shows characteristics of the resulting corpus. The corpus has a total of 114,045 scientific entities, ranging across various domains critical to agriculture and science. Domain-specific entities are well-represented, including *Crop* (11,455 entities), *Farm* (10,197 entities), *Disease* (5,113 entities), and *Chemical* (3,196 entities). From the results of data mining tasks, the corpus contains substantial annotations for key elements like *Task* (6,769 entities), *Method* (8,671 entities), *Target* (1,953 entities), *Dataset* (4,158 entities), *Material* (813 entities), and *Metric* (4,084 entities). To sum up, the SAGRI corpus demonstrates that the corpus provides broad coverage of agriculture-related entities, with ample annotations to support common text mining objectives. The diversity and volume of entity types will facilitate training machine learning models on this dataset across a variety of agricultural domains and applications.

3 TRIAFFINE ATTENTION MODEL FOR NESTED NAMED ENTITY RECOGNITION

In this paper, Nested NER is decomposed into two subtasks: *span representation* and *span classification*. The "span representation" subtask focuses on precisely identifying and delineating the boundaries of nested entity spans, which is crucial for capturing their complex nesting structures accurately. On the other hand, the "span classification" subtask assigns specific entity labels to the extracted spans, considering potential overlaps and nested relationships between entities. By separating these subtasks, the model can systematically process and understand the intricate nature of nested entities, leading to improved accuracy and robustness in NER tasks.

3.1 Deep Triaffine Attention

The deep triaffine transformation is defined by employing vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}$, along with a tensor $W \in \mathbf{R}^d \times \mathbf{R}^{d+1} \times \mathbf{R}^d$. This transformation yields a scalar by applying distinct Multi-Layer Perceptron (MLP) operations to the input vectors and performing tensor-vector multiplications. To maintain the biaffine transformation's characteristics, a constant value of 1 is appended to the input elements.

$$u' = \begin{bmatrix} MLP_a(u) \\ 1 \end{bmatrix} \tag{1}$$

$$w' = MLP_b(w) \tag{2}$$

$$TriAff(u, v, w, W) = W \times_1 u' \times_2 w' \times_3 v' \tag{3}$$

where $\times_n$ is the mode-n tensor vector multiplication and MLP_t is a *t-layer MLP* (0-layer MLP is equal to identify function). The tensor W is initialized using $N(0, \sigma^2)$. In our approach, we use boundary

[2]https://www.journals.elsevier.com/computers-and-electronics-in-agriculture
[3]https://www.springer.com/journal/11119

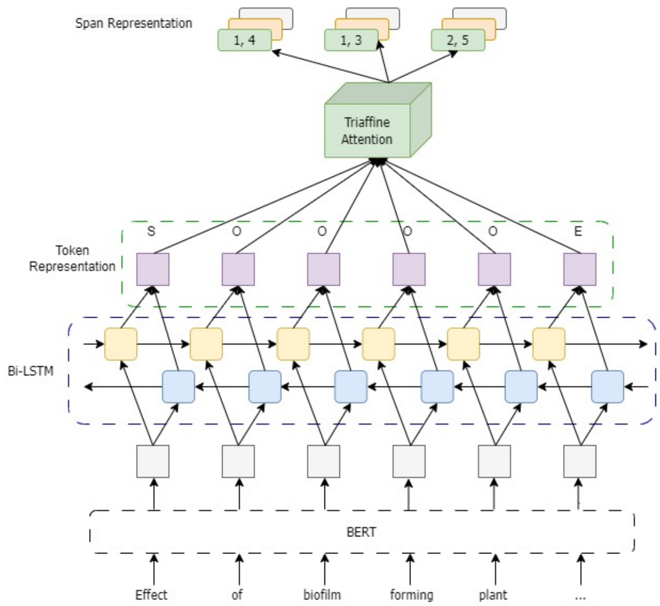

Figure 2: Span Representation Subtask.

