scientific reports

OPEN

Psychomedical named entity recognition method based on multi-level feature extraction and multi-granularity embedding fusion

Zixuan Liu[1], Guofang Zhang[1]✉ & Yanguang Shen[2]

Named Entity Recognition (NER) in psychomedicine is one of the key tasks in natural language processing in psychomedicine. It aims to identify and classify specialized terms in psychomedical texts and provide powerful support for downstream tasks. Psychological medicine texts are characterized by long paragraphs, complex sentences, and scattered knowledge. The current character-based psychomedicine NER model has single embedded information. It lacks structural and phonetic characterization information. Migrating NER models from the general purpose domain to the psychomedical domain are not effective in improving entity recognition accuracy. To solve this problem, we propose a NER method based on multi-level feature extraction and multi-granularity embedding fusion (MFME-NER), which aims to provide an innovative solution. First, three different granularities of embedding information, character embedding, radical embedding and pinyin embedding, are introduced to enrich the semantic representation of the input text. Second, the BERT model is improved. Merging the features of all Encoder layers inside the output. So that the BERT model has multi-layer feature extraction capability (MFE-BERT). The character embedding is pre-trained by MFE-BERT. And the BiLSTM model is utilized for the extraction of features at the character granularity. The features of radical embedding and pinyin embedding are extracted separately by the CNN model, and then feature fusion is performed. Finally, feature vectors at three granularities are integrated using a gated feed-forward neural network attention mechanism (GA-FNNAtention). The experimental results show that MFME-NER achieved 94.26% and 89.63% F1 Score in the self-constructed psychomedical dataset PsyDatase and CBLUE dataset, respectively. The proposed method surpasses the currently used evaluation metrics, thus substantiating its rationality and efficacy.This study can better contribute to the analysis of psychomedical data.

Keywords Named entity recognition, Psychological medicine, Multi-granularity fusion, GA-FNNAttention mechanism, MFE-BERT model

With the increasing stress of life and the importance of physical and mental health, mental health issues are gradually gaining attention. Especially in recent years, the number of people suffering from psychological problems such as social anxiety disorder, digital addiction and sleep disorders has been increasing. Psychological medicine shows an increasingly broad prospect for development through the study and treatment of human psychology, behaviour and various diseases related to physical health. As early as 1958, Carnegie Mellon professors Allen Newell and Herbert A. Simon has made this clear: Psychology is the behavioural basis of human problem-solving and also the root of the development of cognitive behavioural theories of AI. Facing the complexity of knowledge management in psychological medicine and the low efficiency of field information collection and collation, psychological science research and analysis put forward higher requirements for psychomedical data.

Named Entity Recognition (NER) is a key technique in natural language processing, which aims to extract information about entities with specific meanings from text and classify these entities into predefined types (e.g., names of people, organizations, places, etc.). In the field of psychological medicine, NER technology

[1]School of Cyber Security and Computer, Hebei University, Baoding 071000, China. [2]School of Information Science and Electrical Engineering, Hebei University of Engineering, Handan 056038, China. ✉email: xpzemail@163.com

can help researchers quickly and accurately extract information about named entities related to psychological medicine from text data, which helps to accurately construct and analyze the relationship between these entities and to deeply understand the knowledge in the field of psychological medicine. For example, key information such as the names of various psychological disorders and the names of therapeutic drugs can be extracted from psychomedical texts through NER technology. At the same time, NER technology also helps to promote information sharing and collaboration in the field of psychological medicine, providing strong support for precision treatment and personalized medicine. These properties make NER technology an important information processing and analysis tool in the field of psychological medicine.

In recent years, with the continuous iteration of neural network models, the approach of using large-scale pre-trained models and composite neural networks to deal with the named entity recognition task has a more powerful capability of latent feature capture. Representative models are Transformer encoding-based models such as BERT[1], ERNIE[2], and BiLSTM models fused with CRF[3], and so on. However, the direct application of state-of-the-art NER methods to psychomedical entity identification has not been effective. There are mainly two reasons.

(1) These models were originally designed for English, and compared with English, Chinese contains glyph information, pinyin information, and there is no clear demarcation between Chinese words, which poses a great challenge for Chinese named entity recognition[4,5].
(2) General NER models primarily rely on general domain text corpora for training and testing, whereas word distributions, as well as text passage lengths, differ significantly between psychomedical and general domain corpora.

In English, words are generally composed of roots, prefixes and suffixes, etc., and these constituents may be the same or similar in different words. For example, the root “psycho-” in the words “psychotherapy” and “psychology” is related to psychology, the prefix “anxi-” in the words “anxiety” and “anxiolytic” is related to anxiety, and the suffix “-phrenia” in the words “paraphrenia” and “schizophrenia” is related to mental illness. Chinese and English are two distinct languages with notable differences, the expressions and forms of Chinese are very different from English. The Chinese language not only has structural information similar to that of English affixes and roots, but each character also has unique pinyin information. In terms of structural information, the Chinese characters formed by the radical “疒” are mostly related to “medicine”, such as “疾” and “痴”. And the characters formed by the radical “心” are related to “emotion”, such as “虑” and “想”. For this reason, it is particularly important to utilize the structural features of Chinese characters, especially the information of radicals. Radicals are not only rich in semantic cues, e.g., they are closely related to medical concepts when describing organs, symptoms, and pathological states, but also provide additional morphological basis for the model to complement the traditional word vector representation. By incorporating information on radicals into the neural network model, we are able to capture subtle differences in the specialized vocabulary more precisely, thus improving the accuracy and robustness of psychomedical entity recognition. This improvement is especially prominent when dealing with rare and polysemous words, further validating the necessity and effectiveness of structural information in Psychomedical NER. In terms of pinyin information, the same Chinese character has different pronunciations. Each pronunciation represents a different meaning. The word “乐” means “happy” when it is pronounced as “lè” and “sound” when it is pronounced as “yuè”. The word “吓” means “fear” when it is pronounced as “xià”. When pronounced as “hè”, it means “dissatisfied”. The multi-granularity of information in Chinese characters gives them a strong characterization ability. Especially in the field of psychomedicine, where specialized vocabulary accounts for a large portion of the text, the structural and pinyin information is of great significance for semantic modeling.

By investigating the above problems, we propose a multi-level feature extraction and multi-granularity embedding fusion method for NER in the field of psychological medicine. Firstly, the pre-training model of character granularity is improved by merging all Encoder layer features inside the BERT model for output, to make it have multi-level feature extraction capability; secondly, structural information and pinyin information are introduced, and the CNN model is utilized to fuse the semantics of the vectors of the two granularities after local feature extraction respectively, and finally, the gated forward neural network attention mechanism is used to integrate the three granularities of the feature vectors, and the corresponding labels are outputted through the CRF.

In summary, The research presented in this paper makes the main contributions as follows:

(1) We propose an improved pre-training model MFE-BERT for improved multi-layer feature extraction to generate dynamic word vectors from text pre-processing. Based on the feature information of each layer of the fully-connected Transformer Encoder, the output vectors of each layer are finally merged and output, so that the psychomedical entity recognition word vectors are fused with the contextual semantic links, and the character vectors are endowed with richer word-level and semantic information.
(2) The features of Chinese characters themselves are fully utilized to enhance the semantic representation ability. Based on character embeding, structure information and pinyin information are fused. The problem of insufficient auxiliary features and ambiguous lexical boundaries caused by excessively specialized vocabulary in psychomedical texts is solved. Important paraphrase representations are obtained.
(3) We improve the traditional Attention mechanism. The forward neural network attention mechanism[6] is used to capture global word-level information to strengthen the word-level relationship of long text context. Meanwhile, a gating mechanism is used to perform multimodal fusion of semantic representations of three granularity feature vectors. GA-FNNAttention mechanism is formed. It can effectively solve the problem of

semantic dilution of long text passages. Avoid inconsistent labeling of the same psychomedical entities in different utterances.

(4) Based on the self-constructed psychomedical dataset and the publicly available biomedical dataset, the proposed model is compared and analyzed with other existing benchmark models. The experimental results demonstrate the rationality and effectiveness of the proposed model in this paper.

The rest of this thesis is organized as follows. Section "Related work" describes recent work on vertical domains as well as the multi-feature fusion approach to the NER task. Section "Methods" describes our proposed MFME-NER approach. Section "Experiments" describes the experimental details as well as the analysis of the experimental results. Section "Conclusions" concludes as well as provides an outlook on the next steps.

Related work

The traditional NER methods mainly include lexicon and rule-based pattern matching methods, statistical machine learning-based methods such as HMM, ME, SVM, and so on. But there are some problems with each of these approaches. The former relies on the language, style and domain of the text and is weakly shared. For different language system experts need to write specific rules The latter solves the problem that traditional entity recognition is difficult to generalize, but its time complexity is particularly high, which sometimes makes the cost of machine training unaffordable.

In recent years, research experts have been exploring deep learning and have made good progress. Some scholars have tried to combine deep learning with NER. Compared with traditional entity recognition, the entity recognition model integrated with deep learning can effectively reduce the training time of the model while improving the accuracy rate. In particular, it can be solved end-to-end by gradient dimensionality reduction. This makes it suitable for named entity recognition systems in more complex domains.

Pre-trained language model

The pre-trained language model can capture the semantic relationship between words and context information by learning a large amount of text data. It greatly improves the performance of NER. Deep learning-based pre-trained language models can break through the limitations of machine learning methods. It can also improve the accuracy and generalization ability.

