

Research Article

A Multigranularity Text Driven Named Entity Recognition CGAN Model for Traditional Chinese Medicine Literatures

Yuekun Ma ^{1,2,3}, Yun Liu ⁴, Dezheng Zhang ^{1,3}, Jiye Zhang ², He Liu ²
and Yonghong Xie ^{1,3}

¹School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

²School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China

³Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China

⁴Beijing Haidian Hospital, Beijing 100080, China

Correspondence should be addressed to Dezheng Zhang; zdzchina@ustb.edu.cn

Received 13 June 2022; Accepted 8 July 2022; Published 24 September 2022

Academic Editor: Dalin Zhang

Copyright © 2022 Yuekun Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recognition of Traditional Chinese Medicine (TCM) entities from different types of literature is challenging research, which is the foundation for extracting a large amount of TCM knowledge existing in unstructured texts into structured formats. The lack of large-scale annotated data makes unsatisfactory application of conventional deep learning models in TCM text knowledge extraction. Some other unsupervised methods rely on other auxiliary data, such as domain dictionaries. We propose a multigranularity text-driven NER model based on Conditional Generation Adversarial Network (MT-CGAN) to implement TCM NER with small-scale annotated corpus. In the model, a multigranularity text features encoder (MTFE) is designed to extract rich semantic and grammatical information from multiple dimensions of TCM texts. By differentiating the conditional constraints of the generator and discriminator of MT-CGAN, the synchronization between the generated tag labs and the named entities is guaranteed. Furthermore, seeds of different TCM text types are introduced into our model to improve the precision of NER. We compare our method with other baseline methods to illustrate the effectiveness of our method on 4 kinds of gold-standard datasets. The experiment results show that the standard precision, recall, and F1 score of our method are higher than the state-of-the-art methods by 0.24~8.97%, 0.89~12.74%, and 0.01~10.84%. MT-CGAN is able to extract entities from different types of TCM literature effectively. Our experimental results indicate that the proposed approach has a clear advantage in processing TCM texts with more entity types, higher sparsity, less regular features, and a small-scale corpus.

1. Introduction

Named Entity Recognition (NER) in Traditional Chinese Medicine (TCM) texts is an important task of TCM knowledge extraction, which refers to extracting the instances of TCM domain concepts from large-scale unstructured TCM texts and identifying their concept types. The grammatical form of classical Chinese and semantic peculiarities of TCM texts leads to the complexity of NER in the TCM field being higher than in other fields.

At present, there are deficiencies in various mainstream NER methods for general domain texts and TCM texts. Since the domain characteristics of TCM texts are not considered, the NER methods of generic domain texts have weak

generalization ability and unsatisfactory recognition results in TCM texts. Compared with other methods, the NER models based on deep learning in the field of TCM texts have achieved better results, but they tend to need a large-scale training corpus. The highly specialized nature of TCM texts leads to the high cost of training corpus annotations. The lack of large-scale annotated data makes unsatisfactory application of conventional deep learning models in TCM text knowledge extraction. How to use the limited annotated corpus to realize NER for massive TCM texts has become an urgent problem to be solved.

Researchers have actively explored in NER domain with a lack of annotation data. Jia et al. [1] considered NER in TCM as a text span detection task, classifying all text spans

and obtaining the relevant knowledge entities from the sentences based on the classification results. The approach does not require a large-scale annotated corpus but is based on a high-quality domain dictionary. Han Yuanbo Tao [2] combined Generative Adversarial Networks (GAN) with the BiLSTM-Attention-CRF model to solve the problems of lack of annotation data for domain NER tasks and inconsistent annotation of entities in the same document. However, this model requires a large-scale crowdsourced annotated dataset, although the cost of the crowdsourced annotation dataset is lower than that of a professional annotation dataset.

To solve the above problems, this paper introduces the generative adversarial idea to the NER tasks and proposes a Multigranularity Text Features Fused NER model based on CGAN (MT-CGAN). The model takes advantage of the fact that the Conditional Generation Adversarial Network (CGAN) model can generate data with the same distribution characteristics as the real data [3], and self-attention is better at capturing the internal correlation of data or features. Using this model, we extract features at different granularities of TCM texts and improve the domain NER performance under the condition of a small amount of annotated corpus. We summarize the contribution as follows:

- (i) A differential feature-constrained NER model in TCM texts is proposed based on the generative adversarial idea. The model solves the problems of lack of annotation data in TCM texts and the inapplicability of NER models in other domains in the NER tasks of TCM texts.
- (ii) A Multigranularity Text Features Encoder (MTFE) is designed to extract grammatical and semantic features embedded in different dimensions and feature spaces of text from words, sentences, paragraphs, and chapters, respectively, to enhance the multidimensional feature learning capability of the model.
- (iii) An improved U-NET is added to the generator of the MT-CGAN to hierarchically extract the features of the seed label. The extracted features are also passed to the decoding layer via jump links to achieve feature enhancement of the generation process. This further reduces the randomness of the generated content and speeds up the convergence of the model.

The paper is organized as follows. Section 2 introduces the work related to the NER tasks in TCM texts and multigranularity features extraction. Section 3 describes the details of the MT-CGAN. Section 4 describes the experiments for the performance evaluation of the MT-CGAN. Section 5 concludes the paper.

2. Related Work

2.1. NER of TCM Texts Based on Deep Learning. Now, deep learning has become a mainstream technique for solving NER tasks because of its ability to discover hidden features. Deng et al. [4] combined BiLSTM and CRF and applied it to the Chinese medicine text NER, and the model improved the

accuracy of NER because of the powerful ability of the LSTM model to extract global contextual semantic information. Li et al. [5] proposed a model called BiLSTM-Att-CRF by integrating attention into BiLSTM networks and proved that this model can avoid the problem of information loss caused by distance. An et al. [6] proposed a Multihead Self-attention-based BiLSTM-CRF model (MUSA-BiLSTM-CRF) for Chinese clinical named entity recognition. The proposal of the Transformer model [7] opened the era of dominance of pretrained models in NLP [8, 9]. As a typical representative of pretrained models, BERT [10] has shown good performance in many NLP tasks [11, 12]. Qu et al. [13] applied the combination of the BERT model and BiLSTM-CRF to the NER task of TCM texts, and the performance was improved compared with the BiLSTM-CRF model. Qu et al. [13] constructed a named entity recognition model based on Bert-BiLSTM-CRF to solve the problem of fuzzy entity recognition and less labeled data in the field of traditional Chinese medicine. Zhang et al. [14] proposed a semisupervised embedded Semi-BERT-BiLSTM-CRF model. Xiong et al. [15] proposed a method based on a relational graph convolutional network (RGCN), which utilized multisource knowledge in a unified manner for Chinese Clinical NER. All of these NER models employed large-scale annotated samples or pretrained models to maximize performance.

