基于大语言模型多维度特征增强的医学命名实体识别方法

蒋明翀[1]，叶娜[1]，徐波波[1*]，
[1]沈阳航空航天大学　计算机学院
jiangming2@stu.sau.edu.cn, yena_1@126.com
xubobo@stu.sau.edu.cn

摘要

医学命名实体识别在医疗信息提取和知识图谱构建中至关重要，但因医学领域的专业性和复杂性，面临数据稀缺、特征不显著及上下文利用不足的挑战。本文提出LLM-MedNER方法，充分利用大语言模型（LLM）的预训练知识，通过提示工程生成语义等价但表达多样的增强文本，并提取多维度特征，包括关键字集合、语义描述、词性信息及医学实体关联特征，从而显著提升模型的特征表达能力。方法采用双通道MacBERT-BiGRU编码模块并行学习原始文本特征与大语言模型增强特征，通过交叉注意力机制融合不同语义特征。随后，引入自适应多粒度扩张卷积层，通过不同膨胀率的一维卷积捕获多尺度的局部上下文信息，进一步丰富词表示。并在输出层引入Biaffine 模块实现实体边界及类型的精准识别。对比实验表明，LLM-MedNER 在多个医学命名实体识别数据集上的表现优于现有基线方法；消融实验进一步证实各模块的有效性。

关键词： 医学命名实体识别；大语言模型；双通道编码；交叉注意力机制；自适应卷积层

A Medical Named Entity Recognition Method Enhanced by Multi-Dimensional Features Based on Large Language Models

Mingchong Jiang[1], Na Ye[1], Bobo Xu[1*]
[1]School of Computer Science, Shenyang Aerospace University
jiangming2@stu.sau.edu.cn, yena_1@126.com
xubobo@stu.sau.edu.cn

Abstract

Medical named entity recognition is crucial in medical information extraction and knowledge graph construction. However, due to the specialization and complexity of the medical field, it faces challenges such as data scarcity, lack of significant features, and insufficient utilization of context.This paper proposes the LLM-MedNER method, which leverages the pre-trained knowledge of large language models (LLMs).By employing prompt engineering, it generates semantically equivalent but expressively diverse augmented texts and extracts multi-dimensional features, including keyword sets, semantic descriptions, part-of-speech information, and medical entity association features. This approach significantly enhances the model's feature representation capabilities.The method employs a dual-channel MacBERT-BiGRU encod-ing module to learn in parallel both original text features and features enhanced by large language

第二十四届中国计算语言学大会论文集，第460页-第473页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

models. Different semantic features are inte-grated using a cross-attention mechanism.Subsequently, an adaptive multi-granularity dilated convolution layer is introduced to capture mul-ti-scale local contextual information through one-dimensional convolutions with different dilation rates, further enriching word representa-tions.The Biaffine module is introduced in the output layer to achieve accurate recognition of entity boundaries and types. Comparative exper-iments indicate that LLM-MedNER outperforms existing baseline methods on multiple medical named entity recognition datasets. Ablation studies further verify the effectiveness of each module.

Keywords: Keywords Medical Named Entity Recognition , Large Language Model , Dual-Channel Encoding , Cross-Attention Mechanism , Adaptive Convolutional Layer

1 引言

近年来，医学命名实体识别（NER）在医疗信息提取、知识图谱构建及辅助诊断等领域扮演着愈发关键的角色。医学文献中蕴含大量专业实体信息，如基因、药物和疾病等，NER的核心任务是从文本中精准划定这些实体的边界并归类至预定义类别。这一技术的进步显著提升了医学信息处理的效率，并为下游任务提供了精确的语义支持(Ohno et al., 2024)。

尽管NER领域的研究已取得长足进展，但医学文本的独特性和复杂性仍带来诸多挑战。首先，数据稀缺与特征不显著并存：医学子领域因专业性强且受隐私限制，高质标注数据集难以获取。其次，医学文本表达形式多样且高度依赖上下文，传统方法如支持向量机（SVM）、隐马尔科夫模型（HMM）和条件随机场（CRF）虽在特定场景下有效，但在复杂医学文本中表现受限(Chen et al., 2023)。近年来，深度学习特别是预训练模型的兴起为医学NER注入新活力，其多头自注意力机制能捕捉长距离依赖，增强上下文建模能力。然而，这些方法在医学应用中仍存不足：训练中对非实体部分的文本表达灵活性不足，影响收敛效率(Sun et al., 2023)；对上下文信息的挖掘不够深入，未能充分利用非实体部分的潜在语义(Wan et al., 2020)。

