IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 June 2024, accepted 13 August 2024, date of publication 16 August 2024, date of current version 26 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3444772

RESEARCH ARTICLE

ANeTCM: A Novel MRC Framework for Traditional Chinese Medicine Named Entity Recognition

YUANYU FENG[1] **AND YAN ZHOU**[2]
[1]The Affiliated Hospital of Guizhou Medical University, Guiyang 550004, China
[2]Department of Children's Health Care, Anhui Provincial Children's Hospital/Children's Hospital of Fudan University Anhui Hospital, Hefei 230022, China
Corresponding author: Yan Zhou (15256215380@163.com)

ABSTRACT Traditional Chinese medicine (TCM) named entity recognition for supporting downstream tasks is receiving increasing attention. However, mainstream named entity recognition models applied to the TCM domain are still affected by the following two challenges: lack of domain knowledge and imbalance between entity classes. Therefore, we propose ANeTCM, a model that enhances both domain knowledge and inter-entity balance. Specifically, we first use a large number of TCM medical case data to continuously pretrain Roberta and enhance its domain knowledge. Secondly, the sequence annotation is converted into a machine reading comprehension task, and gated linear units are incorporated to further enhance the model's feature learning capability. Finally, the weights of the samples are adjusted using a normal distribution to address the imbalance of entity classes. We conducted extensive experiments on two TCM named entity recognition datasets and selected several competitive models. The experimental results show the effectiveness of our model.

INDEX TERMS Traditional Chinese medicine, named entity recognition, machine reading comprehension, gated linear units, normal distribution.

I. INTRODUCTION

Traditional Chinese medicine (TCM) represents thousands of years of Chinese medical tradition, and it is of great significance to protect and preserve this precious cultural heritage [1]. By better organizing and recording information on herbs, remedies, and names of diseases in traditional Chinese medicine, it helps to preserve and pass on traditional knowledge. Chinese medicine still has a wide range of applications in contemporary medical and pharmaceutical research [2]. Combining traditional Chinese medicine with modern medicine promotes the modernization of Chinese medicine and improves therapeutic efficacy. While herbs and medicinal plants in traditional Chinese medicine may contain potential new drugs, the identification and collation of these plant or animal resources can greatly expedite the research and development of new drugs.

The associate editor coordinating the review of this manuscript and approving it for publication was Ugur Guvenc.

Named entity recognition (NER), an important fundamental task in natural language processing (NLP), is applied to many unstructured texts [3]. In recent years, NER tasks have achieved good performance based on the pre-trained model BERT [4], [5]. Applying NER technology in TCM is helpful for supporting downstream tasks, such as TCM knowledge graph construction [6] and Q&A [7]. This can effectively promote the development of TCM in modern times. However, traditional Chinese medicine named entity recognition (TCM-NER) [8], as a special field in NER, is characterised by its specialisation, leading to the fact that generic NER methods cannot be well adapted to the TCM-NER task. Continuous pre-training is widely used in the professional field to enhance domain knowledge [9]. The era of large model based also helps in acquiring domain knowledge and learning [10].

However, there is currently no specialized large model in the TCM-NER domain, and training large models heavily relies on hashrate resources. At the same time, the imbalance in the category of TCM entities can worsen the decline in recognition performance. As a result, we faced the following limitations when conducting the TCM-NER task: (1) The model's lack of knowledge about the domain leads to an inability to efficiently identify specialized entity classes. (2) The problem of imbalance in the TCM-NER task's entity classes exacerbates the difficulty of identifying sparse entity classes.

To address the above issues, we propose ANeTCM, a model that improves domain knowledge and achieves a balance among entity classes. We obtained TCM-BERT by continuously pre-training Roberta [11] using a large amount of medical case data,[1] thereby acquiring a substantial amount of domain knowledge. The model feature learning capability is then further enhanced by machine reading comprehension [12] and gated linear units [13]. Finally, the loss function is improved by drawing inspiration from the normal distribution function in order to enhance the ability of recognizing sparse entity categories.

Experiments show that ANeTCM is effective for TCM-NER tasks. We conducted comparative experiments on the two datasets provided by the "Chinese Medicine Instruction Manual Entity Recognition Challenge"[2] and GitHub.[3] Our approach achieves a significant performance improvement compared to several competitive models. Specifically, we make the following contributions:

- We introduce the ANeTCM, A model for enhancing TCM-NER domain knowledge and balancing entity classes. This approach improves TCM entity class recognition performance in TCM-NER tasks.
- The experiments showed that ANeTCM was more effective than generic NER model in the TCM-NER task.
- We model outperforms other pre-trained models and loss functions in the TCM-NER task.

II. RELATED WORK

Traditional NER methods are classified into probabilistic statistical models [14] and deep learning models [15], [16]. In recent years, deep learning-based models have made significant advancements. Converting the NER task to an MRC task enhances the feature learning capability of the model by introducing external knowledge [12], [17]. As the field of NER has become more specialized, generic approaches no longer fulfill the requirements of specific domains. Zhang et al. [18] proposed a distant supervision via back-labeling approach for the TCM-NER task. Xu et al. [19] proposed a character-level based approach for recognizing named entities in Chinese medicine. Ma et al. [20] proposed a multi-granularity text-driven NER model based on Conditional Generative Adversarial Network (MT-CGAN) to implement TCM-NER using a small annotated corpus. Ma et al. [21] proposed a TCM-NER model with lexicon information and text local feature enhancement of text. We employ the MRC architecture as the backbone network for our ANeTCM model.