Word2Vec[7] and Glove[8] are commonly used word embedding techniques in natural language processing. They can pre-train word vectors with a large amount of unlabeled data to improve the performance of NER models. ELMo[9] utilizes a deep BiLSTM model to form a contextual word embedding. The embedding contains different meanings of words in different contexts. GPT[10] uses a Transformer network structure, and is trained with large-scale natural language data. Thus, it is better adapted to NPL tasks. In 2018, Google team Jacob Devlin et al.[1] proposed the Bidirectional Encoder Representation from Transformers (BERT) language preprocessing model. The deep bidirectional representation of unlabeled text is pre-trained by jointly regulating the left and right contexts in all layers. This enables the model to more effectively capture the relationships between sentences. Baidu improved BERT and released ERNIE[2] language model. It can model the semantic relationships in massive text and enhance the semantic understanding of downstream models. To achieve the lightweight of pre-trained models, Albert[11] reduces the number of model parameters by sharing layer parameters and cross-layer parameter sharing on the basis of BERT. In terms of the training time and utilization of computational resources, it has better performance. The BioBERT pre-trained model proposed by Lee et al.[12] is capable of accurately determining the boundaries of biomedical entities to recognize longer entities by pre-training domain-specific languages on a large biomedical corpus. There are many other biomedical domain-specific pre-trained language models that are also trained on biomedical corpora, such as DMNER[13], VANER[14], BioBBC[15], TFEBioNER[16], and GERBERA[17]. The proposal and application of pre-trained language models have greatly benefited NER tasks.

Deep learning based NER

With the continuous and rapid development of computer neural networks and the increasing popularity of high performance computing, deep learning-based methods have shown significant advantages in the field of NER. Collobert et al.[18] used Convolutional Neural Networks (CNN) to extract features from the input sequence, and the experimental data showed the improvement in the performance of the model. Habibi et al.[19] utilized the BiLSTM-CRF model for the NER in the biomedical field. Thei r drug recognition results are still at the state of the art today. For fine-grained features, the DIE-CDK[20]approach effectively mitigates the problem of losing fine-grained information in different situations by fusing local features extracted in the target domain with the global class semantic knowledge obtained from the general knowledge domain. Guo et al.[21] based on BERT by fusing the BiLSTM and CRF model. A high-precision NER intelligent recognition method was constructed. Ren et al.[22] integrated the attention mechanism into the BERT-BiLSTM-CRF model, so that the model could effectively learn the context structural features. Guo et al.[23] proposed a Dual Semantic Correlation Alignment (DSCA) approach, which effectively supplements the missing semantic information in the target domain and improves the detection accuracy through context-dependent semantic alignment, and instance-level alignment through class-dependent semantic alignment. Mo et al.[24] redefined the CrossNER task as identifying the relationship between token pairs and processed the semantic contrast between source sentences, code-switched sentences, and target sentences through a multi-view contrastive learning framework to align representations between different languages. Hu et al.[25] proposed a novel deep medical NER method co_decision_NER (CDN) based on collaborative decision strategy (CDS), which can not only identify standard medical entities but also non-standard medical entities, effectively alleviating the serious out-of-vocabulary (OOV) problem faced by HQA services when performing medical NER tasks. Lou et al.[26] proposed a dictionary-based matching graph network (DMGN), which projects all possible dictionary-based entity combinations in the text onto a directed graph and uses a bidirectional GCN (BiGCN) to encode the forward and backward graphs. Compared with a simple masking approach, DMGN significantly improves the performance of biomedical named entity recognition. In addition, in recent years, patterns using span-based permutations have also achieved good results in NER tasks[27,28].

Deep learning-based NER reduces the difficulties associated with manual feature selection compared to traditional statistical machine learning methods. The extraction of text data is more adequate. It reflects outstanding advantages in entity recognition tasks.

NER for multi-feature fusion

Character vectors alone may suffer from insufficient information contained in the embedding layer. Li et al.[29] fused glyph and phonetic features, decomposed Chinese characters into Five-Stroke components to represent structural features, and proposed an improved phonetic system, thus solving the problem of character substitution in Chinese named entity recognition. Qiu et al.[30] combined a GeoBERT geological pre-training model with a variety of features (pinyin, radicals, and position vectors) and trained it with a BiLSTM-Attention model, and the experimental results show that the model outperforms ten benchmark models on the constructed dataset. Meng et al.[31] similarly introduced Chinese glyph features through the unique Tianzige-CNN through image classification as an auxiliary task. Xuan et al.[32] proposed the NER method for fused glyph networks. In addition to using a novel CNN for feature representation of the glyph information, the method can capture the graphical information and the interaction information between neighboring graphs by CGS-CNN. For the limited sample case in feature representation learning, the researchers proposed a discriminative prototype based on pairwise decoupled contrast learning (DP-DDCL) approach[33], which constructs more discriminative category prototypes by introducing domain knowledge consisting of CLIP and attribute knowledge graphs to obtain explicit and implicit semantics. Li et al.[34] proposed the FLAT model, based on the Transformer, to design a special positional encoding for the mesh structure that translates into a planar structure composed of spans. The recently proposed Three Level Feature Enhancement (THFE) approach[35] reduces the problem of feature loss by enhancing semantic information at all levels hierarchically through three spaces: super-resolution enhancement, semantic enhancement, and hierarchical enhancement, using sub-pixel convolution, sparse self-attention, channel attention, and spatial attention mechanisms. Wu et al.[36] proposed a CrossTransformer to combine character features with structural features. Li et al.[37] proposed a Chinese clinical named entity recognition (CNER) model based on multi-feature fusion and multi-scale local context enhancement, which enhances the semantic representation of text by combining pinyin, radicals, parts of speech, word boundaries, achieving excellent results on two benchmark datasets.

Although these methods have achieved good results in a variety of fields, when facing named entity recognition in the field of psychological medicine, the above NER methods do not take into account the specialization of the field of psychological medicine, especially for the characteristics of long paragraphs, complex sentences and scattered knowledge of psychological medicine texts, which may suffer from poor named entity recognition accuracy and missed detection. In order to more fully capture the diverse information embedded in the data and to solve the problem that single-level feature extraction may miss details or abstract information, we designed a multilevel feature extraction module. Based on the feature information of each layer of the fully-connected Transformer Encoder, this module combines the output vectors of each layer, extracts low-level features and high-level semantic features at different network levels, and gives the word vectors richer word-level and semantic information. In addition, considering the unique pinyin and structural features of Chinese characters, we further propose a multi-granularity embedding fusion method, which enables the model to retain fine-grained local information and take into account coarse-grained global features when constructing the final embedding vectors by fusing feature representations of different granularities. Such a design not only improves the richness of feature representation, but also enhances the discriminative ability of the model in complex scenarios, which verifies the effectiveness of the method in improving the performance of the model. To further enhance the adaptive ability of feature fusion and capture complex nonlinear features and long-distance dependencies in the input data, we introduce the GA-FNNAttention mechanism, which strengthens the word-level relationships by adjusting the semantic weights of the character-level features as well as the fusion features of pinyin and structural information, and at the same time, utilizes a gated multimodal fusion module to perform the feature fusion, reducing the possibility of introducing noise.

In summary, we propose a multi-level feature extraction and multi-granularity embedding fusion method for NER in the field of psychomedicine, and its innovativeness is mainly reflected as follows:

(1) In order to further enhance the capability of the model and to cope with the complex features in the field of psychomedicine, we propose a multi-granularity embedding fusion method.By extracting features at different network levels, our approach is able to capture not only fine-grained local information at lower levels, but also more abstract semantic information at higher levels. This hierarchical feature extraction approach enables the model to better capture semantic information of different representations, thus improving the recognition ability and accuracy of complex texts.

(2) Multi-granularity embedding fusion can effectively integrate feature representations of different granularities. By fusing fine-grained pinyin and structural embeddings with coarse-grained character embeddings, the model is able to take into account both the details and the whole when constructing the final embedding vectors, which enhances the model's discriminative ability in diverse scenarios. Compared with previous single-granularity approaches, this multi-granularity fusion design is able to capture the multi-granularity and multi-dimensional information in the text in a more comprehensive way, especially in scenarios such as the psychomedical field, where the information is complex and variable, and demonstrates stronger adaptability and robustness.

(3) The GA-FNNAttention mechanism introduces adaptive semantic weight adjustment, which is able to flexibly allocate attention among different feature dimensions, further enhancing the model's ability to model long-distance dependencies and complex nonlinear features. The gated multimodal fusion module, on the other hand, effectively reduces the noise problem that may occur during the multimodal information fusion process, enabling the model to handle complex data inputs more stably.

Methods

In this section, we will introduce the MFME-NER model in detail. The model structure is shown in Fig. 1.

Character granularity feature extraction

Character embedding

The entity information of long texts in psychomedicine usually consists of a large number of Chinese characters, which themselves have independent semantics and may exhibit different meanings in different contexts. The character-level-based modeling approach can avoid the errors caused by Chinese word splitting and improve the accurate recognition of entity boundaries. Characters can be directly converted to vectors using Word2Vec via a vector lookup table. They can also be converted to vectors with more semantic features using some pre-trained models. We chose the second way. In this approach, the text we enter is treated as a sequence of characters $s = \{c_1, c_2, \cdots, c_n\} \in V_c$, where V_c represents the set of characters. Characters are converted from input text to feature vectors by an improved pre-training model- MFE-BERT. The MEF-BERT model is described in detail below.

MFE-BERT pre-training model

We base it on the standard version of BERT released by Google, where the number of Transformer Encoder blocks is 12 layers containing 12 multi-headed self-attention. Since the Transformer Encoder adopts the form of feed-forward propagation of feature vectors, the semantic information contained in the process of unidirectional transmission decreases layer by layer, accompanied by each layer of Encoder has input the semantic information of the sequence of other tokens, then the feature information of the current token will be diluted, which may ultimately result in the problem of the semantic incompleteness of the output feature vectors.

The MFE-BERT model, based on the original full connection, outputs the information processed in the 11 Encode layers to the top layer, at this time the model has feature information of different abstraction strengths of the input sequence, splices the feature vectors with contextual semantic information in the 12 layers through the Concat function. The vectors are then mapped for dimensionality reduction by full concatenation. Finally, the model outputs a feature vector with deeper semantic information. The method ensures that everything from shallow lexical features to deep semantic features can be fully utilized, and captures contextual semantic relationships through character-level embedding to improve the richness and accuracy of semantic representation, especially when dealing with complex semantic features in psychological medicine. The structure of the improved BERT model is shown in Fig. 2.