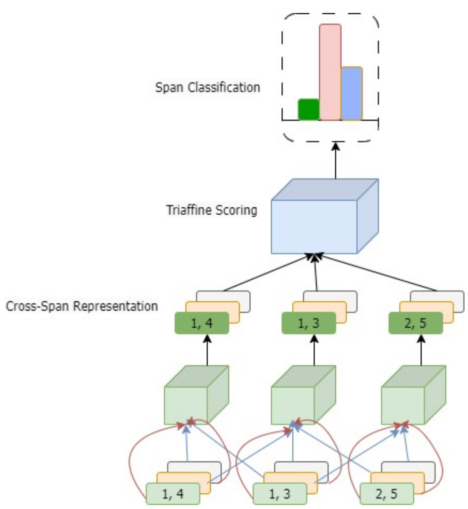

Figure 3: Span Classification Subtask.

representations as u and v. Inside tokens or span representations are used as w. We denote the tensors in the triaffine attention as W_r and triaffine scoring as V_r, which decouples attention weights and scores for different labels.

3.2 Text Encoding

To encode the text, this study uses BERT [3] and represents input text by concatenating its word embedding, contextualized word embedding, and part-of-speech (POS) embedding based on Ju et al. [6]; Shen et al. [16], and Tan et al. [19]. Using the pre-trained language model, this model first creates the contextual embedding $X = [x_1, x_2, ..., x_N]$ for text x_i^c with N tokens.

$$x_1^c, x_2^c, ..., x_N^c = PLM(x_1, x_2, ..., x_N) \tag{4}$$

To obtain the token representations h_i they are concatenated x_i^c with word embedding x_i^w, part-of-speech embedding x_i^p, character embedding x_i^{ch}. The concatenated embedding X is fed into a BiLSTM.

3.3 Triaffine Attention for Span Representations

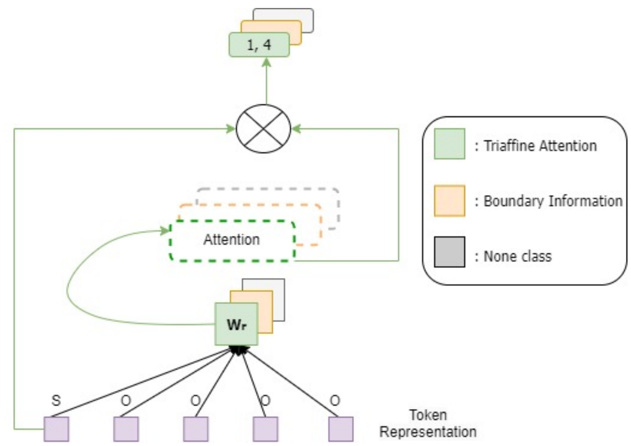

Figure 4: Triaffine Attention.

This study uses the Triaffine mechanism [30], which is depicted in Figure 4, to fuse heterogeneous elements for a better representation of the span. It learns the label-wise span representation $H_{i,j,r}$ with the Triaffine attention $\alpha_{i,j,k,r}$ for the span to interact tokens with labels and boundaries (i, j).

$$s_{i,j,k,r} = \text{TriAff}(h_i, h_j, h_k, W_r) \tag{5}$$

$$\alpha_{i,j,k,r} = \frac{exp(s_{i,j,k,r})}{\sum_{k'=i}^{j} exp(s_{i,j,k',r})} \tag{6}$$

$$\mathbf{h}_{i,j,r} = \sum_{k=i}^{j} \alpha_{i,j,k,r} MLP(\mathbf{h}_k) \tag{7}$$

Boundary representations $(\mathbf{h}_i, \mathbf{h}_j)$ and the labelwise parameters (W_r) can be viewed as attention queries, and tokens (h_k) can be viewed as keys and values. Compared with the general attention framework (additive or multiplicative attention), this Triaffine attention permits all high-order interactions between heterogeneous queries and keys.

3.4 Triaffine Attention for Cross-span Representations

Informed by how interactions between spans manifest in a nested context, the proposed approach integrates information from related spans to establish cross-span representations. Specifically, attention queries are formulated based on the boundaries of a span and its corresponding label, while attention keys and values encompass related spans, including the original span. By applying a tri-affine attention mechanism, it is able to obtain label-specific cross-span representations $\beta_{i,j,g,r}$, denoted as $h^c_{i,j,r}$, for a given span (i, j), as outlined in Equation 7

$$q_{i,j,g,r} = \text{TriAff}(\mathbf{h}_i, \mathbf{h}_j, \mathbf{h}_{i_g,i_g,r}, W_r) \tag{8}$$

$$\beta_{i,j,g,r} = \frac{exp(q_{i,j,g,r})}{\sum_{g'} exp(q_{i,j,g',r})} \tag{9}$$

$$\mathbf{h}^c_{i,j,r} = \sum_{g} \beta_{i,j,g,r} MLP(\mathbf{h}_{i_g,i_g,r}) \tag{10}$$

where (i_g, i_g) are the related spans.