With the advent of the pre-trained model era [22], models have been able to achieve good performance on domain tasks. Roberta [11] achieves significant performance improvement across NLP tasks by increasing the amount of data used for training and eliminating the next sentence prediction objective from the BERT model. Lawformer [23] is pre-trained for the legal text domain and significantly improves the performance of natural language processing tasks in the legal domain. Liu et al. [24] proposed a two-stage transfer learning model to generate TCM prescriptions from a few medical records and TCM documentary resources. However, pre-training models for TCM-NER tasks are currently lacking or not available as open source.

Meanwhile, the imbalance in entity categories has been a challenge for the NER task [25], [26] introduced a smoothing factor based on the cross-entropy function to reduce the contribution of the weights of the easily recognizable samples. Li et al. [27] propose a gradient equalisation mechanism that performs a corresponding normalisation based on the proportions of the sample gradient mode-length distributions, allowing for a more balanced contribution of the various types of samples to the updating of the model parameters. In order to achieve a more balanced distribution of data, various methods have been proposed to over-sample the minority categories [28] and under-sample the majority categories [29].

However, the studies conducted by the aforementioned researchers on domain knowledge and data imbalance have not been incorporated into the TCM-NER task, taking into account the specific characteristics of the domain. We propose the ANeTCM model for TCM-NER, which enhances the benefits of TCM in terms of domain knowledge acquisition and balancing entity classes.

III. METHOD

Our ANeTCM approach utilizes large amount of medical case data to continuously pretraining model, and convert the sequence labelling task into a machine reading comprehension task (see Figure 1). We detail the methodology in this follows.

A. READING COMPREHENSION DATA LABELLING

We are building a machine reading comprehension model based on a large-scale pre-trained BERT model. Given an input $X = x_1, x_2, \ldots, x_n$, where n is denoted as the nth word in the sentence, and then every phrase is found in X. Firstly, the dataset is converted into a ternary form of (Q, A, C), where Q denotes the question generation template, A uses $x_{start,end}$ to denote the start and end positions of the entities in the sentence, and C represents the input text. Second, a problem $q = q_1, q_2, \ldots, q_m$ is generated for the TCM

[1] https://www.yiankb.com/
[2] https://tianchi.aliyun.com/competition/entrance/531824/information
[3] https://github.com/cshan-github/TCM_NER_datasets

TABLE 1. Conceptual description of all entities.

dataset	Entity	Descriptions
TCM-SYN	Drug	中药名称，指在中医理论指导下，用于预防、治疗、诊断疾病并具有康复与保健作用的物质。中药主要来源于天然药及其加工品，包括植物药、动物药、矿物药及部分化学、生物制品类药物。
	Drug ingredient	中药组成成分，指中药复方中所含有的所有与该复方临床应用目的密切相关的药理活性成分。
	Disease	指人体在一定原因的损害性作用下，因自稳调节紊乱而发生的异常生命活动过程，是特定的异常病理情形，而且会影响生物体的部分或是所有器官。
	Symptom	指疾病过程中机体内的一系列机能、代谢和形态结构异常变化所引起的病人主观上的异常感觉或某些客观病态改变。
	Syndrome	中医学专用术语，概括为一系列有相互关联的症状总称，即通过望、闻、问、切四诊所获知的疾病过程中表现在整体层次上的机体反应状态及其运动、变化，简称证或者候，是指不同症状和体征的综合表现，单一的症状和体征无法表现一个完整的证候。
	Disease group	疾病涉及有人体组织部位的疾病名称的统称概念，非某项具体医学疾病。
	Food	指能够满足机体正常生理和生化能量需求，并能延续正常寿命的物质。对人体而言，能够满足人的正常生活活动需求并利于寿命延长的物质称之为食物。
	Food group	中医中饮食养生中，将食物分为寒热温凉四性，同时中医药禁忌中对于具有某类共同属性食物的统称。
	Person group	中医药的适用及禁忌范围内相关特定人群。
	Drug group	具有某一类共同属性的药品类统称概念，非某项具体药品名。
	Drug dosage	物在供给临床使用前，均必须制成适合于医疗和预防应用的形式，成为药物剂型。
	Druge taste	药品的性质和气味。
	Drug efficacy	药品的主治功能和效果的统称。
TCM-DIS	SYM	指疾病过程中机体内的一系列机能、代谢和形态结构异常变化所引起的病人主观上的异常感觉或某些客观病态改变。
	CAU	指的是在一定原因作用下自稳调节紊乱而发生地异常生命活动过程，并引发一系列代谢、功能、结构的变化，表现为症状、体征和行为的异常。
	HER	用以预防、治疗及诊断疾病的物质。在理论上，药物是指凡能影响机体器官生理功能及细胞代谢活动的化学物质都属于药物的范畴。
	PRE	物在供给临床使用前，均必须制成适合于医疗和预防应用的形式，成为药物剂型。
	EFF	药品的主治功能和效果的统称

entity, where m is denoted as the mth word in the problem. A triple $(q_y, x_{start,end}, X)$ can be obtained by generating a problem q_y. Therefore, we constructed the problem of prior knowledge of entities using entity definitions, as shown in Table 1.

B. ANETCM FRAMEWORK

Our model mainly consists of three modules: continuously pre-training module, feature extraction Module and normal distribution sampling module.