Compared with other deep learning models with end-to-end training, GAN can learn the distribution of real samples and has better robustness, which can alleviate the problem of insufficient labeled samples in deep learning.

2.2. Conditional Generation Adversarial Network. Generative Adversarial Networks (GAN) is a generation framework. Because it can use the adversarial sample [16] to generate data with the same distribution as the real target, it is applied to a variety of generation tasks. For example, Cao [17] used it to restore images and graphs. The original GAN has some limitations, which are mainly reflected in the uncontrollable process of generator generating samples. For this reason, Mirza and Simon [18] proposed CGAN. It aims to change the random generation of the target object and control the data generation process by incorporating specific conditions into the generator and discriminator inputs. It solves the problem that GAN cannot generate specific attributes in the NLP. Feng [19] proposed a Conditional Wasserstein Generative Adversarial Network model (CWGAN) for NER tasks by combining CGAN and a modified Wasserstein Generative Adversarial Network (WGAN-GP). They verified the effectiveness of the CWGAN on the generic domain NER tasks. However, this model does not perform well in the NER task of TCM texts.

At the same time, by analyzing the characteristics of current NER models based on the generative adversarial idea, we find that these models are all based on feature extraction from a single granularity of text. As the carrier of knowledge, text with different granularity contains different information. Text features can be mined more comprehensively through multigranularity text feature extraction.

2.3. Multigranularity Features Extraction. It was found that a multigranularity cognitive mechanism was used to improve the NER task. Zhang and Yang [20] designed a Lattice-LSTM model to encode characters and words, respectively, which achieved the utilization of information on both words and order of words and avoided the problem of word separation errors. Gao et al. [21] proposed contextual multifeature embedding (CMFE) and used it to construct a multifeature semantic fusion model (MFSFM) to realize citation entity recognition. Sun [22] referred to a method of recognizing Chinese clinically named entities based on multistrategy fusion. Li et al. [23] proposed a Flat-Lattice Transformer model, which transformed the Lattice structure into Flat-Lattice to link the semantic information of characters and words together for NER.

Multigranularity feature extraction can obtain richer semantic information than single-granularity feature extraction. In the NER tasks, whether the semantic information can be fully extracted is the key to the performance of the model. Therefore, we propose the MT-CGAN to solve the problem of NER in TCM texts.

3. Model

In this paper, we view the NER tasks in TCM as sequence annotation tasks and use the generative adversarial idea to construct the MT-CGAN model for implementing NER in TCM texts in a small-scale corpus. The description of the tasks is as follows: given an annotated corpus text, $X = \{x_0, x_1, \dots, x_i, \dots, x_n\}$ is the text sequences, $L = \{l_0, l_1, \dots, l_i, \dots, l_n\}$ is the sequence of named entities labels for the corresponding text sequence X , where the x_i represents a character, and x_i represents the entity category label corresponding to x_i . The role of the generator of the model is to generate the sequence of annotations $L' = \{l'_0, l'_1, \dots, l'_i, \dots, l'_n\}$ corresponding to. The discriminator guides the model training by scoring L' so that the generator can generate a sequence of entity labels that approximates L . The following is a detailed description of the model.

3.1. Model Structure. The MT-CGAN model proposed in this paper consists of the generator and the discriminator, and the structure is shown in Figure 1. The generator of the MT-CGAN consists of the MTFE and a modified U-NET (conditionally enhanced U-NET, C-U-NET). The discriminator of the MT-CGAN consists of multiple Convolutional Neural Networks (CNN) [24], normalization operations of each layer, and activation functions.

In this paper, the text attention data generated by the MTFE is used to replace the white noise input of the CGAN generator as the main input of the model generator in the text feature extraction process. Moreover, to improve the quality of label sequence generation, we fuse the multigranularity text features extracted by the MTFE with the TCM text entity distribution features extracted by the encoder of the C-U-NET and as the conditional input Y to the model's generator.

In the discriminator part of the MT-CGAN, we use a small-scale of real sample labels L as the conditional input of

the discriminator, which improves the ability of the model to discriminate authenticity and accelerates the convergence speed of the model training.

The generator of the MT-CGAN generates a sequence of entity labels in vector form, and a Conditional Random Field (CRF) [25] is added to which to convert the vector into an intuitive symbolic representation form. When the MT-CGAN training reaches a steady state, the model loads the trained generator model to generate a sequence L' of text labels. Then L' is used as the input for the CRF to obtain the mapping of text labels T . Finally, the generated text labels T are synchronized with the corresponding text to form the final text annotation sequence.

3.2. Multigranularity Text Features Encoder. The MTFE of the MT-CGAN consists of four feature encoders with different granularity of characters, sentences, paragraphs, and chapters to extract text features from different dimensions.

The Character-Granularity Feature Encoder (CGFE) is implemented based on the GRU [26] network and is used to extract the position information h_c among characters in text. The Sentence-Granularity Feature Encoder (SGFE) is realized based on Transformer's encoder for extracting text features h_s embedded in the textual context within a sentence. The Paragraph-Granularity Feature Encoder (PGFE), also based on Transformer's encoder, is introduced into this model to extract the deeper latent semantic and syntactic features h_p contained among different sentences within a text segment. A local central feature within a chapter is important to the NER in TCM texts, so the Article-Granularity Feature Encoder (AGFE) based on CNN is designed to achieve the extraction of central features h_a from the global scope of the chapter text.

The SGFE and PGFE adopt the attention mechanism that can generate attention data for the corresponding granularity of text while extracting deep semantic features of the text context. Because these data reflect the contextual relevance among tokens, we fused the attention data generated by the SGFE and PGFE to form the main input to the generator of the MT-CGAN.

3.2.1. Character-Granularity Feature Encoder. To ensure that the generated sequence of text labels is consistent with the original text sequence, this paper designs the CGFE based on GRU to extract the relative position feature h_c of characters in TCM texts and uses h_c to constrain the order of the entities label generating.

As shown in Figure 2(a) CGFE consists of an embedding layer, GRU, a fully connected layer, and an activation function. The GRU that can retain information in a sequence for long periods is used to encode character position information in text. It includes a reset gate and an update gate. The extent to which the preceding information is retained is determined by both. The smaller the value of the reset gate is, the more previous information is discarded, while the larger the value of the update gate, the more is retained.