3.5 Triaffine Scoring for Span Classification

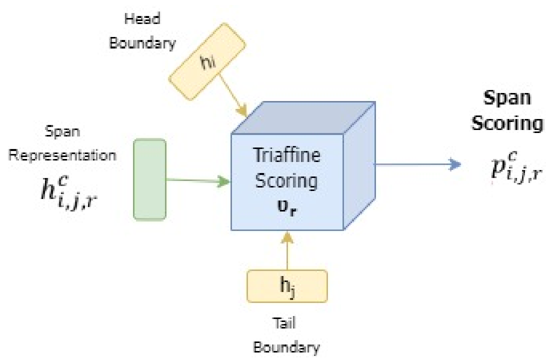

Figure 5: Triaffine Scoring.

In order to determine the entity type of a given span, the model utilizes label-wise scores based on cross-span representations. This approach incorporates boundary information, which is effective in previous studies [5], [28]. Triaffine scoring is employed to classify spans using both boundary information and cross-span representations. This involves estimating the log probabilities $p^c_{i,j,r}$ of a given span (i, j) for label r using boundaries $\mathbf{h}_i$, $\mathbf{h}_j$ and cross-span representations $h^c_{i,j,r}$.

$$p^c_{i,j,r} = TriAff(\mathbf{h}_i, \mathbf{h}_j, \mathbf{h}^c_{i,j,r}, V_r) \tag{11}$$

Since $h^c_{i,j,r}$ are composed by $\mathbf{h}_{i_g,i_g,r}$, it can decompose Equation 11 into following if and only if the layer of MLP transformation on $h^c_{i,j,r}$ is 0:

$$t_{i,j,g,r} = TriAff(\mathbf{h}_i, \mathbf{h}_j, \mathbf{h}_{i_g,i_g,r}, V_r) \tag{12}$$

$$\mathbf{h}^c_{i,j,r} = \sum_{g} \beta_{i,j,g,r} t_{i,j,g,r} \tag{13}$$

Figure 5 shows the mechanism of triaffine scoring. The model also applies similar decomposition functions in the auxiliary span classification task, which applies the triaffine scoring on boundary representations and intermediate span representations $\mathbf{h}_{i,j,r}$ to estimate log probabilities $p_{i,j,r}$ as intermediate predictions.

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Training Strategy

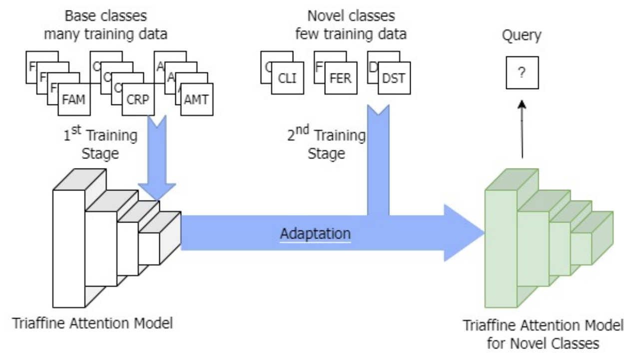

Figure 6: Training strategy.

The Triaffine attention model has a consistent performance in the evaluated datasets, and it can be observed from Table 4. However, several challenges emerged due to factors such as limited data, dataset imbalance, and inconsistent annotations within the SAGRI dataset, resulting in suboptimal outcomes when employing the Triaffine attention model. In addition, the scarcity of resources posed a significant issue in NER, Nested NER, and NLP tasks.

To overcome these challenges, the study incorporated few-shot learning as a technique to address the imbalanced data in the SAGRI dataset. Few-shot learning involves training a model with a limited number of examples to enhance its generalization capability to unseen instances. By leveraging this approach, the aim was to improve the model's performance in recognizing and classifying entities with limited instances, thereby mitigating the impact of data imbalance on overall performance.