1) CONTINUOUSLY PRE-TRAINING

We proposed continuous pre-training module that builds upon a foundational pre-trained language model, such as BERT (Bidirectional Encoder Representations from Transformers) or its derivatives. The core idea is to further pre-train this model on a large corpus of medical case texts, which are rich in TCM-specific terminology and context. This process can be mathematically represented as follows:

Given a pre-trained language model M_0 with parameters θ_0, we aim to optimize these parameters through additional training on a TCM-specific corpus D_{TCM}. The objective is to minimize the loss function $\mathcal{L}$ over the corpus, thus refining the model parameters to θ_{TCM}:

$$\theta_{\mathrm{TCM}} = \arg\min_{\theta} \mathcal{L}\left(\mathcal{M}(\theta), \mathcal{D}_{\mathrm{TCM}}\right) \tag{1}$$

The corpus D_{TCM} consists of annotated medical case reports, clinical notes, and other relevant documents sourced from various TCM institutions and publications. Each document in the corpus is tokenized and processed to align with the input requirements of the pre-trained model. Special attention is given to the accurate representation of TCM-specific entities such as drug, syndromes, symptoms, and food.

The continuous pre-training involves a multi-step procedure. We employ the MLM objective, where certain tokens in the input are masked, and the model is trained to predict these tokens based on the surrounding context. Formally, for a given sequence $X = x_1, x_2, \ldots, x_n$, a subset of tokens x_m is masked, and the model aims to maximize the likelihood:

$$\mathcal{L}_{MLM} = -\sum_{x \in \mathbf{x}_m} \log P\left(x|\mathbf{x}_{\backslash m}; \theta\right) \tag{2}$$

Additionally, the NSP task is incorporated to enhance the model's understanding of sentence-level coherence. Given two sequences x_A and x_B, the model is trained to predict whether x_B follows x_A:

$$\mathcal{L}_{NSP} = -\sum_{(\mathbf{x}_A, \mathbf{x}_B)} \log P(\text{ isNext } |\mathbf{x}_A, \mathbf{x}_B; \theta) \tag{3}$$

2) FEATURE EXTRACTION

The feature extraction module is designed to leverage the strengths of MRC and GLU models to effectively capture and utilize the intricate features present in TCM texts. This hybrid approach ensures a robust representation of contextual and domain-specific information, which is essential for accurate NER.

The MRC model is employed to understand and extract relevant information from the text by posing it as a question-answering task. Given a context C and a query Q, the MRC model predicts the start and end positions (s, e) of the answer span within the context. The probability distributions for the start and end positions are computed as:

$$P_s = \text{softmax}\left(\mathbf{W}_s\left[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\right]\right) \tag{4}$$

$$P_e = \text{softmax}\left(\mathbf{W}_e\left[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\right]\right) \tag{5}$$

where h_i represents the hidden state of the context at position i. W_s and W_e are learnable weight matrices.

GLU models are used to further refine the extracted features by gating mechanisms that control the information flow. The GLU layer can be defined as:

$$\text{GLU}(\mathbf{X}) = (\mathbf{X} \cdot \mathbf{W}_1 + \mathbf{b}_1) \odot \sigma\left(\mathbf{X} \cdot \mathbf{W}_2 + \mathbf{b}_2\right) \tag{6}$$

where X is the input feature matrix, W_1 and W_2 are weight matrices, b_1 and b_2 are bias vectors, σ is the sigmoid activation function, and $\odot$ denotes element-wise multiplication.

The integration of MRC and GLU models involves a multi-stage process where the MRC model first extracts preliminary features from the TCM text. These features are then passed through the GLU layers to refine and enhance their representation.

3) NORMAL DISTRIBUTION SAMPLING

The normal distribution-based (ND) loss function is designed to address class imbalance by assigning different weights to the loss contributions from different classes. The probability density function (PDF) of the normal distribution is used to calculate these weights. Given the logits y produced by the model, the softmax output $\check{y}$ is adjusted using the normal distribution:

$$y_{\text{true normal}} = \begin{bmatrix} \frac{f(0;y_{\text{true}},\sigma)}{f(0;y_{\text{true}},\sigma)+f(1;y_{\text{true}},\sigma)}, \\ \frac{f(1;y_{\text{true}},\sigma)}{f(0;y_{\text{true}},\sigma)+f(1;y_{\text{true}},\sigma)} \end{bmatrix} \tag{7}$$

where $f(x; y_{ture}, \sigma)$ represents the probability density function (PDF) of the normal distribution. Here, x is a discrete value (0 or 1 in this case), y_{ture} is the true label, and σ is the standard deviation of the distribution.

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{8}$$

where μ is the mean and σ is the standard deviation.

The weighted loss function is then defined as:

$$\begin{aligned} L_{\text{weighted\_normal}}\left(y_{\text{true}}, \hat{y}\right) = &-w_0 \cdot y_{\text{true\_normal}}[0] \cdot \log(\hat{y}[0]) \\ &- w_1 \cdot y_{\text{true\_normal}}[1] \cdot \log(\hat{y}[1]) \end{aligned} \tag{9}$$

where w_0 and w_1 are weights derived from the normal distribution to balance the loss contributions.

IV. EXPERIMENTAL AND ANALYSIS

In the following, we evaluate it on the two TCM-NER datasets and compare it with other competitive models. Finally, we present the experimental results and the experimental results are all model replication results.