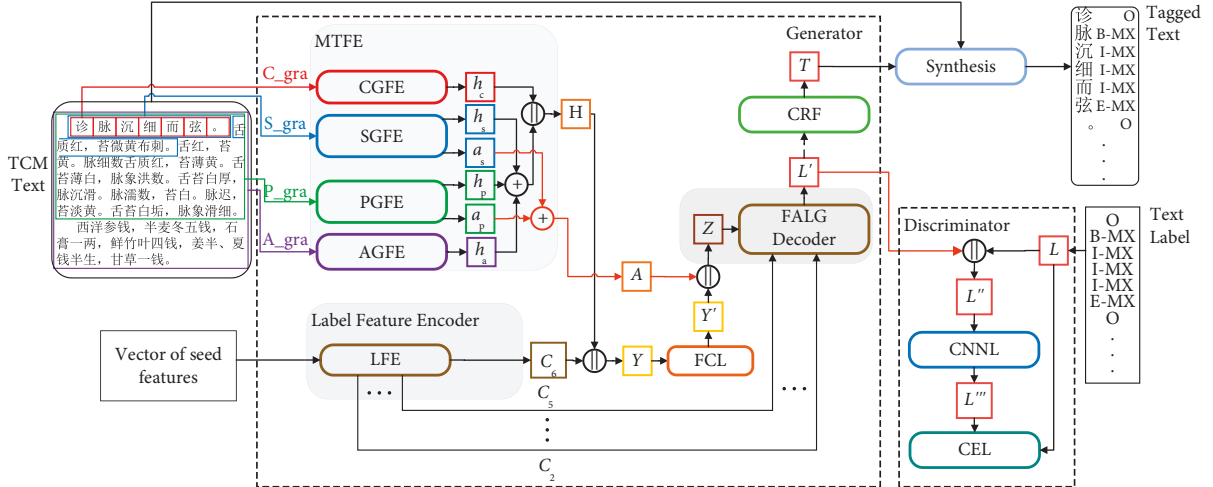


FIGURE 1: Model framework. \parallel is a splicing operation, $+$ is a contraposition addition operation. C_gra , S_gra , P_gra , and A_gra denote characters, sentences, paragraphs, and chapters of text, respectively. FCL represents a fully connected layer. CNNL represents the convolutional neural network Layer, and CEL denotes the cross-entropy loss function.

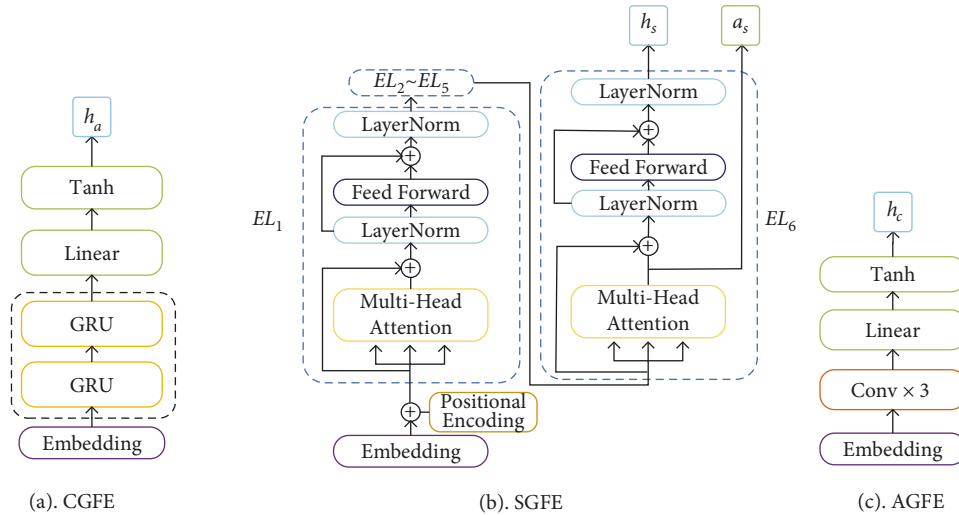


FIGURE 2: Network structure of each component of the MTFE. (a) CGFE (b) SGFE (c) AGFE.

The equation of reset gate is as follows:

$$r_i = \sigma(W^{(r)}e_i + U^{(r)}p_{i-1}), \quad (1)$$

where r_i is the reset gate, e_i is the vector of the current input token, $W^{(r)}$ and $U^{(r)}$ are the reset gate weight matrices, and p_{i-1} is the relative position information of the previous token.

The equation of the update gate is as follows:

$$z_i = \sigma(W^{(z)}e_i + U^{(z)}p_{i-1}), \quad (2)$$

where z_i is the gating controlling the update gate, $W^{(z)}$ and $U^{(z)}$ are the weight matrices of the update gate.

The linear transformation parameters of the update gate and the reset gate are changed in the MT-CGAN model to obtain the candidate hidden state p'_i and the output

p_i , respectively. p'_i and p_i are calculated by (3) and (4), respectively.

$$p'_i = \text{Tanh}(We_i + r_i \odot Up_{i-1}), \quad (3)$$

where \odot is the Hadamard product and the data is mapped between $(-1, 1)$ by the Tanh activation function.

$$p_i = (1 - z_i) \odot p_{i-1} + z_i \odot p'_i, \quad (4)$$

where $(1 - z_i) \odot p_{i-1}$ represents the forgetting of the relative position information of the previous character and $z_i \odot p'_i$ represents the remembering of the current position information.

The text character sequence passes through two GRU layers to obtain the character relative position feature sequence P :

$$P = p_1, p_2, \dots, p_i, \dots, p_n. \quad (5)$$

After the full concatenation and Tanh activation operations, the final character position feature h_c of the text sequence is obtained.

$$h_c = \{h_c^0, h_c^1, \dots, h_c^i, \dots, h_c^n\}, \quad (6)$$

3.2.2. Sentence-Granularity and Paragraph-Granularity Feature Encoder. The entities in the field of TCM texts have their own characteristics. According to Yurui [27], concepts in the field of TCM texts are characterized by heterogeneity, and they have obvious multimeaning and fluidity, and the same concept could have different functions. The boundaries between the categories of entities, attributes, and relations are ambiguous. For example, the word 经脉 appears 29 times throughout the text of *Qian Jin Fang*, and its connotation and scope vary in different contexts, indicating the meanings of Channel of Qi-Blood, Channel Number, and Menses, respectively. Therefore, to achieve accurate NER on TCM texts, the context of the text must be fully used.

The contextual semantic spaces constructed within a sentence and among different sentences are different, and the different semantic spaces have different representations for TCM knowledge and have different roles for NER in the TCM domain. The token level and sentence level semantics are equally important. Therefore, it is necessary to obtain deep-level potential semantic features of text from different scope contexts by using text of both sentence and paragraph granularities as objects.