现有的医学NER方法主要存在以下局限性：首先，传统特征增强方法仅关注单一文本表示增强，缺乏多维度特征提取，且采用单通道架构，无法实现不同特征表示间的有效互补(Bai et al., 2020)；此外，现有方法(Hu et al., 2022b)主要融入外部知识，未能充分利用大语言模型的深度语义理解能力；最后，如(An et al., 2022)多采用固定窗口处理，难以适应医学实体长度变化大的特点。

本文提出一种新型医学命名实体识别方法——LLM-MedNER，针对上述问题提出了三个核心创新：第一，多任务提示策略引导大模型提取多维度医学特征（关键字、语义描述、词性和关联特征），这与现有仅关注单一文本表示的方法有本质区别，能够深度挖掘医学文本的语义层次；第二，设计的双通道编码架构通过并行编码和交叉注意力机制实现特征互补，相比单通道融合方法能够更好地利用不同表示间的协同效应；第三，针对医学实体长度变化大的特点，提出自适应多粒度扩张卷积层，相比固定窗口方法更具适应性，能够灵活捕获不同尺度的上下文信息。

LLM-MedNER充分利用大语言模型在海量医疗语料上的预训练优势，通过精心设计的提示策略生成增强文本，并结合双通道编码网络提升识别性能。系统首先识别文本中的潜在医学实体，包括疾病名称、症状和治疗方法等，确保这些实体的语义和相对位置保持不变；然后仅改写实体外的上下文部分，通过同义词替换、句式变换或语序调整等方式丰富表达，同时保证增强文本与原始文本的医疗意图和整体语义一致，避免引入歧义或改变实体含义。这一过程为模型提供了更丰富的语言表达和上下文特征。此外，我们还利用大语言模型提取多维度特征，包括关键字、语义描述、词性信息及实体间关联关系，并通过内化医疗领域知识，捕捉实体间的隐含语义联系，如疾病与治疗的因果关系、症状与疾病的伴随关系，进一步增强特征表达的深度与多样性，提升模型对复杂医学文本的理解能力。网络结构设计上，LLM-MedNER采用

* 通讯作者
©2025 中国计算语言学大会
根据《Creative Commons Attribution 4.0 International License》许可出版
基金项目：国家自然科学基金项目(U1908216)

双通道MacBERT-BiGRU模块并行处理原始文本与增强文本，提取深层语义特征，并通过交叉注意力机制融合双路径信息。在此基础上，引入自适应多粒度扩张卷积层，通过多个不同膨胀率的一维卷积并行提取多尺度的局部上下文信息，进一步丰富词级表示；最后，在输出层加入Biaffine模块，通过二次仿射变换精确建模词间依赖，实现高精度的实体边界和类型预测。实验结果表明，LLM-MedNER在多个医学数据集上的精度和召回率均超越现有方法，消融实验进一步验证了特征增强与网络结构设计的协同效应。

2 相关研究

命名实体识别是一项核心的自然语言处理任务，旨在从非结构化文本中识别并抽取具有特定意义的实体(何玉洁et al., 2021)。早期的NER 研究主要集中于基于规则的方法，这类方法通常依赖专家制定的规则和领域特定的词典，以识别和分类实体(向晓雯et al., 2005)。例如，利用正则表达式和词汇匹配技术(丁建平et al., 2024)能够在结构化程度较高或文本格式相对固定的场景下取得较好的效果。然而，随着应用领域和文本风格的逐渐多样化，基于规则的方法在维护成本、适用性和泛化性方面的局限也随之显现(雷宇翔and 廖涛, 2025)。

为突破人工规则在可扩展性和自适应性方面的限制，研究者开始将统计学习理论引入到NER 任务中。典型的统计模型包括隐马尔可夫模型(Bikel et al., 1998)、条件随机场(杨荣根et al., 2022)和支持向量机(Imam et al., 2021)。这些方法通过特征工程提取文本中的多种特征，以训练分类器进行实体识别。尽管统计学习方法在一定程度上减少了对手工规则的依赖，但其依然需要大量的人工特征设计，且在处理大规模、多样化或跨语言的数据时，性能表现不够稳定(Goyal and Singh, 2024)。