A. DATASETS

In order to verify the effectiveness of our method, we conducted experimental comparisons on the two TCM-NER datasets. The two datasets are publicly available datasets provided by the TCM-NER Big Data Competition and GitHub. We counted the number of each type of entity in the data, as shown in Figure 2 and Figure 3. The TCM-SYN dataset include drug (药品), drug ingredient (药物成分), disease (疾病), symptom (症状), syndrome (证候), disease group (疾病分组), food (食物), food group (食物分组), person group (人群), drug group (药品分组), drug dosage (药物剂型), drug taste (药物性味) and drug efficacy (中药功效). The TCM-DIS dataset include SYM (症状), CAU (病因), HER (药草), PRE (制剂), EFF (功效). In our experiments, we converted the original dataset into the format required by the MRC task, dividing the ratio by 8:2.

B. APPROACHES COMPARED

We compared the proposed ANeTCM framework with the following baseline and competitive models. At the same time, experimental comparisons were also made between other pre-trained models and algorithms used to mitigate sample imbalance.

- **Baseline models:** This method incorporates classical baseline models including BiLSTM-CRF [30], BERT-softmax [22], BERT-BiLSTM-CRF [31] and MRC [12].
- **Competitive models:** These three competitive models are applied to the public and TCM-NER domains,

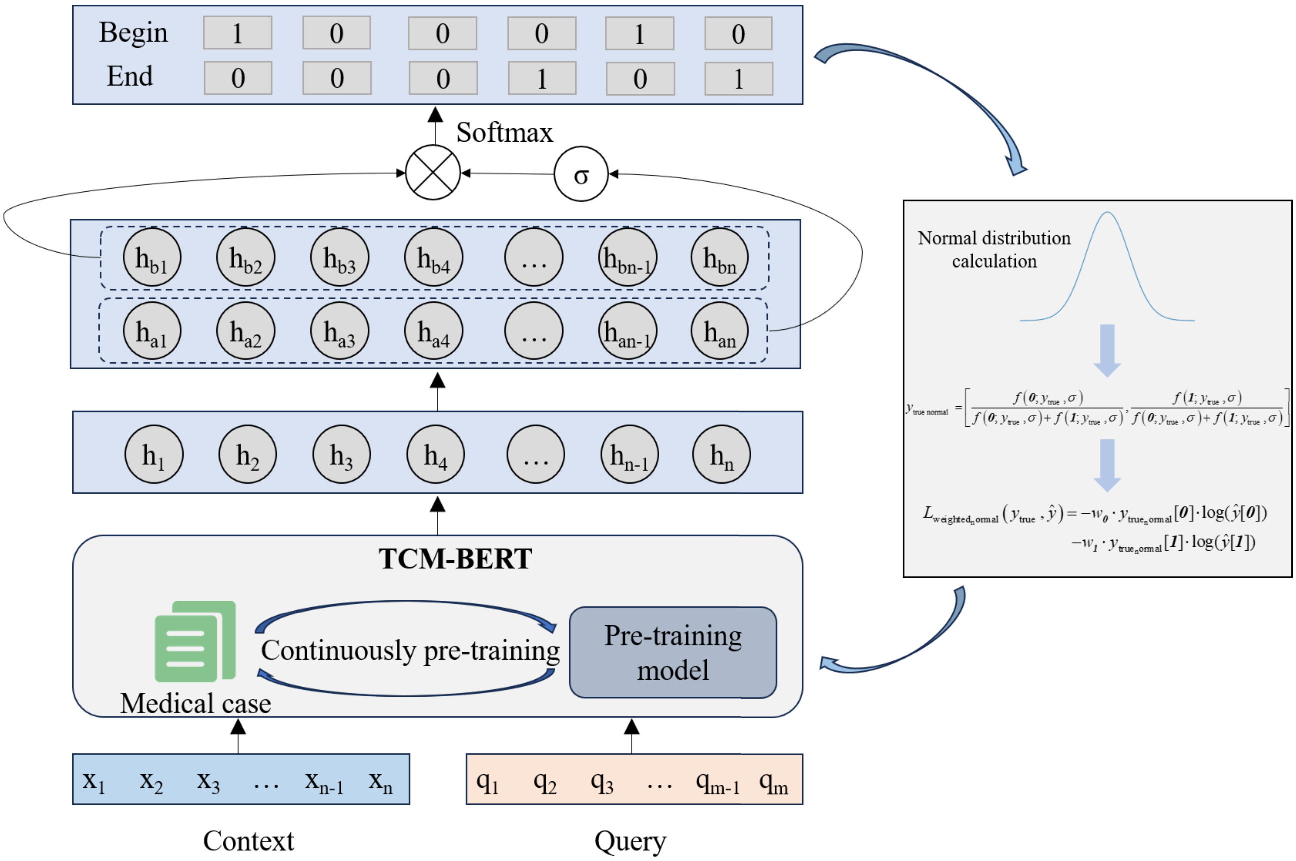


FIGURE 1. The architecture of the ANeTCM.