It is essential to clarify the specifics of the few-shot learning within your method. In this context, few-shot learning refers to training the model with a limited number of labeled examples in the context of Nested NER. This technique is designed to enable the model to generalize effectively when dealing with entities that have only a few instances in the dataset. By incorporating few-shot learning, the model becomes more adept at recognizing and classifying such entities, ultimately improving its overall performance and reducing the sensitivity to data imbalance.

In addition, it is noteworthy to emphasize the resource requirements of the Triaffine attention model, as it necessitates a minimum of 16GB VRAM GPU for inference. This poses a challenge, particularly for smaller research groups. The utilization of few-shot learning was deemed essential to address the issues associated with data imbalance in the SAGRI dataset and the resource constraints posed by the Triaffine attention model.

As a result, the study employed few-shot learning as a technique to address the imbalanced data within the SAGRI dataset, as illustrated in Fig 6. This approach involved training the model with a limited number of samples to facilitate generalization to new instances. By incorporating few-shot learning, the model's capability to recognize and classify entities with only a few instances in the dataset was substantially enhanced, thus mitigating the impact of data imbalance on model performance. Moreover, this technique optimized the training process and reduced the computational resources required. In essence, the use of few-shot learning proved to be a valuable strategy for addressing data imbalance in the SAGRI dataset.

During the training procedure, a two-stage training process was employed utilizing the SAGRI dataset, which consisted of both inter-domain and intra-domain components, as proposed previously. Initially, entities with a count exceeding *5,000* were selected as base classes in the first stage. Subsequently, 16-way 5-shot subtasks were randomly sampled from the SAGRI dataset for training. Among the sampled subtasks, 25,301 were assigned to the training set, while 8,479 were assigned to the validation set. The validation set was used to assess the performance of the framework and fine-tune it during the training process.

4.2 Results

The evaluation of the trained model revealed significant improvements in its ability to recognize and classify new classes. The evaluation results of the Triaffine model, as shown in Table 2, highlight promising improvements in its ability to recognize and classify labels within the SAGRI dataset. While addressing data imbalance with few-shot learning, the model's performance on new classes exceeded expectations. The incorporation of few-shot learning has notably enhanced the model's performance on new classes, as demonstrated by the higher F1 scores in the "Triaffine + Fewshot" column compared to "Triaffine" across various labels. For Layer 1 labels, we observed that the "Triaffine + Fewshot" variant consistently achieved competitive or slightly improved F1 scores compared to the base "Triaffine" model. For instance, in new classes like MIO, NUT, FER and DST, the few-shot enhanced model outperformed its counterpart. In Layer 2, while there was a slight decrease in performance for METHOD the "Triaffine + Fewshot" model maintained or slightly improved F1 scores for other labels like TARGET, MATERIAL and METRIC. It demonstrated enhanced accuracy in identifying and classifying instances from the new classes, marking a notable advancement.

Table 3 provides a comprehensive evaluation of the Triaffine model on two distinct datasets, SAGRI and GENIA. In the SAGRI dataset, the performance of the model is segmented by layers (Layer 1 and Layer 2) and the combination of both (Layer 1 & 2). This segmentation allows for a detailed analysis of the capabilities of the model across different aspects of the data. The F1 scores, which balance precision and recall, indicate the model's effectiveness in identifying specific labels within each dataset and layer. Furthermore, the inclusion of the GENIA dataset allows for a cross-dataset comparison, showcasing the model's versatility in handling diverse text data.

A comparative analysis of various models on the SAGRI dataset is shown in Table 4. The results show that the performance of the Triaffine model and its enhanced variant, Triaffine + Fewshot, underscoring their effectiveness in comparison to other state-of-the-art models. The precision-recall trade-off is evident here, as the Triaffine + Fewshot model achieves a higher F1 score for recognizing new classes, but with a slight trade-off in performance on existing classes, a crucial aspect to consider in model evaluation.