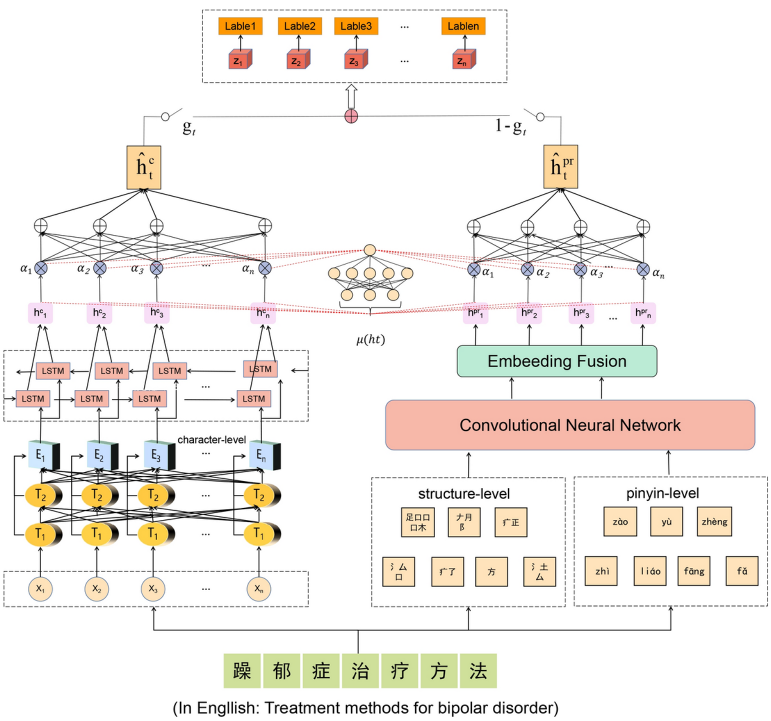

Fig. 1. The process of recognizing psychomedical entities using the MFME-NER model.

The specific expressions are as follows: The vector $E_n = \{e_1, e_2, \ldots, e_n\}$ enters the encoder layer after linear variation to obtain the query matrix Q, the contextual relationship matrix of characterization K and the content matrix V. The product of the Q-matrix and the K-transpose matrix corresponds to the degree of interconnectedness of the individual words after being calculated by the scaling factor $\sqrt{d_k}$with the softmax function. Then click on the content matrix V to get the attention score value:

$$Attention\left(Q, K, V\right) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Next, Concat splice the attention scores of each layer and dot-multiply the additional weight matrix W^O. At this point, the feature vectore_{ij} with contextual semantic information is obtained:

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{2}$$

Next, Concat splice the attention scores of each layer and dot-multiply the additional weight matrix W^O. At this point, the feature vectore_{ij} is obtained with contextual semantic information:

$$e_{ij} = Concat\left(head_1, \cdots, head_{12}\right)W^o \tag{3}$$

Finally, the feature vectors output from the 12-layer Encoder are concatenated by the Concat function, and the character embedding with multi-layer feature representation can be output:

$$\mathrm{ce}_i = Concat\left(e_{i1}, e_{i2}, \cdots, e_{i12}\right) \tag{4}$$

In order to make the dimension of the feature vector trained by the MFE-BERT model correspond to the downstream task dimension, we perform full link mapping dimensionality reduction on the vectors, where b_i is the position bias vector:

$$x_i = \tanh\left(ce_i + b_i\right) \tag{5}$$

MFE-BERT models character-level embeddings to improve semantic expressiveness. While traditional BERT adopts unidirectional transfer and uses the last layer of output, MFE-BERT makes character embeddings contain richer semantic information by incorporating features from multilayer Transformer encoders, and this improvement enhances the model's ability to understand psychomedical terminology and makes it better able to adapt to the complex expressions of the terminology specialized in this field.

Bidirectional LSTM model

LSTM utilizes gating units to alleviate the gradient explosion and gradient vanishing problems that arise during RNN model training[38]. Long-term memory is realized and sequence-dependent information can be obtained efficiently. LSTM state transmission direction is unidirectional from forward to backward, which can only obtain current as well as forward semantic features through vectors, and cannot utilize the following information at the same time, however, there is a strong correlation between the psychomedical text entities, and the state at the current moment not only depends on the input vectors at that moment, but also relates to the state at the next moment[39]. Therefore, BiLSTM models that can extract bidirectional semantic features are used for modeling.

Bidirectional long- and short-term memory networks can effectively utilize information in both directions of context simultaneously. Since psychomedical texts usually contain long sentences and complex syntactic structures, BiLSTM can effectively model character-level contextual relationships through bi-directional

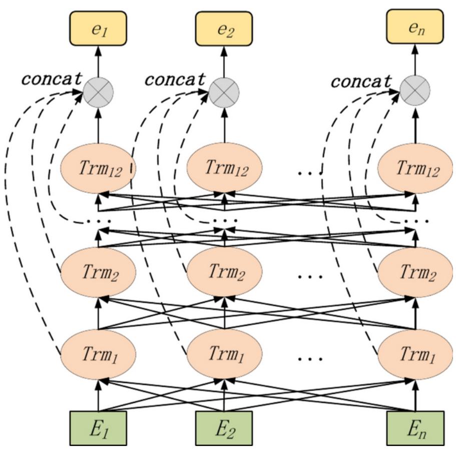

Fig. 2. Diagram of the MEF-BERT model.

information flow, capture long-range dependencies, and improve the accuracy of entity recognition. The basic idea is to present each forward sequence and backward sequence to two separate hidden states to capture the information of the two directions separately, and then connect the two hidden states as the final output. For the input feature embedding$\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_\mathrm{t}\}$, its corresponding implicit state sequence is$\mathrm{H} = \{\mathrm{h}_1, \mathrm{h}_2, \cdots, \mathrm{h}_\mathrm{t}\}$. The BiLSTM encoder generates two sequences of states at moment t in two directions, forward and backward$\left\{\overrightarrow{\mathrm{h}_1}, \overrightarrow{\mathrm{h}_2}, \cdots, \overrightarrow{\mathrm{h}_\mathrm{t}}\right\}and\left\{\overleftarrow{\mathrm{h}_1}, \overleftarrow{\mathrm{h}_2}, \cdots, \overleftarrow{\mathrm{h}_\mathrm{t}}\right\}$, which is the hidden state$H_\mathrm{t}$of BiLSTM after being connected. It is expressed as follows:

$$H_\mathrm{t} = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \tag{6}$$

For character-grained psychomedical entities, most of which consist of more than three consecutive Chinese characters, more contextual information needs to be taken into account. The output vectors of the BiLSTM model can capture bi-directional semantic features of psychomedical texts over longer distances, and can more accurately recognize psychomedical entities with character-level granularity.

Pinyin granularity and structural granularity feature extraction

Structural embedding

Chinese characters are hieroglyphics. Fixed structures have similar representations. For example, "疒" and "月" are mostly used in medicine. With the help of the hanyuzidian1 toolkit, Chinese characters can be decomposed into the corresponding structural parts, and the embedding representation corresponding to the structural sequences can be obtained by using Word2Vct training, which enables the model to learn the semantic features of the structural information in different contexts, and thus improves the model's capability of recognizing complex vocabulary and proper nouns. Through the learning of structural components, the model can recognize words containing specific radicals. In addition, structural embedding helps to solve the ambiguity problem caused by polysemous words. For example, the radical "心" is mostly related to emotion and mood in psychology, while "疒" is mostly related to physical diseases, and the model can better locate the semantics through structural features. The structural embeddings are described as follows:

$$x_i^{rd} = e^{rd}\left(f_{radical}\left(s_i\right)\right), i \in Z \cap i \in [0, n-1] \tag{7}$$

where $f_{radical}$ denotes a function that maps the input sequence to a structural sequence, e^{rd} denotes a vector lookup table corresponding to the structural sequence, and x_i^{rd} denotes a structural vector corresponding to the *i*th character in the sequence.

Pinyin embedding

In psychomedical entities, the pronunciation of different Chinese characters usually represents different semantics. Similarly, deep learning models can learn to map the semantics of Chinese characters with the help of their pinyin features. We obtain the corresponding pinyin sequences of the text through the pinyin(https://www.qqxiuzi.cn/zh/pinyin/) toolkit, and use Word2Vct training to obtain the corresponding embedding representations of the pinyin sequences.

Pinyin information, as a phonetic feature of Chinese characters, plays an important role in enhancing the semantic modeling of Chinese characters. Pinyin of Chinese characters not only helps to disambiguate homographs, but also provides additional contextual clues to the model. The introduction of pinyin information allows the model to capture the morphological features of Chinese characters while incorporating information at the phonetic level, making the model more accurate in recognizing polyphonic and polysemantic characters in different contexts. Pinyin information can enhance the phonetic information of homographs, and is able to differentiate different semantics corresponding to different pinyin, which reduces the confusion of the model when dealing with polyphonic words. In addition, pinyin embedding helps to model the context of phonological information, and the introduction of pinyin enables the model to understand sentences not only by the shape of the characters, but also by the pronunciation information of the characters, which is especially important for specific medical contexts (e.g., symptom, disease name, etc.). The specific formulation of pinyin embedding is as follows:The specific expressions are as follows:

$$x_i^{py} = e^{py}\left(f_{pinyin}\left(s_i\right)\right), i \in Z \cap i \in [0, n-1] \tag{8}$$

where f_{pinyin} denotes a function that maps the input sequence to a pinyin sequence. According to the serial number of the resulting pinyin sequence, the corresponding pinyin vector can be obtained from the vector lookup table e^{py} , where x_i^{py} is the pinyin vector corresponding to the *i*th character in the sequence.