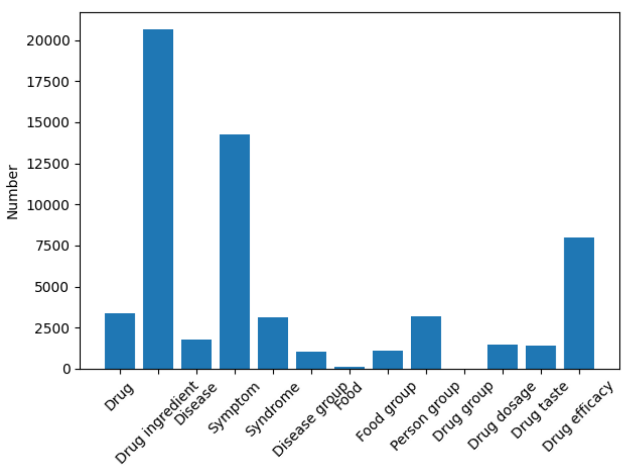


FIGURE 2. Statistical chart of the TCM-SYN dataset.

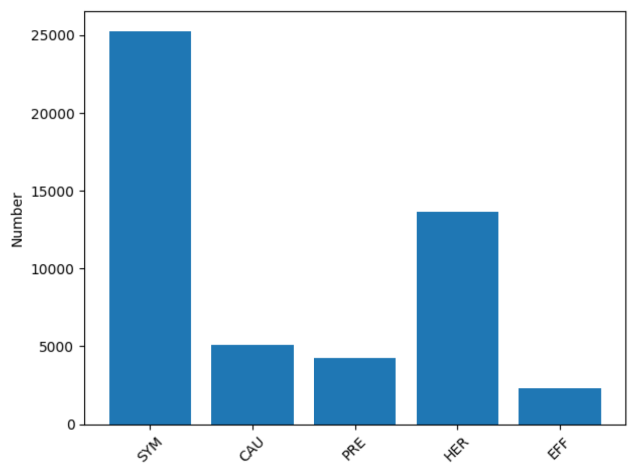


FIGURE 3. Statistical chart of the TCM-DIS dataset.

respectively. Yan et al. [32] proposed formulating the NER subtask as a task of generating entity-spanning sequences and addressing it using a unified sequence-to-sequence (USTS) framework. The PIQN model [17] establishes a globally learnable instance query, eliminating the need for an external knowledge base and manual construction of instance queries to extract entities from sentences in a parallel manner. Zhao et al. [33] proposed a dynamic optimisation-based ensemble learning (DOEL) approach for Chinese medicine named entity recognition.

- **Pre-train models:** This comparison of pre-trained models on the included BERT$_{base}$ [22], BERT$_{wwm}$ [34], Roberta [11], Macbert [35].

- **imbalance algorithms:** The loss balance algorithms included BCE_{loss}, $Focal_{loss}$ [26], GHM_{loss} [27] and $DOEL_{loss}$ [33].

Our model and baseline models were implemented with Pytorch on a server with one A100. To ensure fair comparisons, we applied the same hyper-parameter setting used by [33] for all of the methods. The hyper-parameters are shown in Table 2. The evaluation indicators mainly use Precision (P), Recall (R) and F1. Recall represents the ratio of true positives among all actual positives, precision denotes the ratio of true positives among all predicted positives, and F1 score represents the harmonic mean of recall and precision.

TABLE 2. Hyper-parameters used in all models.

Parameter	Value
Optimizer	AdamW
learning rate	0.000002
Batch size	40
Maximum sentence length	512
Random seed	2023
Dropout	0.1

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$F1 = \frac{2-P-R}{P + R} \tag{12}$$

where TP represents true positive, FN represents false negative, and FP represents false positive.

C. EXPERIMENT RESULTS

1) MAIN EXPERIMENTAL ANALYSIS

To validate the effectiveness of our model for the TCM-NER task, we compared other mainstream models and the experimental results are shown in Table 3.

Firstly, we compared different classical models BiLSTM-CRF,BERT-softmax, BERT-BiLSTM-CRF and MRC on the TCM-SYN and TCM-DIS datasets, which achieved state-of-the-art performance. The reasons are as follows: **(i)** methods such as BiLSTM-CRF, BERT-softmax and BERT-BiLSTM-CRF perform well in general-purpose NER tasks, but the terminology and linguistic structure of the TCM domain are unique. These methods rely on pre-trained general-purpose language models and cannot fully utilize domain-specific information. By introducing the MRC mechanism, we can transform the NER task into a Q&A question and utilize the question template to effectively guide the model to focus on domain-specific entities. This approach is especially well-suited for handling intricate terminology in the TCM domain. **(ii)** BERT, as a pre-trained deep learning model, effectively captures contextual information through a bidirectional Transformer structure. However, BERT, when combined with softmax or BiLSTM-CRF, still struggles to handle long-distance dependencies and complex contexts, despite its advancements in sequence annotation. The strategy of combining with MRC enhances the model's recognition ability in complex contexts by guiding the model to understand entity relationships in specific contexts through constructive questions. Additionally, GLU further assists the model in effectively filtering and controlling the information flow when dealing with long sequences, thereby enhancing the model's ability to understand and remember the context. **(iii)** The loss functions of traditional Named Entity Recognition (NER) models mainly consist of cross-entropy or Conditional Random Field (CRF) losses. However, these may not be effective in handling category imbalance or complex sequences. We have developed an enhanced loss function within the MRC framework, integrated with domain knowledge, to effectively address the challenges of category imbalance and intricate entity boundary identification. By weighting different categories and boundaries during the training process, the model can more accurately identify named entities in the TCM domain.