深度学习技术的迅猛发展为NER 任务带来了新的突破。循环神经网络（Recurrent Neural Networks, RNN）及其变体长短期记忆网络（Long Short-Term Memory, LSTM）(Naik et al., 2024)和双向LSTM （BiLSTM）(项恒et al., 2024)不断被用于NER 任务，这些模型能够自动学习文本的时序特征，避免了复杂的特征工程。结合条件随机场层（CRF Layer）的端到端架构，如Mekki等人(Mekki et al., 2024)提出BiLSTM-CRF 模型，进一步提升了序列标注的准确性和鲁棒性。此外，卷积神经网络（Convolutional Neural Networks, CNN）也被广泛应用于NER任务，用于提取局部特征。例如，Sawinder等人(Kaur et al., 2020) 将LSTM和CNN结合，构建了一个CNN-LSTM神经网络结构，通过CNN 提取词级和字符级的局部特征。Junting等人(Lin and Liu, 2022)进一步提出了BiLSTM-CNNs-CRF模型，结合了双向循环神经网络、CNN和CRF，充分利用了各模块的优势，提升了实体识别的效果。随着深度学习方法的不断发展，预训练语言模型（Pre-trained Language Models, PLM）的出现进一步提升了NER 的性能。BERT （Bidirectional Encoder Representations from Trans-formers）(Park et al., 2023)通过在大规模无监督语料上进行双向训练，获得了强大的上下文语义理解能力，其变体如RoB-ERTa(顾泽et al., 2023)和MacBERT(Yuan et al., 2024)等在多个NER 基准数据集上表现优异，显著减少了对人工特征工程的依赖，并在医学和法律等专业领域展现了良好的适应性(Zhang et al., 2023)。例如Chen等人(Chen et al., 2024)采用基于ALBERT-BiLSTM-CRF 的半监督的中文电子病历NER 模型，通过ALBERT 进行特征提取，双向长短期记忆网络加条件随机场来训练数据，有效提高模型准确度。

继预训练模型之后，大型语言模型（Large Lan-guage Models, LLM）如GPT 系列(García-Barragán et al., 2024)、T5(Chung et al., 2024)和ChatGLM(Zeng et al., 2022) 迅速崛起，这些模型拥有更多的参数，能够通过海量多领域数据进行深层次的语言模式学习和语义学习(段宇锋and 谢佳宏, 2025)。LLM 能够通过其庞大的参数和预训练的知识，在文本中提取出多种不同类型的特征。例如，提取关键字、实体之间的关系、语义描述、句法结构等。LLM也可以通过提示（Prompt）实现零样本或小样本的推断与生成(徐春et al., 2023)。对于NER 而言，LLM所具备的多样化生成能力可用于构造或改写训练样本，以缓解标注数据不足的问题。此外，LLM所具备的提取特征能力可以帮助NER 模型更好地理解复杂文本中的潜在信息，并且在某些场景下能显著提升实体识别的鲁棒性与泛化性。

3 LLM-MedNER框架

LLM-MedNER网络框架如图1所示，主要由LLM文本增强模块、双通道编码器层、多粒度扩张卷积层、共预测器层四个模块组成。以下将详细描述各部分的具体实现和细节。

(c) 2025 中国中文信息学会计算语言学专业委员会

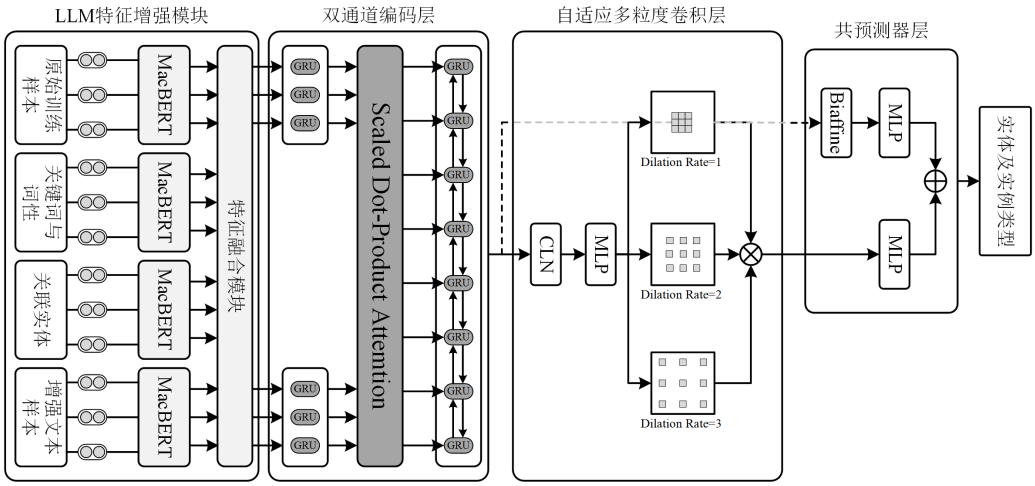

Figure 1: LLM-MedNER网络框架

3.1 LLM特征增强模块

文本增强模块通过大语言模型生成增强文本并提取多维度特征，以增强模型对医疗文本的表示能力。每个输入样本包含一条医疗文本$T = \{x_1, x_2, \ldots, x_N\}$，其中$x_i$表示文本中的词，N为文本长度。大语言模型通过多任务提示策略生成增强文本$T' = \{x'_1, x'_2, \ldots, x'_M\}$，其中M为增强文本的长度，并提取多维度特征，包括关键字集合$K = \{k_1, k_2, \ldots, k_P\}$，关键字描述特征$H_{\text{desc}}$，医学实体关联特征$H_{\text{rel}}$，词性特征$H_{\text{pos}}$。最终，生成原始文本特征$H_{\text{aug}}$和增强特征$H_{\text{aug}}$，为后续的双通道编码器层提供输入。