CNN

When performing psychomedical NER tasks, it is especially important to capture local and global features of fine-grained embeddings, especially the semantic information of radicals and pinyin, which often contains many local features. For example, the radical "疒" is often associated with illness, while pinyin embedding helps to differentiate the different meanings of polyphonic characters, and also helps to disambiguate homographs. Using the convolutional and maximum pooling layers in the convolutional neural network model (CNN) can effectively extract the most representative part of the psychomedical text as a feature vector, and then through the multi-layer convolutional operation edge can be well learned to the global features of the textual text information, and its parallel computing capability can well deal with the multi-dimensional data task, and it can cope with the complex feature extraction process.

For the input sequence of structures $\mathrm{X}^{\mathrm{rd}}=\left\{x_1^{rd},x_2^{rd},\cdots,x_n^{rd}\right\}$, the convolutional layer is mainly responsible for the feature extraction of the sequence. The convolutional layer contains multiple feature extraction convolutional kernels, and the *d*-dimensional vectors are obtained after processing the structure vectors, and then the *s*-row *d*-dimensional vector matrix can be obtained. To reduce the number of model training parameters and improve the training efficiency, the neurons in the convolutional layer are connected to the neurons in the previous layer through the convolutional kernel, and then the ReLU activation function is used to obtain the output of multiple feature maps. The specific operation is shown below:

$$h_j^{l(rd)}=\mathrm{Re}LU\left(\sum\nolimits_{i=M_j}k_{ij}^{l(rd)}x_i^{l-1(rd)}+b_j^l\right) \tag{9}$$

$k_{ij}^{l(rd)}$ denotes the magnitude of the weights corresponding to the $x_i^{l-1(rd)}$ structure vectors, and b_j^l denotes the feature offset vector. The activation function is computed on its set to obtain $h_j^{l(rd)}$, the feature matrix of the convolutional layer. Then dimensionality reduction is performed by the maximum pooling method. The feature vectors are obtained:

$$h_{\max}^{\mathrm{rd}}=\max p\left(\left[\begin{array}{c}h_1^{l(rd)}\\ \vdots\\ h_j^{l(rd)}\end{array}\right]\right) \tag{10}$$

The feature extraction of pinyin embedding is the same as that of structural embedding, which is done by the CNN model. Similarly, we can get the feature vector of pinyin embedding:

$$h_{\max}^{\mathrm{pinyin}}=\max p\left(\left[\begin{array}{c}h_1^{l(pinyin)}\\ \vdots\\ h_j^{l(pinyin)}\end{array}\right]\right) \tag{11}$$

Structural and pinyin granularity embedding fusion

To minimize the loss of structural and pinyin feature information, we directly concatenate the two granularity embeddings obtained and input them into a fully connected layer for feature fusion:

$$x^{rd\&pinyin}=Concat\left(h^{rd},h^{pinyin}\right)W+b \tag{12}$$

where h^{rd} and h^{pinyin} are the structural embedding and pinyin embedding extracted by CNN features respectively, $Concat\left(\right)$ is the vector fusion function, and *b* is the position bias vector.

GA-FNNAttention

So far, there is still a problem for entity recognition of long texts in psychomedicine: character labels are not consistent across long texts. Although we believe that long short-term memory networks enable long-term memory and efficient access to sequence-dependent information, Urvashi Khandelwal et al.[40] demonstrated that the balance of the number of contextual characters that can be efficiently linked is 200, whereas our psychomedical research text paragraphs are much more than this number. In long paragraphs of text, the same psychomedical entities whose positions are far apart are probabilistically assigned different entity labels by the algorithm, and there is a problem of semantic dilution of long sequences. And the current research mainly focuses on the character-level self-attention mechanism, which cannot better match the case of multi-granularity feature fusion in this paper, which makes it difficult to improve the correct rate of the model. Therefore, this paper uses the gated forward neural network attention mechanism (GA-FNNAttention), while focusing on semantic representations in characters, pinyin, and structures.

GA-FNNAttention is a multimodal model using gated forward neural network attention mechanisms. GA-FNNAttention consists of two parts, firstly, the FNNAttention mechanism adjusts the semantic weights of character-level features as well as fusion features of pinyin and structural information to strengthen the word-level relationships, and secondly, the gated multimodal fusion module is utilized to perform feature fusion to reduce the possibility of introducing noise in multimodality. Finally, a feature vector based on multi-granularity feature fusion is obtained based on the combination of features with different weights. We use the GA-FNNAttention mechanism mainly to more effectively capture the complex nonlinear features and long-distance dependencies in the input data. Traditional feedforward neural networks may face the problems of insufficient information transfer and gradient vanishing when processing complex features. The introduction of the gating mechanism enables the network to adaptively regulate the information flow, thereby suppressing noise information while maintaining the transmission of key information. This mechanism not only improves the expressiveness of the model, but also enhances its stability and robustness in deep network structures. The GA-FNNAttention structure diagram is shown in Fig. 3.

(1) We effectively utilize the semantic information between contexts of long text passages through FNNAttention to obtain the global information of multi-granularity feature vectors, which are integrated with the current multi-feature vectors for computation, and allocate the attention mainly to the keywords to obtain the feature embeddings with more adequate semantics:

$$\hat{h}_t^c = \sum_{i=1}^{T} \frac{\exp\left(u\left(h_t^c\right)\right)}{\sum_{i=1}^{T} \exp\left(u\left(h_k^c\right)\right)} h_t^c \tag{13}$$

The character-level vector h_t^c with global feature information is obtained by computing the adaptive weighted average of the state sequence $\hat{h}_t^c$. Where $\mu\left(\right)$ is the forward neural network self-learning function that learns only through the state sequence h_t^c, which is a representation of the state sequence with bi-directional semantic information outputted through the BiLSTM layer. And k is the value obtained after the neural network self-learning.

The same operation allows us to obtain structural and pinyin granularity vectors with local feature information through the CNN layer $\hat{h}_t^{pr}$:

$$\hat{h}_t^{pr} = \sum_{i=1}^{T} \frac{\exp\left(u\left(h_t^{pr}\right)\right)}{\sum_{i=1}^{T} \exp\left(u\left(h_k^{pr}\right)\right)} h_t^{pr} \tag{14}$$

(2) We utilize a gating mechanism to fuse character features, structural and pinyin features, which determines the weights from the character feature vectors and structural pinyin feature vectors in the neural network through gating:

$$h_{\hat{h}_t^c} = \tanh\left(W_{\hat{h}_t^c} \hat{h}_t^c + b_{\hat{h}_t^c}\right) \tag{15}$$

$$h_{\hat{h}_t^{pr}} = \tanh\left(W_{\hat{h}_t^{pr}} \hat{h}_t^{pr} + b_{\hat{h}_t^{pr}}\right) \tag{16}$$

$$g_t = \sigma\left(W_{g_t}\left(\hat{h}_t^{pr} \oplus \hat{h}_t^c\right)\right) \tag{17}$$

$$z_t = g_t \hat{h}_t^c + \left(1 - g_t\right) \hat{h}_t^{pr} \tag{18}$$

$W_{\hat{h}_t^c}$, $W_{\hat{h}_t^{pr}}$ and W_{g_t} denote the weight matrices respectively. $b_{\hat{h}_t^c}$, $b_{\hat{h}_t^{pr}}$ denote the bias vectors corresponding to character granularity, structure and pinyin granularity. $\sigma\left(\right)$ is the logistic sigmoid activation function. $\oplus$ denotes the splicing of the three granularity features. g_t is the gate switch. z_t is the multi-granularity fusion of character embedding, structure and pinyin embedding features. It is also the input vector of the CRF conditional random field.

CRF label prediction loss function optimization

CRF is a classical discriminative model developed based on the undirected graph model. Compared with the traditional model that handles the task of independent sequence labeling for each character label, the CRF layer can well constrain the dependencies between labels. An optimal sequence is obtained by modeling the sequence labels.

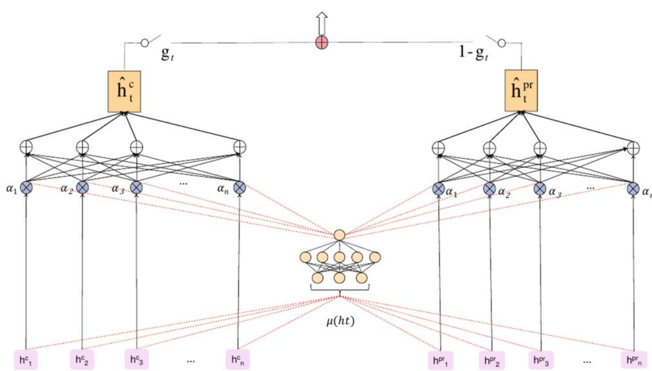

Fig. 3. Diagram of the GA-FNNAttention model.

To be able to avoid overfitting in the final result of the entity recognition model, the CRF loss function is optimized to incorporate the penalty mechanism:

$$L = -\left(S_t - \log\left(e^{S_1} + e^{S_2} + \cdots + e^{S_N}\right)\right) + \alpha \left\|\theta\right\|_2^2 \tag{19}$$

where S_t is the score of the true labeled path in the sequence and $P_{\text{total}} = \left(e^{S_1} + e^{S_2} + \cdots + e^{S_N}\right)$ is the total score of all predicted labeled paths. When training a CRF model, usually our goal is to minimize the loss function, hence the addition of a minus sign. θ is the trainable parameter in the entity recognition model. α is the hyperparameter determined by the custom cross-validation method as a way to penalize the parameters in the model can reduce the generation of overfitting.

Experiments

We conduct a large number of experiments on a self-constructed psychomedical dataset and a publicly available Chinese medical dataset, respectively. The recognition performance of the MFME-NER model is evaluated by experimentally comparing it with the current mainstream NER model. In this section we focus on the experimental dataset, evaluation metrics, experimental environment, parameter settings, experiments and corresponding analysis.