Secondly, We compared the mainstream models USTS, PIQN and DOEL and also obtained state-of-the-art performance. The reasons are as follows: **(i)** Compared to USTS and PIQN, although they both perform well in the generic domain, they do not incorporate the domain-specific characteristics required for TCM named entity recognition. **(ii)** Although DOEL improves the loss function by incorporating TCM named entity recognition features, it still fails to address the lack of domain knowledge, resulting in a model that is inferior to ours in terms of recognition performance.

2) BASELINE PRE-TRAINING MODEL EXPERIMENTAL ANALYSIS

The lack of domain knowledge leads to poor recognition performance. Therefore, we have utilized various baseline models for experimental comparisons to confirm the effectiveness of pre-training through continuous pre-training, as shown in Table 4.

We can see that compared to other pre-trained models, we obtained the highest performance by using a large amount of medical case data for continuous pre-training. We are state-of-the-art compared to other pre-trained models. The reasons are as follows: **(i)** The medical case data contains rich TCM-related terminology and expressions. The model is pre-trained on this domain-specific data to better capture the linguistic features and expertise in the TCM domain. **(ii)** Continuous pre-training enables the model to understand the context and usage of TCM-related words more deeply, and improves the accuracy of the model in recognizing named entities. **(iii)** By pre-training on specialized domain data, the model implicitly learns the TCM knowledge system and logic, enhancing the accuracy and consistency of named entity recognition.

TABLE 3. Comparative experimental results of different mainstream models.

Model	TCM-SYN			TCM-DIS		
	Precision(%)	Recall(%)	F1(%)	Precision(%)	Recall(%)	F1(%)
BiLSTM-CRF [30]	81.25	81.04	80.22	56.72	62.34	59.39
BERT-softmax [22]	82.76	70.94	76.93	54.66	59.88	57.15
BERT-BiLSTM-CRF [31]	84.76	84.81	84.16	60.22	69.83	64.67
MRC [12]	85.28	84.83	84.52	61.73	70.15	65.67
USTS [32]	86.72	87.31	87.01	64.85	68.11	66.44
PIQN [17]	90.28	91.64	90.96	**70.86**	65.92	68.30
DOEL [33]	85.81	85.47	85.08	-	-	-
ANeTCM	**93.61**	**95.21**	**94.40**	67.21	**73.95**	**70.42**

TABLE 4. Comparative experimental results of different baseline models.

Model	TCM-SYN			TCM-DIS		
	Precision(%)	Recall(%)	F1(%)	Precision(%)	Recall(%)	F1(%)
BERT-base	91.32	91.89	91.60	65.65	64.23	64.93
BERT-wwm	92.41	93.10	92.75	66.12	66.61	66.36
Roberta	**93.94**	94.24	94.09	**68.77**	71.47	70.09
Macbert	93.49	93.73	93.60	68.13	70.66	69.37
TCM-BERT	93.61	**95.21**	**94.40**	67.21	**73.95**	**70.42**

3) COMPARATIVE EXPERIMENTAL ANALYSIS OF LOSS FUNCTIONS

At the same time, we aim to further verify the effectiveness of our proposed loss function in mitigating data imbalance. We conducted an experimental comparison with other loss functions, as shown in Table 5.

From the table, it can be seen that our proposed loss function obtains state-of-the-art performance compared to other loss functions. The reasons are as follows: **(i)** The cross-entropy loss function is one of the most commonly used loss functions, but it lacks a specific mechanism to address data imbalance. When the categories are unbalanced, the model will tend to make better predictions for the majority class and ignore the minority class. This leads to poor identification of rare classes by the model. **(ii)** Focal loss alleviates this problem to some extent. It reduces the influence of the majority class samples by giving greater weights to the difficult-to-classify samples, thus improving the recognition accuracy of the minority class. However, the Focal loss function is still limited in its effectiveness when dealing with extremely unbalanced data, as it cannot fully address the issue of scarcity of minority class samples. **(iii)** GHM loss enables challenging-to-train samples to receive more focus by adjusting the gradient. This method partially balances the importance of the samples; however, in practice, tuning the parameters of GHM is more complex, and the impact is not always significant when handling extremely imbalanced data.

Therefore, this enhanced loss function not only effectively addresses the data imbalance issue but also enhances the model's capability to identify samples from minority classes. This suggests that by analyzing the data characteristics in detail and combining the advantages and disadvantages of each loss function, our new loss function is able to better balance the imbalance phenomenon in the data of TCM named entities recognition, thus improving the overall recognition effect.

4) ANALYSIS OF ABLATION EXPERIMENTS

We conducted ablation experiments in order to verify the validity of each module of the model, as shown in Table 6.

We found that the various modules of the model were effective in improving the model performance. The reasons are as follows: MRC provides powerful context-awareness and fine-grained semantic analysis, enabling the model to understand and answer complex questions more accurately. Second, GLU selectively conveys essential information through its gating mechanism, filters out irrelevant noise, and introduces non-linear factors to enhance the expressive capacity and learning efficiency of the model. This also helps to alleviate the issue of gradient vanishing, ensuring the stability of the training process. Finally, the enhanced loss function is optimized specifically for the particular task. This enables better handling of the distribution of sample errors, accelerates the model's convergence process, and enhances training efficiency and final performance. This combination of multiple modules not only leverages their respective advantages but also enhances overall performance through synergy, resulting in improved

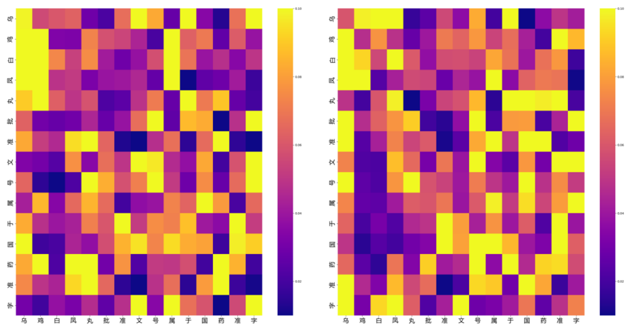
TABLE 5. Comparative experimental results of different loss function.