3.1.1 增强文本生成

大语言模型在特征增强模块中的首要任务是生成与原始医疗文本语义等价但表达方式不同的增强文本，从而扩充训练数据的多样性，为模型提供更丰富的语言表达形式。大语言模型利用其在海量医疗语料上的预训练知识，能够深刻理解医疗文本中的专业术语和上下文关系，例如疾病、治疗方法、症状等概念之间的关联。生成增强文本的过程不仅需要保持语义一致，还需通过多样化的表达方式增强模型的泛化能力。在预测阶段，大语言模型识别文本中的潜在医学实体并进行增强。增强过程需满足：保持潜在实体的语义和位置不变，确保与原始文本语义一致；仅重写实体外的上下文，通过同义词替换、句式变换、语序调整等方式丰富表达；保证医疗意图和整体语义一致，避免歧义或改变实体含义。增强文本T'随后通过MacBERT编码为特征表示：

$$H_{\text{aug}} = \text{MacBERT}\left(T'\right) = \left\{h^1_{\text{aug}}, h^2_{\text{aug}}, \ldots, h^M_{\text{aug}}\right\} \in \mathbb{R}^{M \times d_h} \tag{1}$$

其中d_h为MacBERT的隐藏层维度，表示每个词的特征向量维度。原始文本T同样通过MacBERT编码为：

$$H_{\text{org}} = \text{MacBERT}(T) = \left\{h^1_{\text{org}}, h^2_{\text{org}}, \ldots, h^N_{\text{org}}\right\} \in \mathbb{R}^{N \times d_h} \tag{2}$$

增强文本生成提示模板如表1所示。

3.1.2 关键词及词性信息提取

大语言模型从医疗文本T和增强文本T'中提取关键字及其相关特征，生成关键字集合$K = \{k_1, k_2, \ldots, k_P\}$，其中$P$为关键字数量。关键字通常是医疗文本中的核心术语或短语，往往是潜在医学实体的核心组成部分。大语言模型利用其在医疗领域的预训练知识，能够准确识别这些术语，例如疾病名称，症状描述或治疗方法等。提取关键字后，大语言模型进一步生成多维度特征，包括关键字描述特征和词性特征，以丰富模型对关键信息的理解能力。大语言模型首先基于上下文和预训练知识识别T和T'中的关键字集合$K = \{k_1, k_2, \ldots, k_P\}$。关键字集合K通过MacBERT编码为特征表示：

$$H_K = \text{MacBERT}(K) = \left\{h^1_K, h^2_K, \ldots, h^P_K\right\} \in \mathbb{R}^{P \times d_h} \tag{3}$$

(c) 2025 中国中文信息学会计算语言学专业委员会

用户输入	GLM输出
根据给定的医疗文本，生成一个语义等价但表达方式不同的增强文本。要求如下： 1.识别文本中的潜在医学实体（如疾病名称、症状、治疗方法等），并确保这些实体的语义和在文本中的相对位置保持不变； 2.仅改写潜在医学实体外的上下文部分，通过同义词替换、句式变换或语序调整等方式丰富表达； 3.确保生成的增强文本与原始文本的医疗意图和整体语义保持一致，避免引入歧义或改变实体含义。 输入文本：患者因高血压接受药物治疗效果不佳。输出格式：增强文本：增强文本	增强文本：患者由于高血压而接受的药物治疗未能取得理想效果。

Table 1: 增强文本生成提示模板

为每个关键字k_i 生成简要的语义描述信息，解释其在医疗文本中的含义。例如，对于关键字 " 发热 " ，描述可能是 " 体温升高，通常为疾病症状 " 。所有描述信息集合记为 " Descriptions " ，通过MacBERT 编码为句子级特征向量：

$$H_{\text{desc}} = \text{MacBERT}(\text{ Descriptions }) = \left\{h^1_{\text{des}}, h^2_{\text{desc}}, \ldots, h^P_{\text{desc}}\right\} \in \mathbb{R}^{P\times d_h} \tag{4}$$