Datasets

Psychological medicine dataset

There is a lack of standard-type labeled data resources in the field of psychological medicine, the dataset is batch crawled from 26 psychological medicine websites such as YouMental, YouLaiDoctor, and Medical Encyclopedia by using regular expressions and XPath to extract the relevant text data, preprocessed and cleaned the psychological medicine text, and finally obtained the psychological medicine text dataset containing 3927 entities by manual labeling. The categories and quantities are as follows: the number of "alternate name" is 77, the number of "pathogenic site" is 166, the number of "symptom" is 924, the number of "check" is 313, the number of "department" is 385, the number of "disease" is 627, and the number of "susceptible crowd" is 170.

CBLUE open data set of biomedical texts

To validate the effectiveness of the proposed model on public dataset for psychomedical entity recognition, a part of Alibaba's publicly available CBLUE biomedical text dataset applicable to the corresponding domain without entity nesting is used for testing. Its predefined categories are the same as the self-constructed psychomedical dataset, and the corpus contains 2755 sentences for the training set, 936 sentences for the validation set, and 927 sentences for the test set. The statistics of PsyDataset and CBLUE dataset are shown in Table 1.

To expand richer categories in subsequent studies, the psychomedical entity labeling adopts BIO's sign system to label the collected entities one by one, where B (Begin) represents the beginning of a psychomedical entity, I (Intermediate) represents the middle portion of a psychomedical entity, and O (other) represents the other portion. The experiments were labeled with seven entity types, including psychological disease, susceptible crowd, symptom, alternate name, pathogenic site, check, and department. And the corresponding labeling examples are shown in Table 2.

To facilitate the intuitive display of the content and structure of the dataset used in this article, we will show some data examples in PsyDataset in Table 3. (The meaning of the Chinese characters in Table 3 is "Half a month ago, the patient developed personality changes, became irritable and easily angered, had no hallucinations or delusions of persecution, could not take care of himself in basic daily life, and had visited the neurology clinic.".)

Evaluation metrics

Precision, Recall and F1 Score are commonly used as evaluation metrics for psychomedical entity recognition performance, and the application of these metrics in psychomedical NER tasks helps to assess the model's ability to recognize different entities. In this paper, we mainly use the F1 Score as the performance index judgment, and the index calculation formula is shown as follows:

$$precision = \frac{TP}{TP+FP} \times 100\% \tag{20}$$

$$recall = \frac{TP}{TP+FN} \times 100\% \tag{21}$$

Dataset	Type	Train	Dev	Test
PsyDataset	Character	60.23k	16.59k	17.35k
	Entity	2.31k	0.79k	0.83k
CBLUE	Character	105.21k	45.16k	43.81k
	Entity	2.93k	1.36k	1.32k

Table 1. Statistics of the datasets.

Entity	Label of beginning position	Label of intermediate position	Other
Disease	B-disease	I-disease	–
Susceptible crowd	B-susceptibleCrowd	I-susceptibleCrowd	–
Symptom	B-symptom	I-symptom	–
Alternate name	B-alternateName	I-alternateName	–
Pathogenic site	B-pathogenicSite	I-pathogenicSite	–
Check	B-check	I-check	–
Department	B-department	I-department	–
None	–	–	O

Table 2. Representation of the dataset annotations.

Psychomedical-label			
半 O	、 O	, O	曾 O
月 O	易 B-symptom	日 O	到 O
前 O	怒 I-symptom	常 O	神 B-department
出 O	, O	基 O	经 I-department
现 O	无 B-symptom	本 O	内 I-department
性 O	幻 I-symptom	生 O	科 I-department
格 O	觉 I-symptom	活 O	门 I-department
改 O	和 O	不 O	诊 I-department
变 O	被 B-symptom	能 O	就 O
, O	害 I-symptom	自 O	诊 O
暴 B-symptom	妄 I-symptom	理 O	° O
躁 I-symptom	想 I-symptom	, O	

Table 3. Representation of the dataset annotations.

Environment	Hardware configuration
System	Window10
CPU	Intel Core i7-10700 CPU
GPU	NVIDIA GeForce RTX 3070(8G)
RAM	16GB
TensorFlow	1.15
Python	3.7

Table 4. Experimental environment.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{22}$$

where TP denotes true cases, FP denotes false positive cases, and FN denotes false negative cases.

Experimental environment

Table 4 shows the specific environmental configurations of all experiments in the Psychological Medicine NER study. This study is based on stable hardware facilities and software framework construction, the specific configuration is as follows: the experiment is run under the Windows 10 operating system environment, the hardware part of the Intel Core i7-10700 CPU, whose multi-core design provides efficient parallel computing capability for the model training; the graphics processor is selected NVIDIA GeForce RTX 3070 (8G The graphics processor is NVIDIA GeForce RTX 3070 (8G memory), which significantly improves the computing efficiency of deep learning tasks with its powerful CUDA acceleration; the system memory is 16GB, which ensures the fast loading and processing of large-scale text data. At the software level, the experiment relies on the TensorFlow 1.15 deep learning framework to build the model; the programming language is Python 3.7, which combines with related scientific computing libraries (e.g., NumPy, Keras, etc.) to achieve efficient development and optimization of the model.

Parameter	Value
Character_dim	768
Pinyiun_dim	200
Structural_dim	200
Max_length	300
Kernel_size	3
Dropout	0.3
Batch_size	32
Epoch	30
Learn_rate	1e−4
Optimizer	Adam
Activation_function	ReLU

Table 5. Experimental environment.

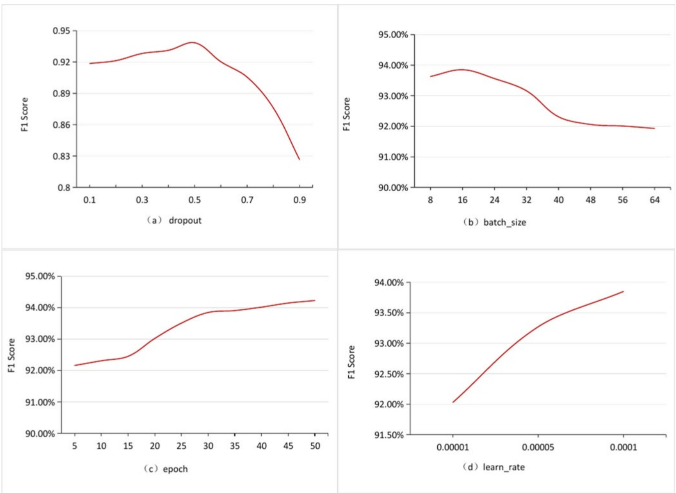

Fig. 4. Effect of different hyperparameters on model performance. (**a**) shows the trend of F1 Score under different dropouts. (**b**) shows the trend of F1 Score under different batch sizes. (**c**) shows the trend of F1 Score under different epochs (**d**) shows the trend of F1 Score under different learn rates.

HyperParameter of the psychomedical entity recognition model

The main hyperparameter settings of the MFME-NER model are shown in Table 5.

In this paper, the character embedding dimension is set to 768 to ensure that the vector has rich semantic information; the dimensions of pinyin embedding and structure embedding are both set to 200 to achieve a balance between representation ability and computational cost, avoid information redundancy, and are suitable for extracting local features of structural information and enhancing the structural modeling ability of text. In view of the long input sequence of psychological medical text, the maximum sentence length is set to 300. In addition, through experimental testing of hyperparameters such as dropout rate, batch size, and number of iterations on the psychological medical dataset, the following parameter configurations are determined: the dropout rate (droupout) is set to 0.3 to prevent overfitting and ensure the generalization ability of the model (as shown in Fig. 4a); the batch size is set to 32 (as shown in Fig. 4b); the number of iterations is set to 30 (as shown in Fig. 4c); the learning rate is 1e-4 to control the learning speed of the model, avoid gradient explosion or disappearance, and ensure stable convergence (as shown in Fig. 4d). In addition, the convolution kernel size (Kernel_size) is set to 3 to better adapt to the local dependency feature extraction of the Chinese NER task. The optimizer selects Adam to balance the convergence speed and generalization ability, while the activation function adopts ReLU to alleviate the gradient vanishing problem and enhance the nonlinear expression ability of the model.

For the characteristics of longer input sequences in the psychomedical text dataset, the maximum length of the sentence is set to 300, and the parameters of the model are determined through experimental testing of hyperparameters such as dropout rate, batch size, epoch number and learning rate in the psychomedical dataset. The specific parameters for entity recognition are: Adam is chosen as the optimizer, the learning rate is 1e-4, and the activation function is ReLU to prevent the overfitting phenomenon during model training. The introduction dropout is set to 0.3 (as in Fig. 4a), the batch size" is set to 32 (as in Fig. 4b), the number of epoch is set to 30 (as in Fig. 4c), and the learning rate is 1e-4 (as in Fig. 4d).

In Fig. 4a, the dropout is set to prevent the model from overfitting, and the optimal performance of the model is achieved at dropout = 0.5, and the model performance is only second to the optimal at dropout = 0.4. When dropout<0.3, the model performance is decreased, which is due to the overfitting phenomenon caused by too low a dropout rate. The model performance is too low when dropout>0.6, which is due to the high dropout rate that causes the model network to converge slower and less stable, and unable to obtain sufficient semantics. In Fig. 4b, the model performance is optimal when the batch size is around 16, and it is difficult to achieve convergence after decreasing the model batch size, so the model performance decreases. Blindly increasing the model batch size can reduce the number of training times, but it will cause the lack of memory capacity and reduce the accuracy rate. In Fig. 4c, the model F1 Score increases with the number of iterations, and when the number of iterations is greater than 30, the model F1 Score increases slowly and the efficiency decreases, so the number of model iterations is set to 30. In Fig. 4d, the three indicators are highest when lr = 1e - 4; when lr>1e-4, the parameter update is too fast, which leads to the model can not be converged in the training process; when lr<5e -4, the learning rate is too small, and the model only finds the local optimal solution rather than the global optimal solution during the optimization process, so the performance is not optimal, and the overall trend of change can be seen from the figure is increasing.