Transformer can find latent recognition patterns from data and has been proven to have the ability to extract deep semantic features embedded in a text [28]. To benefit from its self-attention mechanism, the Transformer can give the same attention to all elements of the input text sequence and can understand the relationship among different elements of input sequences, although the elements may be far away from each other [29]. Transformer models can model the relationships among words in long sentences or paragraphs, so the sentence-granularity and paragraph-granularity text feature encoders proposed in this paper are both implemented based on the encoder of Transformer.

In the following, we use the SGFE as an example to illustrate how attention data and contextual features of text are extracted based on Transformer. As shown in Figure 2(b), the SGFE is composed of a stack of six identical encoding layers EL_i , where $i = 1, \dots, 6$ and represents the index of the encoding layers. We can obtain a_i and h_i through each encoding layer. a_i and h_i represent the attention data and contextual feature data, respectively. The outputs of the SGFE are a_6 and h_6 obtained from EL_6 . Each encoding layer has two sublayers. The first is a multihead self-attention mechanism, and the second is a fully connected feedforward network. The input of EL_1 is a vector matrix $E_s = \{e_s^0, \dots, e_s^i, \dots, e_s^n\}$, e_s^i denotes the vector corresponding to the i th word in the input text sequence.

The following describes how the attention data and contextual feature data of a text sequence are calculated.

(1) *Attention Data.* Multihead self-attention consists of 8 attention layers running in parallel, each attention layer calculates different attention data fragments of the input text from different semantic spaces, and then the eight fragments are concatenated to form the attention data. The equation for calculating the attention data fragment is as follows:

$$Z_h = \text{softmax}\left(\frac{Q_h \times K_h}{\sqrt{d_k}}\right) \times V_h, \quad (7)$$

where Q_h represents the query matrix, K_h represents the key matrix, V_h represents the value matrix obtained by transforming E_s through the linear matrices $W_h^Q \in \mathbb{R}^{d_k \times d_{mo\ de\ l}}$, $W_h^K \in \mathbb{R}^{d_k \times d_{mo\ de\ l}}$ and $W_h^V \in \mathbb{R}^{d_v \times d_{mo\ de\ l}}$ respectively, and h represents the index of the attention header, $d_k = d_v$.

The output data of the multihead self-attention are concatenated by each attention data fragment according to the following:

$$a_i = \text{concat}(Z_1, Z_2, \dots, Z_8) \times W_O, \quad (8)$$

where $W_O \in \mathbb{R}^{hd_v \times d_{mo\ de\ l}}$ is the matrix weight.

(2) *Contextual Feature Data.* The calculation of the output h_i for EL_i is as follows:

$$h_i = \text{LayerNorm}((\text{LayerNorm}(a_i + f)) + \text{Linear}(\text{LayerNorm}(a_i + f))), \quad (9)$$

where *Linear* is the fully connected operation, *LayerNorm* represents the layer normalization operation, when $i = 1$, f represents the vector matrix E_s , and when $i = 2, \dots, 6$, it represents the output of the previous encoding layer h_{i-1} .

3.2.3. Article-Granularity Feature Encoder. Usually, different TCM articles focus on different themes. The core sentence of a chapter is considered the carrier of the overall idea of the chapter, which exists in some parts of the chapter. The CNN is a feedforward network model structure based on artificial neural network. It has the advantages of local connection and weight sharing, which greatly reduces the number of parameters needed to learn in the network. Through multilayer nonlinear transformations, a CNN can learn the implicit features in a large-scale text [30]. So the AGFE based on CNN is designed to extract the core features h_a of a chapter-level text.

As shown in Figure 2(c), the AGFE consists of three one-dimensional convolutional layers, a maximum pooling layer, and a fully connected layer. Firstly, each one-dimensional convolution operation can obtain the features of the corresponding text fragment, and then after maximum pooling and fully connected operation with dropout, the final output of the AGFE h_a is obtained.

3.3. Seed of Entity Distribution Features. Different types of TCM texts cover different types of entities, and the distribution of entity label sequences reflects different patterns, which can enhance the effectiveness of entity label generation. The model proposed in this paper can obtain entity

tagging distribution features based on the annotated corpus, which we call the seed of entity distribution features.

We refer to the classification criteria of the 12 first-level categories given in the *General catalogue of Ancient Chinese Medicine books* and select three basic categories of Chinese medical texts, with the additional category of comprehensive Chinese medical texts as the research object, then analyze the entities contained in above 4 categories of texts to summarize 19 typical entity types in total as shown in Table 1.

As shown in Figure 3, the following are the construction steps of the seed for a certain type of TCM text.

Step 1: Using word embeddings of 19 typical entities to generate 19 initial vectors v_i , ($i = 1 \sim 19$) independent of text type.

Step2: Let f_i ($i = 1 \sim 19$) be the number of times entity type i occurs in the annotated corpus, and use f_i to calculate the probability distribution of each type of entity p_i , ($i = 1 \sim 19$). The calculated probabilities are then multiplied with the corresponding initial entity vectors v_i ($i = 1 \sim 19$) and spliced into a vector as the representation that reflects the distribution features of the entities in this corpus, which is used as the input to the Label Seed Feature Encoder (LSFE).

3.4. The Generator of the MT-CGAN. In Figure 4, the H represents text features and the A represents sentence and paragraph-grained text attention features, both generated by the MTFE. Red arrows indicate convolutional layers (including Convolution, Normalization, and activation). Green arrows indicate transposed convolutional layers (including Transpose Convolution, Normalize, and Activate).

3.4.1. The Generator Overview. The generator of the MT-CGAN consists of the MTFE and C-U-NET, and it is used to generate the sequence of entity labels corresponding to TCM texts. The C-U-NET is composed of the LSFE, the Multi-Granular Feature Fusion Layer (MGFFL), and the Feature Augmented Label Generation Decoder (FALG Decoder). The C-U-NET model is the main framework of the generator network. Overall, the encoding and decoding processes of the generator form a hierarchically symmetrical U-shaped structure as shown in Figure 4.

The encoder of the generator consists of the MTFE and LSFE of the C-U-NET. The role of the LSFE is to encode the features of seeds that reflect the distribution patterns of different categories of TCM entities and to output them hierarchically to the decoder, as shown in Figure 4. The decoder of the C-U-NET acts as the decoder of the generator. There is no data from other spaces introduced between the encoder and decoder of the traditional U-NET [31]. By adding a MGFFL in the middle of its encoder and decoder, we have improved the U-NET to form the C-U-NET.