其中$h^i_{\text{desc}} \in \mathbb{R}^{d_h}$ 表示第i 个关键字的描述特征向量。关键字描述特征为模型提供额外的背景知识，帮助更准确地理解术语在医疗文本中的作用。关键词信息提取提示模板如表2 所示。

在完成增强文本生成后，随后进行词性特征提取步骤，大语言模型分析关键字集合K 中每个关键字的词性（如名词，动词，形容词等），生成词性集合$POS = \{\text{pos}_1, \text{pos}_2, \ldots \text{pos}_P\}$ ，。例如，" 发热 " 可能标注为名词，" 治疗 " 可能为动词。词性集合POS 通过MacBERT 编码为特征表示：

$$H_{\text{pos}} = \text{MacBERT}(POS) = \left\{h^1_{\text{pos}}, h^2_{\text{pos}}, \ldots, h^P_{\text{pos}}\right\} \in \mathbb{R}^{P\times d_h} \tag{5}$$

其中$h^i_{\text{pos}} \in \mathbb{R}^{d_h}$ 表示第i 个关键字的词性特征向量。词性特征增强了模型对关键字语法功能的理解，有助于后续实体识别和关系分析。词性提取可通过提示策略实现，模板可参考表2扩展。

关键词提取提取模板和词性提取模板如表2、表3所示。

用户输入	GLM 输出
根据给定的医疗文本，提取关键词并为每个关键词提供简要的语义描述。要求如下： 1.关键词应为医疗文本中的核心术语或短语，例如疾病名称、症状描述、治疗方法等，通常是潜在医学实体的核心组成部分； 2.为每个提取的关键词生成一个简短的语义描述，解释其在当前医疗文本中的具体含义，描述应基于文本上下文和医疗领域知识； 3.确保关键词及其描述准确反映文本的核心信息。 输入文本：患者因高血压接受药物治疗效果不佳。 输出格式：关键词：{关键词}，描述：{描述}	关键词：高血压，描述：一种血压持续高于正常水平的慢性疾病，在此指患者的主要健康问题。 关键词：药物治疗，描述：通过服用药物来控制或缓解疾病的方法，在此指针对高血压的干预措施。 关键词：效果不佳，描述：治疗结果未达到预期目标的状态，在此指药物未能有效控制高血压。

Table 2: 关键词信息提取提示模板

第二十四届中国计算语言学大会论文集，第460页-第473页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

用户输入	GLM 输出
根据给定的关键词列表，为每个关键词标注其词性。 要求如下： 1.关键词列表来源于医疗文本的提取结果，可能是疾病名称、治疗方法、症状等； 2.为每个关键词标注准确的词性（如名词、动词、形容词等），基于其在原始文本中的语法功能； 3.确保词性标注符合医疗文本的上下文语义。 输入关键词列表：高血压、药物治疗、效果不佳。 输出格式：关键词：{关键词}，词性：{词性} 本的核心信息。 输入文本：患者因高血压接受药物治疗效果不佳。 输出格式：关键词：{关键词}，描述：{描述}	关键词：高血压，词性：名词 关键词：药物治疗，词性：名词 关键词：效果不佳，词性：形容词

Table 3: 词性提取提示模板

3.1.3 医学实体关联特征提取

大语言模型分析医疗文本中潜在医学实体之间的关联关系，提取医学实体关联特征H_{rel}，以增强模型对文本的理解能力。医学实体关联特征旨在捕捉潜在医学实体之间的关系，例如疾病与治疗方法之间的因果关系，疾病与症状之间的伴随关系等。这种特征在医疗文本中尤为重要，因为实体之间的关联往往提供关键上下文信息，帮助模型更准确地识别和分类实体。通过提示策略（模板如表3所示），大语言模型首先识别文本中的潜在医学实体，然后分析它们之间的关联关系，生成描述信息。实体对集合定义为$E = \{(e_{a_1}, e_{b_1}), (e_{a_2}, e_{b_2}), \ldots, (e_{a_R}, e_{b_R})\}$，其中R 为实体对数量，每个实体对$(e_{a_j}, e_{b_j})$ 表示一对相关的医学实体。实体对及其关联关系描述（文本形式）集合记为 " Relations "，通过MacBERT 编码为特征向量：

$$H_{\text{rel}} = \text{MacBERT}(\text{ Relations }) = \{h^1_{\text{rel}}, h^2_{\text{rel}}, \ldots, h^R_{\text{rel}}\} \in \mathbb{R}^{R \times d_h} \tag{6}$$

其中$h^j_{\text{rel}} \in \mathbb{R}^{d_h}$ 表示第j 个实体对的关联特征向量。由于实体对通常由关键字组成，H_{rel} 将在后续双通道编码器层中映射到对应的关键字位置。关联特征提取提示模板如表4 所示。