Experiments of the psychomedical entity recognition model

Baseline comparison experiment

To verify the feasibility of the model in the task of psychomedical entity recognition, we conducted comparative experiments of this model on the self-constructed professional psychomedical dataset PsyDataset and the CBLUE biomedical text open dataset, respectively. The experimental results are shown in Table 6. Among them, CNN-CRF, BiLSTM-CRF, BERT-BiLSTM-CRF and BERT-BiLSTM-ATT-CRF are a few classical NER models, which utilize the context modeling of composite neural networks to enhance the semantic comprehension ability of the model; ALBERT strengthens the performance of NER through parameter optimization strategies; Pure and PL-Marker are both used for NER tasks based on segment alignment, the latter is improved by packing the span grouping; FLAT transforms the grid structure into a planar structure consisting of spans; Soft-Lexicon, FGN, Glyce 、 MECT 、 MFE-NER and MFDE models improve the recognition by introducing feature information such as word lexicon, position vector, Chinese character radicals, structure, and glyphs, and improve the recognition accuracy, and have made important contributions in the field of Chinese NER.

The experimental data show that our model achieves optimal performance metrics on both the self-built PsyDataset psychomedical heart dataset and the publicly available CBLUE medical dataset, with F1 Scores of 94.26% and 89.63%, respectively.

Among several classical models, the CNN-CRF and BiLSTM-CRF models perform mediocrely, which is because these two models are not pre-trained on the text and lack the ability of semantic representation and contextual understanding. BERT-BiLSTM-ATT-CRF applies the attention mechanism on top of BERT-BiLSTM-CRF to help the model capture the long-distance dependencies, which strengthens the model's ability to understand and model contextual relationships, and achieves good results in the PsyDataset and CBLUE datasets, with the F1 Score reaching 93.43% and 88.89%, respectively, but there is still a slight gap from the optimum. In addition, comparing our model with the newly researched PL-Markerl NER model, our model improves the F1 Score by 1.41% and 0.69% in the two datasets respectively, which may beacuse that PL-Markerl is still essentially a span entity recognition method, which performs well in the general-purpose domain, but the long text of psychomedicine usually contains more complex contexts, so there are some limitations in dealing with the features of psychomedical long texts.

	PsyDataset			CBLUE		
Model	P	R	F1	P	R	F1
CNN-CRF[19]	87.19%	86.95%	87.06%	81.56%	81.33%	81.45%
BiLSTM-CRF[20]	88.29%	86.65%	87.46%	82.93%	81.96%	82.16%
BERT-BiLSTM-CRF[21]	94.36%	91.61%	92.96%	89.38%	87.26%	88.31%
BERT-BiLSTM-ATT-CRF[22]	94.89%	92.02%	93.43%	89.65%	88.16%	88.89%
ALBERT[11]	92.36%	93.02%	92.67%	85.91%	86.55%	86.23%
PURE(BERT)[27]	92.95%	90.89%	91.90%	88.90%	87.65%	88.27%
PL-Marker(BERT)[28]	93.22%	92.48%	92.85%	89.32%	88.61%	88.95%
FLAT[34]	93.43%	92.12%	92.77%	87.68%	87.02%	87.34%
Soft-Lexicon[22]	93.11%	91.86%	92.48%	88.34%	87.91%	88.12%
FGN[32]	94.26%	92.27%	93.25%	89.06%	88.32%	88.69%
Glyce[31]	94.25%	93.39%	93.81%	90.23%	88.12%	89.16%
MECT[36]	95.04%	93.25%	94.13%	90.03%	88.56%	89.28%
MFE-NER[29]	93.87%	92.56%	93.21%	88.12%	86.92%	87.51%
MFDE[30]	94.76%	93.31%	94.03%	90.17%	89.06%	89.61%
MFME-NER(Our model)	95.25%	93.29%	94.26%	90.36%	88.92%	89.63%

Table 6. Comparison results of entity recognition for different datasets.

Soft-Lexicon constructs soft lexicon features and introduces them into the representation of all characters, and its recognition performance is average, which we speculate is due to the fact that medical words in the dataset are mostly abbreviated, and the model can easily misrecognize them after introduction.Glyce and FGN both incorporate glyphic features on top of character embedding. Glyce utilizes the unique Tianzige-CNN as an auxiliary task through image classification, and performs slightly better than FGN, achieving F1 Score of 93.81% and 89.16% in the two datasets respectively.MECT, a cutting-edge result of mainstream multi-feature fusion NER, introduces structural and lexical information, and achieves sub-optimal results in PsyDataset. In both datasets, the Precision of MFME-NER reaches 95.25%, which is higher than the 94.76% of MFDE and the 93.87% of MFE-NER, indicating that MFME-NER performs more superior in reducing misidentification and provides more accurate entity prediction. This is attributed to the multi-layer feature extraction and multi-granularity embedding fusion strategy adopted by MFME-NER, which enhances the model's ability to differentiate medical terms, and the FNNAttention mechanism applied to better learn the semantic information between the contexts of long text passages. In addition, MFDE performs well in Recall, probably due to the fact that MFDE incorporates additional position vectors in the feature modeling process, which enables the model to better learn the semantic boundaries of entities when processing text, thus improving the recall rate.

Compared to other multi-feature fusion models, our model is designed with multi-scale feature extraction and deep feature fusion modules based on the characteristics of psychomedical texts, which are able to capture local details and global semantics at different levels, enabling the model to exhibit higher robustness and accuracy when processing long texts. Meanwhile, character, pinyin and structural embedding are organically integrated, and the gated feed-forward neural network attention mechanism is used to realize dynamic weighted fusion, so as to fully capture text features at different granularities. This enables the model to comprehensively and finely express the complex semantic and structural information in long texts of psychological medicine, and significantly improves the overall performance. In summary, the MFME-NER model performs well in the psychomedical NER task, which can well extract the implicit features of sequences and effectively recognize the corresponding entities in the field of psychomedicine.

Convergence analysis of the NER models

To assess the stability of different NER models during training, this study systematically analyzes the dynamics of F1 values of mainstream models over 50 training cycles (Epoch) based on PsyDataset and CBLUE datasets. The effects of different model architectures on the training dynamics are revealed through the two dimensions of convergence speed and stability. The convergence effects of each model in the two datasets are shown in Figs. 5 and 6.

In the PsyDataset : CNN-CRF is limited by the local feature extraction mechanism and has the slowest convergence speed, finally converging to an F1 value of 86.98%, with a standard deviation of 0.41 at the end of the segment, which is less stable.The BERT-BiLSTM-CRF model, by virtue of the ability of pre-training semantic characterization, has a rapid increase in the F1 value within the first 15 Epochs from 65.07 to 90.23%, and then enters a fine-tuning phase, finally stabilizing at 93.02% in the 34th Epoch.MECT combines structural and lexical features, and converges at a speed close to that of the BERT class of models, but due to the diversity of psychomedical terminology, the F1 value fluctuates around 94.26% in the later stages.Glyce relies on glyph features, and the F1 value within the first 30 Epochs rises rapidly to 93.81%, but subsequent performance stagnation due to feature homogeneity, the F1 value in the 50th epoch is 93.69%, which does not break through the upper limit of the pre-trained model.MFME-NER (the present model), through multi-level feature fusion, the model achieves an F1 value of 93.03% in the 26th Epoch, and the speed of convergence has a small increase compared with the other models. The final F1 value is 94.26%, and the standard deviation of the last 10 Epochs is only 0.09, with the smallest fluctuation among the compared models.

In the CBLUE dataset: FGN, a Grapheme Fusion model based on Graph Neural Networks, shows a small oscillation (standard deviation 0.23) in the late convergence period (Epoch 40-50), probably due to the weaker grapheme-semantic associations in biomedical texts than in the field of psychomedical sciences.ALBERT, which reduces computational costs through parameter sharing, has a final F1 value of 86.23%, with fluctuations in the end period (standard deviation 0.19) reflecting its limited characterization ability.FLAT, a Transformer

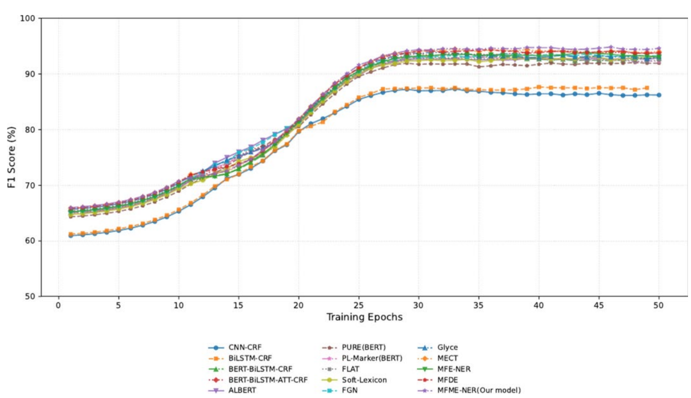

Fig. 5. Convergence analysis on PsyDataset.

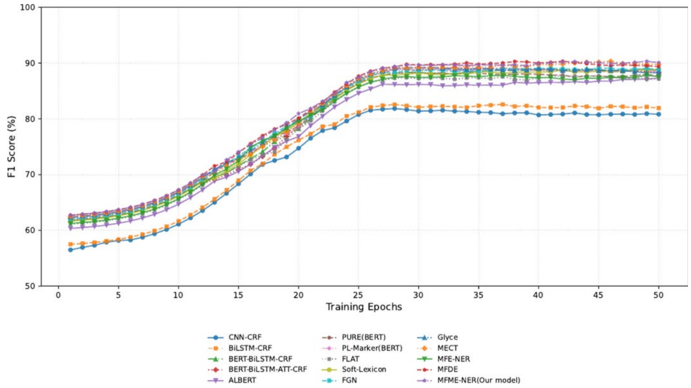

Fig. 6. Convergence analysis on CBLUE dataset.