Table 2: Evaluation (F1 score) of Triaffine on each label of SAGRI

Layer	Label	Triaffine	Triaffine + Fewshot
Layer 1	LOC	77.15	76.47
	ORG	64.97	63.23
	PER	63.76	64.41
	AMT	90.72	90.82
	TIME	89.85	90.13
	ANI	84.31	84.62
	CRP	87.76	88.76
	FAM	94.16	93.88
	DIS	87.05	87.29
	MIO	83.77	86.17
	FOD	86.18	86.97
	NUT	82.25	83.96
	TEM	84.87	84.42
	CLI	93.72	94.46
	CHE	82.1	82.74
	FER	82.71	84.06
	DST	82.18	85.29
	AGRI	89.50	89.59
Layer 2	METHOD	83.53	82.48
	TARGET	52.2	55.47
	DATASET	93.38	92.99
	MATERIAL	74.46	75.52
	TASK	89.86	89.50
	METRIC	90.87	92.59

Table 3: Evaluation of Triaffine on SAGRI and GENIA

Dataset		Precision	Recall	F1 score
SAGRI	Layer 1	84.66	81.66	82.91
	Layer 2	82.83	81.75	81.03
	Layer 1 & 2	83.89	81.33	82.08
GENIA		80.42	82.06	81.23

4.3 Error Analytics

The evaluation of the trained model unveiled significant strides in its capacity to identify and categorize novel classes (Table 2). However, this notable progress came at the expense of a noticeable decline in the model's proficiency in recognizing the pre-existing classes. One potential contributing factor to this issue is the two-stage training strategy, where the model undergoes stage two training, primarily utilizing new-class samples. The diminished performance observed in certain base classes can be elucidated by the inherent trade-off when striving to excel in recognizing new classes.

To tackle this challenge, a reevaluation of the training strategy is recommended. This involves incorporating a broader range of examples from the new classes during the initial training stages to

Table 4: Comparison of different models on SAGRI dataset

Model + Encoder	SAGRI dataset		
	Precision	Recall	F1 score
Other methods			
TreeCRF [5]	78.45	77.56	78.00
Pyramid [22]	75.91	73.33	74.60
Pyramid [22] + ALBERT	81.16	75.72	78.35
BiFlag [10] + LSTM	71.92	70.24	71.07
Seq2seq [18]	79.48	75.62	77.50
BartNER [25] + BART	76.30	79.49	77.86
Sequence to Set [19]	83.47	79.76	81.57
Locate and Label [16]	80.05	77.28	78.64
Triaffine methods			
Triaffine	82.62	**81.66**	81.53
Triaffine + Fewshot	**83.89**	81.33	**82.08**

enhance generalization. Additionally, refining the sampling strategy to ensure a more equitable representation of both base and new classes would mitigate bias and improve recognition accuracy.

The lower performance observed in layer 2 as shown in Table 3 can be explained by the inherent difficulty in recognizing abstract concepts associated with the corresponding labels. Layer 2 labels encompass tasks in data mining, algorithms, predicted labels, datasets, conditions, and evaluation metrics, often involving complex and abstract ideas. The higher level of abstraction presents challenges for the model in accurately recognizing and distinguishing between different layer 2 labels.

Table 5: Comparison of top 10 incorrectly predicted labels

True label	Count	Triaffine		Triaffine + Fewshot	
		To label	Rate(%)	To label	Rate(%)
I-PER	49	I-LOC	16.33	I-ORG	10.2
B-PER	291	B-LOC	8.94	I-DIS	8.65
I-AGRI	544	B-AGRI	4.04	B-AGRI	3.51
B-TASK	723	I-METHOD	3.73	B-TASK	3.6
I-CHE	192	B-CHE	3.65	I-AMT	4.64
B-DATASET	816	I-DATASET	3.31	I-DATASET	3.43
B-FER	62	B-FER	3.23	B-FER	1.61
B-LOC	937	B-FER	3.09	B-ORG	3.93
I-NUT	429	B-NUT	3.03	B-NUT	2.33
I-DST	206	B-DST	2.91	B-DST	2.90

Table 5 provides a comparison of the top 10 labels incorrectly predicted by Triaffine and Triaffine + Fewshot models. Common errors are observed in predicting the "I-PER" and "B-PER" labels, which pertain to personal names or entities associated with individuals.