用户输入	GLM 输出
根据给定的医疗文本和提取的关键词，识别医学实体对并描述它们之间的关联关系。要求如下： 1.从文本中识别潜在的医学实体对，例如疾病与治疗方法、疾病与症状等，实体对应于已提取的关键词； 2.为每个实体对提供一个关联关系的描述，说明它们之间的关系类型（如因果关系、伴随关系、结果关系等），描述需基于文本上下文和医疗知识； 3.确保关联关系准确反映文本中的语义逻辑，避免臆断或偏离文本内容。 输入文本：患者因高血压接受药物治疗效果不佳。 输入关键词：高血压、药物治疗、效果不佳。 输出格式：实体对：{实体1}-{实体2}， 关联关系：{关系描述}	实体对：高血压-药物治疗， 关联关系：因果关系（高血压是患者接受药物治疗的原因，药物治疗旨在控制高血压）。 实体对：药物治疗-效果不佳， 关联关系：结果关系（药物治疗的实施结果是效果不佳，未能有效缓解高血压）。

Table 4: 关联特征提取提示模板

3.2 双通道编码器层

双通道编码器层负责将来自大语言模型（LLM）特征增强模块的多种特征融合成统一的特征表示H，简单来说其输入包括原始文本特征$H_{\text{org}} \in \mathbb{R}^{N \times d_n}$，增强文本特征$H_{\text{aug}} \in \mathbb{R}^{M \times d_h}$，

(c) 2025 中国中文信息学会计算语言学专业委员会

关键字描述特征，医学实体关联特征$H_{\text{rel}} \in \mathbb{R}^{R\times d_h}$ ，词性特征$H_{\text{pos}} \in \mathbb{R}^{P\times d_h}$ 以及关键字集合特征$H_K \in \mathbb{R}^{P\times d_h}$ 。该层通过特征映射，序列编码和交叉注意力机制，逐步融合这些特征，生成最终的融合特征H 。下面详细介绍图1 中特征融合模块的流程。

通过特征映射，将与关键字相关的特征（如H_{desc} , H_{pos} , H_{rel} 融入到原始文本特征H_{org} 和增强文本特征H_{aug} 中对应关键字的位置。假设关键字集合为$K = \{k_1, k_2, \ldots, k_P\}$ ，其在原始文本T 中的位置为$\{p_1, p_2, \ldots, p_P\}$ ，在增强文本T' 中的位置为$\{q_1, q_2, \ldots, q_P\}$ 。对于关键字描述特征H_{desc} 和词性特征H_{pos} ，通过加权映射融入对应位置：

$$\begin{aligned} h_{\text{org}}^{p_i} &\leftarrow h_{\text{org}}^{p_i} + \alpha h_{\text{desc}}^{i} + \gamma h_{\text{pos}}^{i} \\ h_{\text{aug}}^{q_i} &\leftarrow h_{\text{aug}}^{q_i} + \alpha h_{\text{desc}}^{i} + \gamma h_{\text{pos}}^{i} \end{aligned} \tag{7}$$

对于医学实体关联特征H_{rel} ，假设第j 个实体对由关键字k_{a_j} 和k_{b_j} 组成，其映射方式为：

$$\begin{aligned} h_{\text{org}}^{p_{a_j}} &\leftarrow h_{\text{org}}^{p_{a_j}} + \beta h_{\text{rel}}^{j} \\ h_{\text{org}}^{p_{b_j}} &\leftarrow h_{\text{org}}^{p_{b_j}} + \beta h_{\text{rel}}^{j} \\ h_{\text{aug}}^{q_{a_j}} &\leftarrow h_{\text{aug}}^{q_{a_j}} + \beta h_{\text{rel}}^{j} \\ h_{\text{aug}}^{q_{b_j}} &\leftarrow h_{\text{aug}}^{q_{b_j}} + \beta h_{\text{rel}}^{j} \end{aligned} \tag{8}$$

其中，α, β, γ 为可学习权重参数。完成特征映射后，更新后的H_{org} 和H_{aug} 包含丰富语义和关联信息。接着，使用双向GRU（BiGRU）对H_{org} 和H_{aug} 进行编码，捕捉序列上下文信息：

$$\begin{aligned} H'_{\text{org}} &= \text{BiGRU}\left(H_{\text{org}}\right) \\ H'_{\text{aug}} &= \text{BiGRU}\left(H_{\text{aug}}\right) \end{aligned} \tag{9}$$