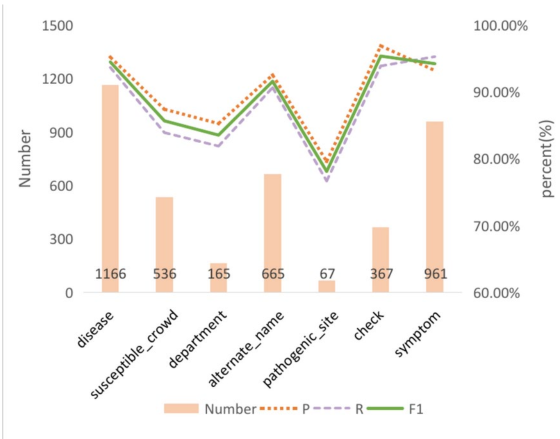

Fig. 7. Results of the number of entities recognized and performance metrics for the psychomedical dataset.

model based on a flat structure, has a significant difference in convergence speed in CBLUE (38 Epochs) versus PsyDataset (30 Epochs), highlighting the importance of domain adaptation for structural design.MFME-NER (the present model) has an outstanding cross-domain generalization ability, with a significant difference in F1 value between the first F1 value within 30 Epochs improves from 62.74 to 89.64%, the convergence speed is comparable to that of PsyDataset (difference<5%), and the standard deviation of the end segment is 0.11, which verifies the compatibility of the multi-granularity features to the domain differences.

Results of entity-level experiments

The MFME-NER model performs performance test experiments on 7 types of 3927 labeled entities in the psychological medical dataset. The experimental results are shown in Fig. 7. It can be seen that the F1 Score and accuracy rate of entities such as "disease", "symptom", "alternate_name", and "check" are above 90%, among which the effect on entity recognition of "check" is the best, with an accuracy rate of 96.92%, a recall rate of 93.91%, and an F1 Score of 95.39%. This is because the expression in the data set is fixed, and the model can easily learn the description of its rules. The experimental F1 Score of the "susceptible_crowd" is 85.69%, and the performance is relatively poor. This is because the expression of some of the entities of the "susceptible_crowd" contains "symptom", such as the expression "commonly seen in men with anxiety", which recognizes "anxiety" as a symptom, which is also the reason for the high recall rate of the entity of "symptom". The effect of entity recognition of "pathogenic_site" and "department" is poor. The main reason is that the number of samples of these two entities in the psychological medical dataset is small, and the training effect cannot be achieved.

Impact of the number of MFE-BERT feature extraction layers

To study the effect of extracting different numbers of Encoder layer features on the entity recognition results of the MEF-BERT pre-training model, six different sets of extracted layers L = [2,4,6,8,10,12] are set up to splice Layer 6 with Layer 12; Layer 1, Layer 5, Layer 9, Layer 12, and so on, with the number of intervals between each layer as similar as possible to maintain the uniformity of the feature distribution. Experiments were conducted on the self-constructed psychomedical dataset using the MFME-NER model with other parameters unchanged, and the experimental results are shown in Table 7.

Model	Level	P	R	F1
MEF-BERT-2	6,12	94.61%	92.58%	93.59%
MEF-BERT-4	1,5,9,12	94.69%	92.75%	93.71%
MEF-BERT-6	1,3,5,7,9,12	94.82%	92.91%	93.85%
MEF-BERT-8	1,3,5,6,7,8,10,12	95.12%	93.06%	94.07%
MEF-BERT-10	2,3,4,5,6,8,9,10,11,12	95.19%	93.17%	94.16%
MEF-BERT-12	1-12	95.25%	93.29%	94.26%

Table 7. Entity recognition results for different numbers of feature extraction layers of psychomedical datasets.

Model	P	R	F1
MFME-NER	95.25%	93.29%	94.26%
-SC Embeding	94.89%	92.93%	93.89%
-Pinyin Embeding	95.03%	93.12%	94.06%
-SC、 Pinyin Embeding	94.83%	92.69%	93.75%
-MEF-BERT	94.76%	92.89%	93.81%
-GA-FNNAttention	93.95%	92.45%	93.19%

Table 8. Ablation trials on the PsyDataset dataset.

As can be seen in Table 7, the performance of the model gradually improves with the increase in the number of feature extraction layers. When extracting all the features of the 12 layers, that is, the scheme adopted by this model, the model performance reaches the best, and the F1 Score is 94.26%. From 2-layer splicing to 8-layer splicing, the rate of model performance enhancement increases with the increase of layers, which is beacuse that MFE-BERT encodes rich linguistic hierarchical information, the bottom network can learn the superficial information features, the middle network can learn the syntactic information features, AND the high-level network is responsible for learning the semantic information features. In most cases, syntactic information can better reflect the representation of text data in NER tasks, so the more intermediate levels, the better the performance of task processing. The splicing level reaches 10 and 12 layers, although the performance improvement rate slows down, the performance is still gradually enhanced, which indicates that only more layers of the joint processing model will have the richer expressive ability, thus making the model more competitive in psychomedical NER tasks.

Ablation experiment

To verify the validity of the modules in our proposed model further, PsyDataset is selected as the experimental dataset for the ablation experiments, and the comparative model of the ablation process is as follows:

- MFME-NER: Our proposed model is shown in Fig. 1.
- -SC: The structural embedding in the MFME-NER model is removed. The model's multi-granularity embedding contains only character embedding and pinyin embedding, and does not contain structural embedding information.
- -PINYIN: Remove the pinyin embedding from the MFME-NER model.
- -MFE-BERT: Replacing the MFE-BERT pre-training for character feature extraction in the MFME-NER model with the unimproved BERT model.
- -SC, PINYIN: Structural embedding and pinyin embedding are removed from the MFME-NER model. At this point, the model is converted from a multi-granularity feature to a character embedding model.
- -GA-FNNAttention: the gated forward neural network attention mechanism in the MFME-NER model is removed. At this point, the three features, character, structure and pinyin, are fused by the Concat splicing function.

Table 8 shows the experimental data comparing each model with our proposed model after ablation treatment. The experimental data shows that our model has the best results. Removing the Pinyin Embeding module has a small effect on the performance of the model, with the F1 Score reduced by only 0.2%, which is beacuse that the Pinyin feature is mainly applied to a small number of Chinese polyphonic characters and has less relevance to other parts. After removing SC Embeding, the F1 Score of the model is reduced by 0.37%, which is beacuse that structural features play a more important role in entity recognition compared to Pinyin features. When we replace the pre-trained model with BERT and remove the SC and Pinyin Embeding modules at the same time, the F1 Score of the model is 93.81% and 93.75%, respectively, and the difference between the two is not obvious, which indicates that our proposed multi-layer feature extraction method is effective, and the MEF-BERT can extract the features with richer semantic information. In addition, after removing the GA-FNNAttention module, the F1 Score of the model changes significantly and decreases by 1.07% relative to our proposed model. This indicates that the FNNAttention mechanism and the gated feature fusion module have a key role in the NER process, and the FNNAttention can fully capture the semantic information between contexts in long text

Sentence	精神分裂症容易引发幻想异常行为 [Schizophrenia is prone to abnormal behaviors that trigger fantasies]
Gold labels	精(B-disease)神(I-disease)分(I-disease)裂(I-disease)症(I-disease)容(O)易(O)引(O) 发(O)幻(B-symptom)想(I-symptom)异(O)常(O)行(O)为(O) [In English: Schizophrenia(disease) is(O) prone(O) to(O) abnormal(O) behaviors(O) that(O) trigger(O) fantasies(symptom)]
BERT-BiLSTM-CRF	精(B-disease)神(I-disease)分(I-disease)裂(I-disease)症(I-disease)容(O)易(O)引(O) 发(O)幻(B-disease)想(I-disease)异(O)常(O)行(O)为(O) [In English: Schizophrenia(disease) is(O) prone(O) to(O) abnormal(O) behaviors(O) that(O) trigger(O) fantasies(disease)]
MFME-NER	精(B-disease)神(I-disease)分(I-disease)裂(I-disease)症(I-disease)容(O)易(O)引(O) 发(O)幻(B-symptom)想(I-symptom)异(O)常(O)行(O)为(O) [In English: Schizophrenia(disease) is(O) prone(O) to(O) abnormal(O) behaviors(O) that(O) trigger(O) fantasies(symptom)]
Sentence	分裂情感性精神病 [schizoaffective psychosis]
Gold labels	分(B-disease)裂(I-disease)情(I-disease)感(I-disease)性(I-disease)精(I-disease) 神(I-disease)病(I-disease) [In English: schizoaffective psychosis(disease)]
BERT-BiLSTM-CRF	分(O)裂(O)情(O)感(O)性(O)精(B-disease)神(I-disease)病(I-disease) [In English: schizoaffective(O) psychosis(disease)]
MFME-NER	分(B-disease)裂(I-disease)情(I-disease)感(I-disease)性(I-disease)精(I-disease) 神(I-disease)病(I-disease) [In English: schizoaffective psychosis(disease)]

Table 9. Case study of the MFME-NER method.

passages and reasonably assign the weights, together with the gated feature fusion module to filter out the noise information. In summary, these experimental results fully prove the effectiveness of our proposed method and reflect the excellent performance of our model.

Case study

To further validate the advantages of our model, two case studies were conducted using the classical BERT-BiLSTM-CRF model as a baseline approach, as shown in Table 9.

In the first case, the baseline approach incorrectly identifies "幻想"(fantasies) as diseases, while our model accurately identifies them as "symptom" types. This may beacuse that the FNNAttention mechanism of our model can take more contextual information into account. The entity is correctly recognized in the context of other passages in the article. Also, in the second case, "分裂情感性精神病"(schizoaffective psychosis) is a full "disease" type entity. Our model correctly identified it. The baseline model, on the other hand, incorrectly segregates it and recognizes "精神病" (psychosis) as a "disease" type entity. We analyze the reason for this. This is due to the important role of the structure Embeding feature in our model. The structure of "忄 " (mood) and " 心" (mood) in the Chinese characters "分裂情感性" (schizoaffective) is often used in the type of "disease" type of entities. The model captures the structural features of each component. The overall meaning of compound words is better understood. Thus, the entity can be recognized correctly.