3.4.2. The C-U-NET. The encoding process of the encoder in the C-U-NET adopts a downsampling network structure. Each encoding layer passes the features extracted from that layer to the corresponding decoding layer, as shown in

Figure 4. The decoding uses an upsampling network structure. The decoding process is enhanced by using feature information passed from the corresponding layer of the LSFE.

(1) *Label Seed Feature Encoder.* The LSFE is composed of 6 layers of CNN. There are a convolutional layer and a layer normalization and a ReLU activation function in each of the first five layers. The sixth layer consists of a convolutional layer and a layer normalization and a tanh activation function. The final layer uses tanh as the activation function to improve training efficiency and to preserve more fully the feature information of the labels.

(2) *FALG Decoder.* The role of the FALG Decoder is to decode feature sequences by convolution and transposed convolution, and it also contains 6 layers. The first layer is composed of a transposed convolutional layer, a layer normalization, and a ReLU activation function. Layers 2 to 5 consist of a transposed convolutional layer, a convolutional layer, two layer normalizations, and two ReLU activation functions. While the 6th layer has one normalization layer only and one of the activation functions has been replaced with tanh.

3.5. The Adversarial Optimization of the Model

3.5.1. The Discriminator. The discriminator is composed of 7 convolutional layers. A layer normalization and a LeakyReLU activation function are added after each layer of convolution, and the Sigmoid activation function is used in the last layer. The LeakyReLU activation function maximizes the retention of features contained in the data. The Sigmoid, as a saturated activation function, is placed at the last layer of the network to compress the range of real values of the model and improve the accuracy of the discriminator. To meet the requirements of CGAN, the conditional and primary inputs of the discriminator are the real label sequence feature L and the generated label sequence feature L' , respectively.

3.5.2. Confrontation Loss Function. To improve the generator accuracy of named entity label sequences, the generator and discriminator with different constraints as input are constructed. The CGAN converts the unsupervised generator model into a supervised or semisupervised generator model and converts the GAN loss probability representation into a conditional probability formula using conditions as restrictions. The discriminator optimizes the generator in terms of the truthfulness of the generated labels. The loss function is as in the following:

$$\min_G \max_D V(D, G) = \mathbb{E}_{L \sim P_{\text{data}}(L)} [\log D(L'|L)] + \mathbb{E}_{Y' \sim P_{Y'}(Y')} [\log (1 - D(G(A|Y')))], \quad (10)$$

where D is the discriminator, G is the generator, \mathbb{E} is the expected value of the sample, $\mathbb{E}_{L \sim P_{\text{data}}(L)}$ is the data sample from the real data, $\mathbb{E}_{Y' \sim P_{Y'}(Y')}$ is the sample from the training data. The loss function of the MT-CGAN is as follows:

TABLE 1: Typical types of entities contained in different types of TCM texts.

Type of literature	Typical entity types and corresponding labels
Canon of medicine	Cognitive method (RSFF), traditional Chinese physiology (ZYSL), traditional Chinese pathology (ZYBL), the principle of treatment (ZZ), method of treatment (ZF)
Medical cases	Chinese materia medica (ZY), symptoms (ZZ), pulse (MX), tongue (SX), formula (FJ), dosis (FJ)
Herbal	Drug property (YX), flavor of medicinals (YW), channel tropism (GJ), virtue (GX), dilantin (YM), symptoms (ZZ)
Comprehensive	Pathogeny (BY), disease (JB), syndrome (ZH), pulse (MX), tongue (SX), formula (FJ), Chinese materiamedica (ZY)

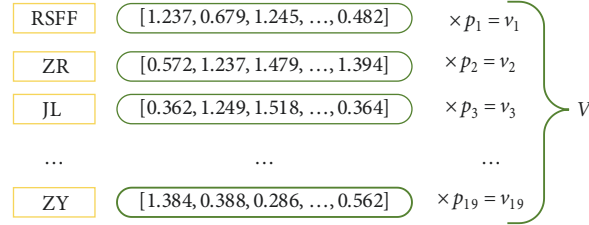


FIGURE 3: A schematic representation of entity distribution features.

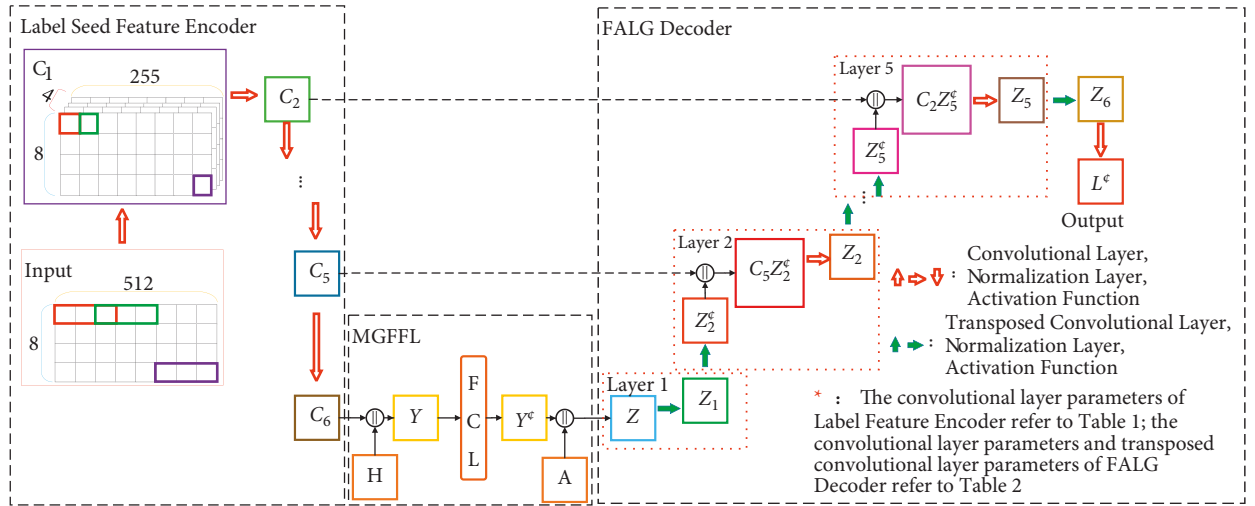


FIGURE 4: The structure diagram of the C-U-NET. FCL represents a fully connected layer.

$$\mathcal{L}(\tilde{T}) = \mathcal{L}_{\text{CrossEntropy}} \left(\min_G \max_D V(D, G) \right). \quad (11)$$

where \tilde{T} is the sequence of entity labels generated by the generator and $\mathcal{L}_{\text{CrossEntropy}}$ is the cross-entropy loss function.