Model	TCM-SYN			TCM-DIS		
	Precision(%)	Recall(%)	F1(%)	Precision(%)	Recall(%)	F1(%)
CE	90.15	91.82	90.75	62.74	68.44	65.46
Focal	92.60	92.81	92.71	64.17	70.39	67.13
GHM	92.85	93.43	93.13	65.24	71.41	68.18
Our	**93.61**	**95.21**	**94.40**	**67.21**	**73.95**	**70.42**

TABLE 6. Ablation experimental results.

Model			TCM-SYN			TCM-DIS		
MRC	GLU	ND	Precision(%)	Recall(%)	F1(%)	Precision(%)	Recall(%)	F1(%)
√			85.28	84.83	84.52	61.73	70.15	65.67
√	√		91.02	92.11	91.52	65.12	71.66	68.23
√		√	92.91	93.47	93.19	64.81	70.36	67.47
√	√	√	**93.61**	**95.21**	**94.40**	**67.21**	**73.95**	**70.42**

reliability and effectiveness of the model in practical applications.



(a) Baseline (b) Our

FIGURE 4. Visual comparison of attention weights.

5) WEIGHT VISUALISATION AND ANALYSIS

Finally, we conducted the attention weight visualization comparison experiment to further validate the model's validity, as shown in Figure 4.

We found that the distribution of attentional weights in the baseline model was more spread out. Attentional weights were distributed across a larger number of positions, indicating that the baseline model lacks a clear focus of attention on the words at each position when processing the input sequence. This may impact its ability to capture long-distance dependencies. However, the weight distribution of our model is more concentrated compared to the baseline model. This suggests that the improved model can explicitly identify and focus on important words or contextual information when processing input sequences. This ability contributes to the enhanced comprehension and generation of the model. Therefore, by comparing the attention weight graphs, we can visualize the advantages of the enhanced model on the attention mechanism, further validating the effectiveness of our model.

V. CONCLUSION

In this paper, we propse a ANeTCM framework that makes the elicitation of TCM domain knowledge and data balance. Specifically, we utilize medical case data for continuous pre-training to acquire domain knowledge, which is then combined with MRC and GLU models to effectively capture important features. In addition, the imbalance of data is effectively alleviated by improving the loss function. Experiments show our method's effectiveness across TCM-NER datasets, achieving state-of-the-art result.

In the further, we will further optimise the time complexity of the model, we also hope that our framework will extend to generalized NER tasks.

ETHICS STATEMENT

The authors use only publicly available datasets and relatively low compute amounts while conducting their experiments to enable reproducibility.

ACKNOWLEDGMENT

The authors thank all the anonymous reviewers for their insightful comments.

REFERENCES

[1] S.-Y. Pan, Q. Nie, H.-C. Tai, X.-L. Song, Y.-F. Tong, L.-J.-F. Zhang, X.-W. Wu, Z.-H. Lin, Y.-Y. Zhang, D.-Y. Ye, Y. Zhang, X.-Y. Wang, P.-L. Zhu, Z.-S. Chu, Z.-L. Yu, and C. Liang, "Tea and tea drinking: China's outstanding contributions to the mankind," *Chin. Med.*, vol. 17, no. 1, pp. 1–40, Dec. 2022.

[2] X. Kang, D. Jin, L. Jiang, Y. Zhang, Y. Zhang, X. An, L. Duan, C. Yang, R. Zhou, Y. Duan, Y. Sun, and F. Lian, "Efficacy and mechanisms of traditional Chinese medicine for COVID-19: A systematic review," *Chin. Med.*, vol. 17, no. 1, p. 30, Dec. 2022.