Conclusions

In this paper, we propose a multi-layer feature extraction and multi-granularity embedding fusion method for psychomedical NER.Aiming at the characteristics of long texts in psychomedicine and the problem of single character granularity embedding information, this paper designs a pre-training model MEF-BERT with the ability of multi-layer feature extraction, and on the basis of feature information of each layer of the fully-connected Transformer Encoder, the output vectors of each layer are finally merged to give more rich word-level and semantic information to the word vectors. On the basis of the feature information of each layer of the fully connected Transformer Encoder, the output vectors of each layer are finally merged to give the word vectors richer word-level and semantic information, and the character, structure and pinyin of Chinese characters are extracted at the same time, and the feature vectors of the three levels are integrated by using the GA-FNNAttention mechanism, so as to reduce the multimodal noise and strengthen the word-level relationship. Experiments on the self-constructed psychomedical text dataset and the CBLUE public dataset show that the MEFE-NER model performs better than existing related methods on the psychomedical NER task, proving the effectiveness of our proposed method.

However, the model still has some limitations. First, the size and diversity of the dataset is limited; the field of psychological medicine covers numerous sub-directions, and the field of psychological medicine covers numerous sub-directions, such as mental disorders, psychotherapy, and neuroscience, etc. The current dataset is deficient in covering the breadth and depth of these sub-directions, which may lead to the model's limited ability to recognize some rare or emerging concepts of psychological medicine when confronted with them. Secondly, the high complexity of the model may lead to a certain impact on the operational efficiency and response speed of the model in practical applications, especially when dealing with large-scale text data. In addition, the model is specifically designed and trained for the field of psychomedicine, and its feature extraction and fusion mechanisms are closely related to the characteristics of psychomedicine texts, and thus may not achieve the same desirable recognition results in cross-domain applications.

We will improve the following aspects in the follow-up study:

(1) Dataset expansion and optimization:Further expand the size of the psychomedical dataset, add text data from more subdisciplines, improve the quality of the data and the accuracy of the annotation, and consider the introduction of multi-source data, such as clinical records, academic papers, and patients' forum discussions, in order to enrich the sources and diversity of the data, so that the model can learn a wider range of psychomedical knowledge and expressions, and enhance its generalization ability.
(2) Model Optimization and Simplification:Under the premise of maintaining the performance of the model, we explore ways to optimize and simplify the model structure. For example, research on more efficient feature extraction algorithms to reduce redundant computation; or model compression techniques, such as pruning and quantization, to reduce the number of parameters and computational complexity of the model, and to improve the model's operational efficiency.
(3) Cross-domain learning ability enhancement: Improve the model to make it have better cross-domain adaptability, introduce domain adaptive technology in the model training process, and enable the model to better understand and recognize entities in different domains by sharing some of the features and knowledge between different domains. Meanwhile, multi-task learning methods are explored to combine the psychomedical NER task with other related tasks such as sentiment analysis and relationship extraction for joint training to improve the comprehensive performance and generalization ability of the model.

Data availability
The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Received: 17 December 2024; Accepted: 17 February 2025
Published online: 15 May 2025

References
1. Kenton, J.D.M.W.C., & Toutanova, L.K. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, p 2 (2019).
2. Sun, Y., Wang, S., Li, Y., et al. Ernie: Enhanced representation through knowledge integration. arXiv preprint arXiv:1904.09223 (2021).
3. Xu, K. et al. Document-level attention-based bilstm-crf incorporating disease dictionary for disease named entity recognition. *Computers in biology and medicine* **108**, 122–132 (2022).
4. Wu, S., Song, X., & Feng, Z. Mect: Multi-metadata embedding based cross-transformer for chinese named entity recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp 1529–1539 (2022).
5. Liao, L. & Xie, S. Chinese named entity recognition based on feature fusion of attention mechanism. *Comput Eng* **49**(4), 256–262 (2023).
6. Raffel, C., & Ellis, D.P. Feed-forward networks with attention can solve some long-term memory problems. arXiv preprint arXiv:1512.08756 (2015).
7. Mikolov, T., Sutskever, I., Chen, K., et al. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26, (2013).
8. Pennington, J., Socher, R., & Manning, C.D. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp 1532–1543, (2014).
9. Peters, M.E., Neumann, M., Iyyer, M., et al. Deep contextualized word representations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp 2227–2237, (2021).
10. Radford, A., Narasimhan, K., Salimans, T., et al. Improving language understanding by generative pre-training, https://openai.com/index/language-unsupervised, (2018).
11. Lan, Z., Chen, M., Goodman, S., et al. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942, (2019).
12. Lee, J. et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **36**(4), 1234–1240 (2021).
13. Bian, J., Jiang, R., Zhai, W., et al. Dmner: Biomedical named entity recognition by detection and matching. In **2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, pp 872–878 (2024).
14. Biana, J., Zhai, W., Huang, X., et al. Vaner: Leveraging large language model for versatile and adaptive biomedical named entity recognition. arXiv preprint arXiv:2404.17835 (2024).
15. Alamro, H. et al. Biobbc: a multi-feature model that enhances the detection of biomedical entities. *Scientific reports* **14**(1), 7697 (2024).
16. Song, L., & Zhang, L. Tfebioner: A new method for biomedical named entity recognition based on table filling for feature extraction. In *Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering*, pp 693–697, (2024).
17. Yin, Y., Kim, H., Xiao, X., et al. Augmenting biomedical named entity recognition with general-domain resources. arXiv preprint arXiv:2406.10671 (2024).
18. Collobert, R., Weston, J., Bottou, L., et al. Natural language processing (almost) from scratch. Journal of machine learning research 12(ARTICLE):2493–2537 (2011).
19. Habibi, M. et al. Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics* **33**(14), i37–i48 (2017).
20. Guo, Y., Yu, H., Ma, L., et al. Die-cdk: A discriminative information enhancement method with cross-modal domain knowledge for fine-grained ship detection. *IEEE Transactions on Circuits and Systems for Video Technology*, (2024).
21. Guo, Z. & Deng, X. Intelligent identification method of legal case entity based on bert-bilstm-crf. *Journal of Beijing University of Posts and Telecommunications* **44**(4), 129 (2021).
22. Ren, Y. et al. Recognition of quantitative indicator of fishery standard using attention mechanism and the bert+bilstm+crf model. *Transactions of the Chinese Society of Agricultural Engineering* **37**(10), 135–141 (2023).
23. Guo, Y. et al. Dsca: A dual semantic correlation alignment method for domain adaptation object detection. *Pattern Recognition* **150**, 110329 (2024).
24. Mo, Y., Yang, J., Liu, J., et al. Mcl-ner: Cross-lingual named entity recognition via multi-view contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp 18789–18797, (2024).

25. Hu, Z. & Ma, X. A novel neural network model fusion approach for improving medical named entity recognition in online health expert question-answering services. *Expert Systems with Applications* **223**, 119880 (2023).
26. Lou, Y., Zhu, X. & Tan, K. Dictionary-based matching graph network for biomedical named entity recognition. *Scientific Reports* **13**(1), 21667 (2023).
27. Cui, Z. et al. Incorporating part of speech information in span representation for named entity recognition. *Applied Soft Computing* **163**, 111844 (2024).
28. Ye, D., Lin, Y., Li, P., et al. Packed levitated marker for entity and relation extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp 4904–4917, (2022).
29. Li, J., & Meng, K. Mfe-ner: multi-feature fusion embedding for chinese named entity recognition. In *China National Conference on Chinese Computational Linguistics, Springer*, pp 191–204, (2024).
30. Qiu, Q. et al. Chinese engineering geological named entity recognition by fusing multi-features and data enhancement using deep learning. *Expert Systems with Applications* **238**, 121925 (2024).
31. Meng, Y., Wu, W., Wang, F., et al. Glyce: Glyph-vectors for chinese character representations. *Advances in Neural Information Processing Systems* 32, (2019).
32. Xuan, Z., Bao, R. & Jiang, S. Fgn: Fusion glyph network for chinese named entity recognition. In *Knowledge Graph and Semantic Computing: Knowledge Graph and Cognitive Intelligence* 28–40 (Springer Singapore, Singapore, 2021).
33. Guo, Y. et al. Dp-ddcl: A discriminative prototype with dual decoupled contrast learning method for few-shot object detection. *Knowledge-Based Systems* **297**, 111964 (2024).
34. Li, X., Yan, H., Qiu, X., et al. Flat: Chinese ner using flat-lattice transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp 6836–6842, (2021).
35. Guo, Y. et al. Thfe: a triple-hierarchy feature enhancement method for tiny boat detection. *Engineering Applications of Artificial Intelligence* **123**, 106271 (2023).
36. Wu, S., Song, X., & Feng, Z. Mect: Multi-metadata embedding based cross-transformer for chinese named entity recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp 1529–1539 (2022).
37. Li, M., Huang, R., & Qi, X. (2024). Chinese clinical named entity recognition using multi-feature fusion and multi-scale local context enhancement. *Computers, Materials & Continua* 80(2)
38. Maimai, A. et al. Uyghur named entity recognition based on bilstm cnn-crf model. *Computer Engineering* **44**(8), 230–236 (2018).
39. Wang, L. et al. Chinese medicine literature word segmentation based on bilstm-crf. *Application Research of Computers* **37**(11), 3359 (2023).
40. Khandelwal, U., He, H., Qi, P., et al. Sharp nearby, fuzzy far away: How neural language models use context. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp 284–294, (2022).

Author contributions
The authors confirm contribution to the paper as follows: Z.L. completed the conception and design of the study, collected data, analyzed the results, and wrote the manuscript. G.Z. participated in the conception and design of the study and analyzed the results. Y.S. collected data. All authors reviewed the results and approved the final version of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to G.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.