3.5.3. Adversarial Training Process. As shown in Algorithm 1, the training of the MT-CGAN is divided into two steps. First, the generator is fixed and the discriminator is trained. Then, the discriminator is fixed and the generator is trained. In the training process, the two steps interact alternatively until reaching a stable state.

4. Experimental Analysis

4.1. Model Parameters and Datasets

Experimental parameters. The MT-CGAN is implemented using Pytorch 1.3.1 library. The Adam optimizer is utilized during the training procedure to solve the problem of gradient sparsity and gradient

oscillation in the field of natural language. The learning rate of the generator and discriminator is set to $1e-4$, the learning rate decay is 0.95, the dropout is 0.3, and the training epoch is set to 20. All models are trained and tested on a single NVIDIA RTX 3090.

Datasets. We use four books Shennong's Classic of Material Medical, Medical Cases of Famous Doctors in Different Periods of China, Miraculous Pivot, and Syndrome in TCM, as the original texts for NER annotation and annotate four types of the corpus as the experimental dataset. The scale of the annotated corpus for each book is shown in Table 2.

The *Shennong's Classic of Material Medical* is one of the four classics of Chinese medicine, was written in the Han Dynasty, and is the earliest known book on herbals, containing 365 species of herb-medicine. The *Medical Cases of Famous Doctors in Different Periods of China* is a collection of Chinese medical cases recorded by famous doctors from all periods of Chinese history. There are more than 18,000 medical cases in total, and each case introduces the patient's

TABLE 2: The scale of annotated corpus of the experimental dataset.

Name	Type	Text size (10,000 words)	Size of annotated corpus (10,000 words)
<i>Shennong’s classic of material medical</i>	Herbal	11.9	2.6
<i>Medical cases of famous doctors in different periods of China</i>	Medical cases	25	10
<i>Miraculous pivot</i>	Canon of medicine	7.9	7.9
<i>Syndrome in TCM</i>	General	564.4	28.6

illness involving the symptoms, pulse, tongue like, and the process of seeking medical treatment. The *Miraculous Pivot*, as an important part of the *Inner Canon of Huangdi*, is the earliest extant medical canon in China. The *Miraculous Pivot* is regarded as the most important classic that has summarized the theory of Jingluo (Channels and Collaterals) and the techniques of acupuncture before the Han Dynasty and contains 9 volumes and 81 chapters. The *Syndrome in TCM* is a comprehensive TCM book formed by collecting and sorting out ancient and modern documents related to medical writings, medical notes and medical cases.

4.2. Experiments and Results. For the TCM NER task, we chose several typical methods with which to compare the proposed approach. We selected two advanced methods: the BiLSTM-CRF architecture [4], BERT-BiLSTM-CRF architecture [13], and a variant of the BERT model architecture that can fuse multigranularity information of text (Roberta-c) [33]. These methods are described as follows.

BiLSTM-CRF [4] consists of the LSTM and CRF and is one of the most popular model architectures for handling NER tasks before pretraining models were proposed.

BERT-BiLSTM-CRF [13] consists of BERT, BiLSTM and CRF. As a representative of pretrained models, BERT has proven to be one of the most advanced models for representation learning in NLP [32]. The combination of BERT and BiLSTM-CRF models is one of the current popular architectures for NER task models.

Roberta-c [33] is a variant of the BERT model and outperforms the BERT model in Chinese NLP tasks. C denotes a word-character set in a self-attention module and it can integrate character and word information.

This section shows the comparative experiments and results analysis of the MGFF-GGAN with other three different types of baseline models on the datasets. The experimental results use Precision (P), Recall (R), and F1 score (F1) as evaluation metrics for the model evaluation. The performance on the test set of the different methods is presented in Tables 3 and 4.

According to the data presented in Tables 3 and 4, the overall performance of the proposed model is superior to that of the baseline models, except for on the *Medical Cases of Famous Doctors in Different Periods of China* dataset. This shows that our model shows overall advantages in various types of TCM texts, which is also the strongest advantage of our proposed model. Although the performance of the MT-CGAN does not improve significantly on both the

Shennong’s Classic of Material Medical and the *Medical Cases of Famous Doctors in Different Periods of China* datasets, even the effect of the F1 value on the *Medical Cases of Famous Doctors in Different Periods of China* dataset decreased by 0.08%, but the precision rate of the MT-CGAN is the highest. And as shown in Table 3 and Table 4, the P values of the proposed model are 90.26%, 90.36%, 77.73%, and 78.45%, respectively, on the four data sets, all of which are higher than other comparison models. It benefits from the fact that the MT-CGAN can extract the contextual semantic relations of text at different granularities.

We also observed that the performance of MT-CGAN on both *Miraculous Pivot* and *Syndrome in TCM* datasets was significantly improved. Especially on the *Syndrome in TCM*, the values of F1 and P of our proposed method are improved by 2.35~5.70% and 3.83~61.2%, respectively. The results indicate that the proposed model has good performance under the condition of small-scale annotated corpus and independent of external resources.

5. Discussion

5.1. Ablation Study. In this section, we investigate the ablation in different aspects of the MT-CGAN to better understand their relative importance.

We used the annotated corpus of the *Syndrome in TCM* as dataset to evaluate combination strategies at different granularities. As shown in Table 5, the F1 value is the highest when the features of the four granularities are used as input together. The F1 value is significantly improved when adding character-granularity features. The main reason is that the MT-CGAN uses GRU to extract the relative position features among characters from the character-granularity, which is used to constrain the order of entity label generation and plays a decisive role in the model.

We conducted a comparative experiment on whether the seed for different types of TCM texts can improve the effectiveness of named entity recognition. The results of the experiment, which are shown in Table 6, show that our method can enhance the effectiveness of entity label generation. When random entity label features are used as input, the F1 value is 75.16. In contrast, the F1 value increased by 2.19 when we used the seed of the corresponding text type as input.

5.2. Comparative Experiments with Different Forms of Main and Conditional Input. We verified the validity of our approach by experimentally comparing the differences with

Input: The fusion of text’s multi-granularity features Z .

Output: The Loss function value \mathcal{L} .