为将增强文本信息融入原始文本，使用交叉注意力机制融合H'_{org} 和H'_{aug} ，以H'_{org} 作为查询（Query），H'_{aug} 作为键（Key）和值（Value），计算注意力权重并通过残差连接生成融合特征：

$$H_{\text{fused}} = H'_{\text{org}} + \text{Attention}(Q, K, V) \tag{10}$$

其中，$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\text{T}}}{\sqrt{d_h}}\right)V$ ，且$Q = H'_{\text{org}}W_Q, K = H'_{\text{aug}}W_K, V = H'_{\text{aug}}W_V$ 。最后，对H_{fused} 再次使用BiGRU 编码，生成最终特征表示：

$$h = \text{BiGRU}\left(H_{\text{fused}}\right) \tag{11}$$

该特征h 融合了原始文本和增强文本的多维度信息，为后续模块提供全面支持。

3.3 自适应多粒度卷积层

自适应多粒度卷积层的用于构建和优化词的表示，以支持后续的实体识别任务。如图1多粒度扩张卷积层所示，包括自适应层归一化（Adaptive Layer Normalization, AdaLN）、多层感知机（MLP）以及多粒度扩张卷积模块三个模块。

首先，通过自适应层归一化（AdaLN）直接对h 进行归一化，生成高质量的词表示$C \in \mathbb{R}^{N\times d_h}$ 。接着，应用多层感知机（MLP）对归一化后的表示进行进一步的非线性变换，得到增强后的词表示c' 。具体来说，对于每个词x_i ，其表示c_i 和c'_i 由以下公式计算：

$$c_i = \text{AdaLN}\left(h_i\right) = \gamma_i \circ \frac{h_i - \mu}{\sigma} + \lambda_i \tag{12}$$

$$c'_i = \text{MLP}\left(c_i\right) = \text{ReLU}\left(W_1 c_i + b_1\right) W_2 + b_2, \tag{13}$$

其中，增益参数γ_i 和偏置λ_i 分别由以下公式计算：

$$\gamma_i = W_a h_i + b_a, \quad \lambda_i = W_B h_i + b_B \tag{14}$$

第二十四届中国计算语言学大会论文集，第460页-第473页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

均值μ 和标准差σ 则通过h_i 的元素计算得出：

$$\mu = \frac{1}{d_h}\sum_{k=1}^{d_h} h_{ik}, \quad \sigma = \sqrt{\frac{1}{d_h}\sum_{k=1}^{d_h}(h_{ik}-\mu)^2}. \tag{15}$$

参数W_1, W_2, b_1, b_2 为可学习的权重和偏置。

为了捕捉不同距离的词之间的交互，采用多粒度扩张卷积。具体而言，应用多个具有不同扩张率$l \in \{1,2,3\}$ 的一维扩张卷积（DConv），每个卷积的输出表示为：

$$D^l = \sigma\left(\mathrm{DConv}_1\left(c'\right)\right), \tag{16}$$

其中，σ 为GELU 激活函数。将所有扩张率的卷积输出拼接起来，得到最终的词表示：

$$D = \left[D^1, D^2, D^3\right] \in \mathbb{R}^{N\times 3d_c}. \tag{17}$$

3.4 共预测器层(Co-Predictor Layer)

为了加强LLM-MedNER的表现，我们设计了一种结合Biaffine的预测层，该层通过联合推理单词间的关系和上下文特征来提高模型的分类能力。Biaffine 根据3.2 所得的词表示矩阵$h \in \mathbb{R}^{N\times d_h}$ ，首先通过线性变换将其映射到两个特征空间，分别为$x \in \mathbb{R}^{N\times d_h}$ 和$y \in \mathbb{R}^{N\times d_{\mathrm{h}}}$ ，公式如下：

$$x = hW_x, \quad y = hW_y \tag{18}$$

其中，$W_x, W_y \in \mathbb{R}^{d_b\times d_{\mathrm{h}}}$ 是可训练的权重矩阵。接着，计算单词间的关系分数矩阵$S_b \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ ，并通过线性变换映射到类别空间$S_b' \in \mathbb{R}^{N\times labels}$ 。计算过程可表示为：

$$S_b' = xW_{b1}y^TW_{b2} \tag{19}$$

其中，$W_{b1} \in \mathbb{R}^{d_h\times d_h}$ 和$W_{b2} \in \mathbb{R}^{N\times \text{ labels}}$ 是可训练的权重矩阵。

同时，根据3.3 所述的卷积操作，得到上下文特征矩阵$D \in \mathbb{R}^{N\times 3d_c}$ ，其通过线性变换DW_d 得到上下文持征分数矩阵$S_d \in \mathbb{R}^{N\times \text{ labels}}$ ，其中，$W_d \in \mathbb{R}^{3\lambda_c\times \text{ labels}}$ 是可训练的权重矩阵。