The challenges associated with personal names' ambiguity and variability often pose difficulties in NLP tasks. The models tend to confuse "I-PER" labels with "I-LOC"/"I-ORG" labels, where "I-LOC"/"I-ORG" represents locational/ organizational entities. This confusion can be attributed to similarities in context and syntactic patterns between personal and locational/organizational names, resulting in higher misclassification rates for these label categories in both models.

Another observed error involves the misclassification of "B-PER" labels as "I-DIS" labels by the Triaffine + Fewshot model. The "I-DIS" label represents disease entities. This misclassification can occur due to the presence of disease-related terms or mentions in the context, leading to incorrect predictions.

It is important to acknowledge that these errors highlight the challenges in disambiguating and correctly classifying certain label categories, particularly those involving personal names. Further improvements in the models' training data, contextual understanding, and entity disambiguation techniques could potentially enhance the accuracy of predictions for these categories.

5 CONCLUSION AND FUTURE WORK

In conclusion, the study sheds light on both the achievements and the limitations inherent in the current model and the training strategy. While significant progress has been made in the model's ability to identify new classes, it comes at a noticeable cost: a reduction in its effectiveness in recognizing pre-existing classes. This trade-off can, in part, be ascribed to the two-stage training strategy, which heavily prioritizes new-class samples. Moreover, acknowledging the presence of abstract concepts within Layer 2 only adds to the intricacy, underscoring the imperative for targeted approaches to enhance recognition in this layer.

To overcome these limitations, future research endeavours should concentrate on incorporating a more diverse range of classes during training and fine-tuning the sampling strategy. Expanding the validation set and exploring advanced techniques such as transfer learning and cutting-edge neural network architectures can further bolster the model's ability to handle imbalanced data while maintaining high performance on existing classes.

Looking ahead, it is imperative for future work to address the identified limitations comprehensively. This can be accomplished through the acquisition of larger and more diverse datasets. Furthermore, the exploration of advanced techniques, including transfer learning and sophisticated neural network architectures, can augment the model's capacity to generalize effectively to new classes. A detailed examination of misclassifications and error patterns offers valuable insights for model refinement. Finally, considering ensemble learning, active learning, and other state-of-the-art methodologies can further elevate the model's proficiency in managing imbalanced data and accurately recognizing both existing and new classes.

REFERENCES

[1] Beatrice Alex, Barry Haddow, and Claire Grover. 2007. Recognising Nested Named Entities in Biomedical Text. In *Biological, translational, and clinical language processing*. Association for Computational Linguistics, Prague, Czech Republic, 65–72.

[2] Payal Biswas, Aditi Sharan, and Sharad Verma. 2016. Named entity recognition for agriculture domain using word net. *Int J Comput Math Sci* 5, 10 (2016), 29–36.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*. 4171–4186.

[4] Alexander Fritzler, Varvara Logacheva, and Maksim Kretov. 2019. Few-shot classification in named entity recognition task. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. 993–1000.

[5] Yao Fu, Chuanqi Tan, Mosha Chen, Songfang Huang, and Fei Huang. 2021. Nested Named Entity Recognition with Partially-Observed TreeCRFs. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 14 (2021), 12839–12847.

[6] Meizhi Ju, Makoto Miwa, and Sophia Ananiadou. 2018. A neural layered model for nested named entity recognition. In *Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Volume 1*. 1446–1459.

[7] Jin-Dong Kim, Tomoko Ohta, Yoshimasa Tsuruoka, Yuka Tateisi, and Nigel Collier. 2004. Introduction to the bio-entity recognition task at JNLPBA. In *Proceedings of the international joint workshop on natural language processing in biomedicine and its applications*. Citeseer, 70–75.

[8] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural Architectures for Named Entity Recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, 260–270.

[9] Nhien-An Le-Khac, Mohand Tahar Kechadi, and Joe Carthy. 2006. Admire framework: Distributed data mining on data grid platforms. In *Conference on Software and Data Technologies (ICSOFT'06), Setubal, Portugal*.

[10] Ying Luo and Hai Zhao. 2020. Bipartite Flat-Graph Network for Nested Named Entity Recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 6408–6418.

[11] CS Malarkodi, Elisabeth Lex, and Sobha Lalitha Devi. 2016. Named Entity Recognition for the Agricultural Domain. *Res. Comput. Sci.* 117, 1 (2016), 121–132.