- (1) **For** number of training iterations **do**
- (2) **for** k steps **do**
- (3) Sample minibatch of mZ samples $\{A|Y'\}^1, \{A|Y'\}^2, \dots, \{A|Y'\}^m$ from Z prior $p_g(A|Y')$
- (4) Sample minibatch of m example $\{L'_1, L'_2, \dots, L'_m\}$ from data generating distribution $p_{\text{data}}(L')$
- (5) Update the discriminator by ascending its stochastic gradient:
- (6) $\nabla_{\theta_d} 1/m \sum_{i=1}^m [\log D(L^i|L) + \log(1 - D(G(A|Y')^i))]$
- (7) **end for**
- (8) Sample minibatch of mZ samples $\{A|Y'\}^1, \{A|Y'\}^2, \dots, \{A|Y'\}^m$ from Z prior $p_g(A|Y')$
- (9) Update the discriminator by ascending its stochastic gradient:
- (10) $\nabla_{\theta_d} 1/m \sum_{i=1}^m \log(1 - D(G(A|Y')^i))$
- (11) **end for**
- (12) The gradient-based updates can use any standard gradient-based learning rule. We used Adam in our experiments.

ALGORITHM 1: The training process. Minibatch stochastic gradient descent training of MT-CGAN. The number of steps to apply to the discriminator, k , is a hyperparameter. We used $k = 1$. Z is the fusion of text’s multigranularity features, seed of entity distribution features and attention data for different types of TCM texts.

TABLE 3: Experimental results under different models on the *Shennong’s classic of material medical* and *medical cases of famous doctors in different periods of China* datasets.

Model	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
	<i>Shennong’s classic of material medical</i>			<i>Medical cases of famous doctors in different periods of China</i>		
BiLSTM-CRF	86.32	87.67	86.99	87.88	89.43	88.65
BERT-BiLSTM-CRF	89.56	90.05	89.80	89.74	92.07	90.89
Roberta-c	89.78	90.15	89.96	90.12	90.59	90.35
MT-CGAN	90.26	89.78	90.02	90.36	91.28	90.81

TABLE 4: Experimental results under different models on the *Miraculous pivot* and *syndrome in TCM* datasets.

Model	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
	<i>Miraculous pivot</i>			<i>Syndrome in TCM</i>		
BiLSTM-CRF	68.76	66.95	67.85	72.38	70.95	71.65
BERT-BiLSTM-CRF	76.65	80.72	78.64	72.33	74.30	73.30
Roberta-c	76.98	80.47	78.68	74.62	75.39	75.00
MT-CGAN	77.73	79.69	78.69	78.45	76.28	77.35

TABLE 5: Experimental results of different combination strategies.

Combination strategies	F1 (%)
Sentence & paragraph	69.49
Sentence & paragraph & chapter	71.25
Character & sentence	74.86
Character & paragraph	73.17
Character & sentence & paragraph	76.48
Character & sentence & paragraph & chapter	77.35

TABLE 6: Comparison of using the seed of entity distribution features or not.

Label seed feature	F1 (%)
Random	75.16
Reflect entity distribution of different types of TCM text	77.35

CGAN on the annotated corpus dataset of the *Syndrome in TCM* from two aspects. (1) Whereas the main input to a CGAN generator is white noise, the main input to the MT-

TABLE 7: Experimental results of different main inputs.

Main input	F1 (%)
White noise	71.62
Attention data	77.35

TABLE 8: Experimental results of different condition inputs.

G	D	F1 (%)	Epoch
L	L	72.79	20–25
Y	L	77.35	15–20

L : the sequence of named entities labels for the corresponding text sequence. Y : the fusion of multigranularity text features and the seed of entity distribution features of the corresponding type of TCM text.

CGAN generator is the attention data of a text. (2) While the generator and discriminator of CGAN use the same data as conditional inputs, the MT-CGAN uses different data as conditional inputs, respectively.

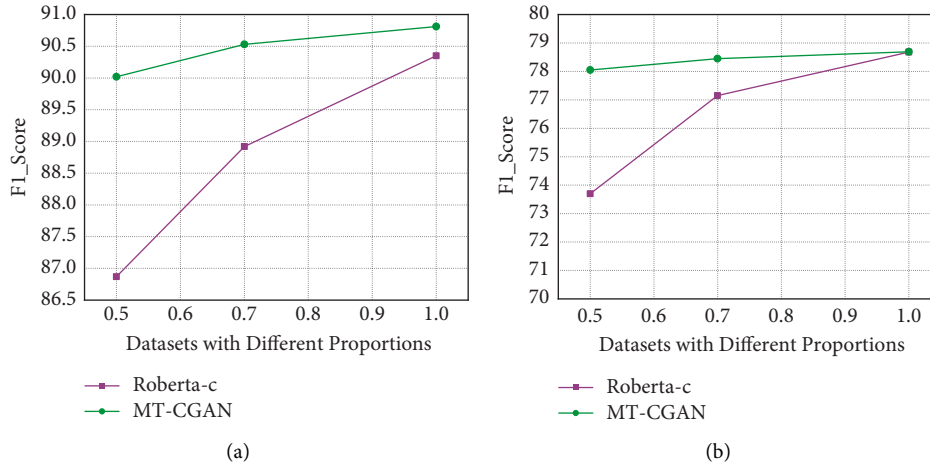


FIGURE 5: Comparison of experimental results on different corpus scale. (a) On *miraculous pivot* dataset (b) on *medical cases of famous doctors in different periods of China* dataset.

As shown in Table 7, the F1 value of the model was improved by 5.73% using the attention data of the text compared to white noise as the main input to the generator. The experimental results validate that the use of textual attention data can guide the model to generate more accurate labels.

As shown in Table 8, compared to the same conditional input, the F1 value of the MT-CGAN is improved by 4.56% and the number of training rounds for the model to reach the steady state is reduced by about 5 epochs. The experimental results demonstrate the method that uses the fused data Y as the conditional input in the model generator can improve the accuracy of the NER and speed up the convergence of the model.

5.3. The Comparison of Model Performance on Different Scales of Annotated Corpus. In this paper, we verified the performance of the MT-CGAN with small-scale samples by reducing the scale of the training dataset. We conducted comparison experiments with the Roberta-c model on different proportions of the original annotated datasets of the *Medical Cases of Famous Doctors in Different Periods of China* and *Miraculous Pivot*, respectively. As shown in Figure 5, we found that the MT-CGAN can maintain relatively stable F1 values, which indicates that the MT-CGAN has better recognition effects than other baseline models on the small-scale annotated corpus. The effectiveness of the MT-CGAN based on generative adversarial ideas for small-scale sample NER tasks was demonstrated.

6. Conclusions

In this paper, we propose the MT-CGAN to achieve TCM NER under the condition of a small-scale annotated corpus. To ensure that the model can learn multiple dimensions of TCM text information, MTFE is designed in MT-CGAN to extract text grammatical and semantic information from four granularity of words, sentences, paragraphs, and chapters. Moreover, to make MT-CGAN get accurate named entity labels when processing different types of TCM texts,

the seed of entity distribution features is introduced as the input of the model. We verify MT-CGAN performance on different types of TCM texts corpora, and the experimental results show that the model can achieve accurate NER tasks. The main advantage of the MT-CGAN is to deal with texts that contain more entity types and has sparse entity distribution, and have a more random and less regular linguistic form. Although our approach outperforms previous methods, our model still fails in some situations. For example, the R and F1 scores of our experimental results are lower than those of certain models, so we will make efforts to improve the R and F1 values of the model in the future.