[3] M. Konkol and M. Konopík, "Named entity recognition for highly inflectional languages: Effects of various lemmatization and stemming approaches," in *Proc. 17th Int. Conf.* Brno, Czech Republic. Springer, Sep. 2014, pp. 267–274.
[4] A. Tikayat Ray, O. J. Pinon-Fischer, D. N. Mavris, R. T. White, and B. F. Cole, "AeroBERT-NER: Named-entity recognition for aerospace requirements engineering using BERT," in *Proc. AIAA SCITECH Forum*, Jan. 2023, p. 2583.
[5] D. Li, L. Yan, J. Yang, and Z. Ma, "Dependency syntax guided BERT-BiLSTM-GAM-CRF for Chinese NER," *Exp. Syst. Appl.*, vol. 196, Jun. 2022, Art. no. 116682.
[6] W. El-Shafai, A. A. Mahmoud, E.-S. M. El-Rabaie, T. E. Taha, O. F. Zahran, A. S. El-Fishawy, M. Abd-Elnaby, and F. E. A. El-Samie, "Traditional Chinese medicine automated diagnosis based on knowledge graph reasoning," *Comput., Mater. Continua*, vol. 71, no. 1, pp. 159–170, 2022.
[7] Y. Xie, "A TCM question and answer system based on medical records knowledge graph," in *Proc. Int. Conf. Comput. Data Sci. (CDS)*, Aug. 2020, pp. 373–376.
[8] J. Shi, M. Sun, Z. Sun, M. Li, Y. Gu, and W. Zhang, "Multi-level semantic fusion network for Chinese medical named entity recognition," *J. Biomed. Informat.*, vol. 133, Sep. 2022, Art. no. 104144.
[9] S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N. A. Smith, "Don't stop pretraining: Adapt language models to domains and tasks," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 8342–8360.
[10] A. Olmo, S. Sreedharan, and S. Kambhampati, "GPT3-to-plan: Extracting plans from text using GPT-3," in *Proc. FinPlan*, 2021, p. 24.
[11] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.
[12] X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, and J. Li, "A unified MRC framework for named entity recognition," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 5849–5859.
[13] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 933–941.
[14] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. ICML*, 2001, p. 3.
[15] L. Liu, J. Shang, X. Ren, F. Xu, H. Gui, J. Peng, and J. Han, "Empower sequence labeling with task-aware neural language model," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 1–8.
[16] P. Cao, Y. Chen, K. Liu, J. Zhao, and S. Liu, "Adversarial transfer learning for Chinese named entity recognition with self-attention mechanism," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 182–192.
[17] Y. Shen, X. Wang, Z. Tan, G. Xu, P. Xie, F. Huang, W. Lu, and Y. Zhuang, "Parallel instance query network for named entity recognition," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 947–961.
[18] D. Zhang, C. Xia, C. Xu, Q. Jia, S. Yang, X. Luo, and Y. Xie, "Improving distantly-supervised named entity recognition for traditional Chinese medicine text via a novel back-labeling approach," *IEEE Access*, vol. 8, pp. 145413–145421, 2020.
[19] H. Xu, H. Liu, Q. Jia, Y. Zhan, Y. Zhang, and Y. Xie, "A nested named entity recognition method for traditional Chinese medicine records," in *Proc. 7th Int. Conf.*, Dublin, Ireland. Springer, Jul. 2021, pp. 488–497.
[20] Y. Ma, Y. Liu, D. Zhang, J. Zhang, H. Liu, and Y. Xie, "A multigranularity text driven named entity recognition CGAN model for traditional Chinese medicine literatures," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, Sep. 2022.
[21] Y. Ma, H. Liu, D. Zhang, C. Gao, and Y. Liu, "A named entity recognition method enhanced with lexicon information and text local feature," *Tehnički vjesnik*, vol. 30, no. 3, pp. 899–906, 2023.
[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, Jun. 2019, pp. 4171–4186.
[23] C. Xiao, X. Hu, Z. Liu, C. Tu, and M. Sun, "Lawformer: A pre-trained language model for Chinese legal long documents," *AI Open*, vol. 2, pp. 79–84, Jan. 2021.
[24] Z. Liu, C. Luo, D. Fu, J. Gui, Z. Zheng, L. Qi, and H. Guo, "A novel transfer learning model for traditional herbal medicine prescription generation from unstructured resources and knowledge," *Artif. Intell. Med.*, vol. 124, Feb. 2022, Art. no. 102232.
[25] X. Wang and Y. Wang, "Sentence-level resampling for named entity recognition," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2022, pp. 2151–2165.
[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
[27] B. Li, Y. Liu, and X. Wang, "Gradient harmonized single-stage detector," in *Proc. AAAI Conf. Artif. Intell.*, Sep. 2019, vol. 33, no. 1, pp. 8577–8584.
[28] J. Byrd and Z. Lipton, "What is the effect of importance weighting in deep learning?" in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 872–881.
[29] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018.
[30] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*.
[31] Z. Dai, X. Wang, P. Ni, Y. Li, G. Li, and X. Bai, "Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records," in *Proc. 12th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2019, pp. 1–5.
[32] H. Yan, T. Gui, J. Dai, Q. Guo, Z. Zhang, and X. Qiu, "A unified generative framework for various NER subtasks," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, 2021, pp. 5808–5822.
[33] Z. Zhao, Y. Qian, Q. Liu, J. Chen, and Y. Liu, "A dynamic optimization-based ensemble learning method for traditional Chinese medicine named entity recognition," *IEEE Access*, vol. 11, pp. 99101–99110, 2023.
[34] Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang, "Pre-training with whole word masking for Chinese BERT," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 3504–3514, 2021.
[35] Y. Cui, W. Che, T. Liu, B. Qin, S. Wang, and G. Hu, "Revisiting pre-trained models for Chinese natural language processing," in *Proc. Findings Assoc. Comput. Linguistics (EMNLP)*, 2020, pp. 657–668.

YUANYU FENG received the degree from Guizhou Medical University, Guizhou, China, in 2021.

She is currently with the Department of Anesthesiology, The Affiliated Hospital of Guizhou Medical University, Guizhou. Her research interest includes smart medical care.

YAN ZHOU received the degree from Sun Yat-sen University, in 2013.

She is currently with the Department of Children's Health Care, Anhui Provincial Children's Hospital/Children's Hospital of Fudan University Anhui Hospital, Hefei, China. Her research interest includes smart medical care.