[12] Quoc Hung Ngo, Tahar Kechadi, and Nhien-An Le-Khac. 2021. Domain specific entity recognition with semantic-based deep learning approach. *IEEE Access* 9 (2021), 152892–152902.

[13] Quoc Hung Ngo, Tahar Kechadi, and Nhien-An Le-Khac. 2022. Knowledge representation in digital agriculture: A step towards standardised model. *Computers and Electronics in Agriculture* 199 (2022), 107127.

[14] Aditi Sharan Payal Biswas. 2016. Named Entity Recognition for Agriculture Domain using WordNet. *International Journal of Computer & Mathematical Sciences* Volume 5, Issue 10 (2016), 29–36.

[15] Massinissa Saoudi, Ahcène Bounceur, Reinhardt Euler, and Tahar Kechadi. 2016. Data Mining Techniques Applied to Wireless Sensor Networks for Early Forest Fire Detection. In *Proceedings of the International Conference on Internet of Things and Cloud Computing*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2896387.2900323

[16] Yongliang Shen, Xinyin Ma, Zeqi Tan, Shuai Zhang, Wen Wang, and Weiming Lu. 2021. Locate and Label: A Two-stage Identifier for Nested Named Entity Recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1)*. 2782–2794.

[17] Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. DiffusionNER: Boundary Diffusion for Named Entity Recognition. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1)*. Association for Computational Linguistics, Toronto, Canada, 3875–3890.

[18] Jana Straková, Milan Straka, and Jan Hajic. 2019. Neural Architectures for Nested NER through Linearization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 5326–5331.

[19] Zeqi Tan, Yongliang Shen, Shuai Zhang, Weiming Lu, and Yueting Zhuang. 2021. A Sequence-to-Set Network for Nested Named Entity Recognition. In *Proceedings of International Joint Conferences on Artificial Intelligence*.

[20] Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of CoNLL-2003, Edmonton, Canada*. Association for Computational Linguistics, 142–147.

[21] Bailin Wang and Wei Lu. 2018. Neural Segmental Hypergraphs for Overlapping Mention Recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 204–214.

[22] Jue Wang, Lidan Shou, Ke Chen, and Gang Chen. 2020. Pyramid: A layered model for nested named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 5918–5928.

[23] Yiran Wang, Hiroyuki Shindo, Yuji Matsumoto, and Taro Watanabe. 2022. Nested named entity recognition via explicitly excluding the influence of the best path. *Journal of Natural Language Processing* 29, 1 (2022), 23–52.

[24] Yucheng Wang, Bowen Yu, Hongsong Zhu, Tingwen Liu, Nan Yu, and Limin Sun. 2021. Discontinuous Named Entity Recognition as Maximal Clique Discovery. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1)*. Association for Computational Linguistics, 764–774.

[25] Hang Yan, Tao Gui, Junqi Dai, Qipeng Guo, Zheng Zhang, and Xipeng Qiu. 2021. A Unified Generative Framework for Various NER Subtasks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1)*. Association for Computational Linguistics, 5808–5822.

[26] Tan Yan, Heyan Huang, and Xian-Ling Mao. 2020. SciNER: A Novel Scientific Named Entity Recognizing Framework. In *Natural Language Processing and Chinese Computing*. Springer International Publishing, 828–839.

[27] Pan Yang, Xin Cong, Zhenyu Sun, and Xingwu Liu. 2021. Enhanced Language Representation with Label Knowledge for Span Extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 4623–4635.

[28] Juntao Yu, Bernd Bohnet, and Massimo Poesio. 2020. Named Entity Recognition as Dependency Parsing. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 6470–6476.

[29] Zheng Yuan, Yuanhao Liu, Qiuyang Yin, Boyao Li, Xiaobin Feng, Guoming Zhang, and Sheng Yu. 2020. Unsupervised multi-granular Chinese word segmentation and term discovery via graph partition. *Journal of Biomedical Informatics* 110 (2020), 103542.

[30] Zheng Yuan, Chuanqi Tan, Songfang Huang, and Fei Huang. 2021. Fusing Heterogeneous Factors with Triaffine Mechanism for Nested Named Entity Recognition. *CoRR* abs/2110.07480 (2021).