Data Availability

Some data can be found within the article. Other data can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant no. 2018YFC1707410).

References

- [1] Q. Jia, D. Zhang, H. Y. Xu, and Y. Xie, "Extraction of traditional Chinese medicine entity: design of a novel span-level named entity recognition method with distant supervision," *JMIR Medical Informatics*, vol. 9, no. 6, p. e28219, 2021.
- [2] Z. G. L. Han Yuanbo Tao, "Domain named entity recognition combining GAN and BiLSTM-attention-CRF," *Journal of Computer Research and Development*, vol. 56, no. 9, pp. 1851–1858, 2018.

- [3] W. Wanliang and L. Zhuorong, "Advances in generative adversarial network," *Journal of Communication*, vol. 39, no. 2, pp. 135–148, 2018.
- [4] N. Deng, H. Fu, and X. Chen, "Named entity recognition of traditional Chinese medicine patents based on BiLSTM-CRF," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, Article ID 6696205, 12 pages, 2021.
- [5] L. Li, J. Zhao, L. Hou, Y. Zhai, J. Shi, and F. Cui, "An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records," *BMC Medical Informatics and Decision Making*, vol. 19, no. S5, p. 235, 2019 Dec 5.
- [6] Y. An, X. Xia, X. Chen, F.-X. Wu, and J. Wang, "Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF," *Artificial Intelligence in Medicine*, vol. 127, p. 102282, 2022.
- [7] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.
- [8] D. Mehta, A. Dwivedi, A. Patra, and M. Anand Kumar, "A transformer-based architecture for fake news classification," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 39, 2021.
- [9] M. Lei, H. Huang, and C. Feng, "Multi-granularity semantic representation model for relation extraction," *Neural Computing & Applications*, vol. 33, no. 12, pp. 6879–6889, 2021.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *Pre-training of Deep Bidirectional Transformers for Language Understanding* NAACL, 2019.
- [11] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the Bert model," *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 33–11, 2021.
- [12] M. N. Asim, M. U. Ghani, M. A. Ibrahim, W. Mahmood, A. Dengel, and S. Ahmed, "Benchmarking performance of machine and deep learning-based methodologies for urdu text document classification," *Neural Computing & Applications*, pp. 1–33, 2020.
- [13] Q. Qu, H. Kan, Y. Wu, and Y. Gao, "Named entity recognition of TCM text based on Bert model," in *Proceedings of the 2020 7th International Forum on Electrical Engineering and Automation*, pp. 652–655, IEEE, Hefei, China, September 2020.
- [14] M. Zhang, Z. Yang, C. Liu, and L. Fang, "Traditional Chinese medicine knowledge service based on semi-supervised BERT-BiLSTM-CRF model," in *Proceedings of the 2020 International Conference on Service Science (ICSS)*, pp. 64–69, Xining, China, August 2020.
- [15] Y. Xiong and B. PengXiangWongChenYanTang, "Leveraging Multi-source knowledge for Chinese clinical named entity recognition via relational graph convolutional network," *Journal of Biomedical Informatics*, vol. 128, no. 2022, p. 104035, 2022.
- [16] P. Xu, Q. Ruan, and W. Xu, "Text adversarial examples generation and defense based on reinforcement learning," *Tehnički Vjesnik*, vol. 28, pp. 1306–1314, 2021.
- [17] Y.. Cao, "Image and graph restoration dependent on generative adversarial network Algorithm," *Tehnički Vjesnik*, vol. 28, pp. 1820–1824, 2021.
- [18] M. Mirza and O. Simon, *Conditional Generative Adversarial Nets*, arXiv preprint arXiv:1411.1784, 2014.
- [19] J. Z. Feng, "Research on generative adversarial networks of named entity recognition," *Journal of Chinese Computer Systems*, vol. 40, no. 6, pp. 1192–1196, 2019.
- [20] Y. Zhang and J. Yang, "Chinese ner using lattice lstm," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 1554–1564, Long Papers).
- [21] J. Gao and F. ZhangCaoHuangLi, "Citation entity recognition method using multi-feature semantic fusion based on deep learning," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 6, 2022.
- [22] J. Sun, "Multi-strategy fusion for medical named entity recognition," in *Proceedings of the 2021 International Conference on Asian Language Processing (IALP)*, IEEE, 2021.
- [23] X. Li, H. Yan, X. Qiu, and X. Huang, "FLAT: Chinese NER using flat-lattice transformer," 6842, Online. Association for Computational Linguistics, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 6836.
- [24] O. Russakovsky and L. DengSuKrauseSatheeshMaHuang KarpthyKhoslaBernsteinBergFei-Fei, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Apr. 2015.
- [25] H. Tseng, P. Chang, G. Andrew, D. Jurafsky, and C. Manning, "A conditional random field word segmenter for Sighan bakeoff 2005," in *Proceedings. Of the Fourth SIGHAN Workshop on Chinese Language Processing*, 2005.
- [26] K. Cho, Bart van Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, *Learning phrase representations using rnn encoder-decoder for statistical machine translation* CoRR, abs/, vol. 1078, 2014.
- [27] X. Yu Rui, "A review of the general research on the concept of traditional Chinese Medicine," *Journal of Traditional Chinese Medicine*, vol. 58, no. 01, pp. 69–72, 2017.
- [28] A. Raganato and J. Tiedemann, "An analysis of encoder representations in transformer-based machine translation," in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 287–297, Association for Computational Linguistics.
- [29] Z. Dai, "Transformer-xl: attentive language models beyond a fixed-length context," arXiv preprint arXiv:1901.02860, 2019.
- [30] W. Tao and D. Chang, "News text classification based on an improved convolutional neural network," *Tehnički Vjesnik*, vol. 26, pp. 1400–1409, 2019.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceeding of the. Med. Image Comput. Comput.-Assisted Intervention*, pp. 234–241, 2015.
- [32] Y. Wang, "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*, vol. Volume 1, 2021.
- [33] Z. Liu and J. LuoZhengLiFuYuZhao, "TCMNER and PubMed: a novel Chinese character-level-based model and a dataset for TCM named entity recognition," *Journal of Healthcare Engineering*, vol. 2021, Article ID 3544281, 10 pages, 2021.