最终，将关系分数S_b' 和上下文特征分数S_d 按元素相加，得到最终分类结果矩阵$S \in \mathbb{R}^{N\times labels}$ ：

$$S = S_b' + S_d \tag{20}$$

通过双仿射预测器捕获了单词间的关系特征，而上下文特征表示则补充了全局语义信息，二者的结合使模型能够更加准确地完成命名实体识别任务。

3.5 损失函数

对于每个词，类型预测损失通过交叉熵来衡量模型预测的实体类型与真实实体类型之间的差异。具体地，类型预测损失定义为：

$$\mathcal{L}_{\text{type}} = -\sum_{i=1}^{N}\sum_{f=1}^{c} y_{i,c}\log\left(p_{i,f}\right) \tag{21}$$

其中，$y_{i,c}$ 是第i 对词在第c 类别上的真实标签，若词对属于实体类型c ，则为1 ，否则为$0, p_{i,f}$ 是S通过Softmax 函数计算得到的第i 对词属于第f 类实体类型的预测概率。

4 实验

4.1 数据集

本文选用了三个公开的医学命名实体识别数据集，即CCKS2019、CCKS2020和CMeEE-V2。这些数据集涵盖了丰富的医学实体类型，包括疾病、症状、药物、检查项目等，每个实体均标注了其在文本中的起始和结束位置。CCKS2019和CCKS2020 数据集由相关机构发布，后者是在前者基础上的扩展版本，增加了样本数量和实体种类，以提升数据的多样性和复杂性。CMeEE-V2数据集则进一步细化了实体类别，涵盖了更为细粒度的医学术语，如手术、诊断方法和治疗方案等。此外，为了保证模型训练与评估的公平性，本文将以上三个数据集分别按照8:1:1的比例划分为训练集、验证集和测试集。

(c) 2025 中国中文信息学会计算语言学专业委员会

4.2 对比模型

为了验证所提方法的有效性，本文选取了多种主流的命名实体识别模型作为对比，包括BERT(李哲and 殷继彬, 2024)、RoBERTa-LSTM(Chen et al., 2025)、RoBERTa-BiLSTM-CRF(陆泽健et al., 2024)、CNN-BiLSTM-CRF(Mengliev et al., 2025)、CNN-CRF(Kachanov et al., 2024)、BERT-CRF(Hu et al., 2022a)以及BiLSTM-CRF(Siami-Namini et al., 2019)。

4.3 参数配置

为了确保实验结果的可重现性和公平性，本文选择ChatGLM-6B-v1.1作为特征增强模块，temperature设置为0.2，top_p值设置为0.85，最大生成长度为1024个token。LLM-MedNER模型的双通道编码器均采用MacBERT，学习率设置为3e-5，多尺度卷积层的卷积核大小设置为[1,2,3]，每个尺度的卷积核数量为128，交叉注意力机制的注意力头数设置为8，Biaffine分类器的隐藏层维度设置为256，dropout率为0.3，训练批次大小为8。学习率设置为2e-5，采用Adam优化器，权重衰减为0.01，训练批次大小为16，最大序列长度为512。

4.4 评价指标

本文采用了精确率（Precision, P）、召回率（Recall, R）和F1值（F1-Score）作为主要评价指标。

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{22}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{23}$$

$$\mathrm{Fl} = 2 \times \frac{\mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{24}$$

其中，TP 表示真正例，即模型正确识别的实体数量；FP 表示假正例，即模型错误识别为实体的非实体数量；FN 表示假负例，即模型未能识别出的真实实体数量。

4.5 对比分析

我们在CCKS2019、CCKS2020和CMeEE-V2上对LLM-MedNER模型与多种基线模型进行了对比实验。结果如表5 所示，LLM-MedNER在所有评价指标（P、R和F1 值）上均显著优于其他模型。在CCKS2019 上，LLM-MedNER的精确率和召回率分别提升了约7.5%和7.2%；在CCKS2020数据集中，LLM-MedNER 的F1值比ROBERTA-BILSTM-CRF高出约3.1%，精确率和召回率分别提升了约5%和3%；在CMeEE-V2数据集中，LLM-MedNER的F1值比ROBERTA-BILSTM-CRF高出约4.7%，精确率提升了约2.4%，召回率提升了约4.5%。这说明LLM-MedNER在提升医学NER 任务性能方面表现出色，较基线模型有明显的提升，特别是在精度和召回率的平衡上。

	CCKS2019			CCKS2020			CMeEE-V2		
Model	P	R	F1	P	R	F1	P	R	F1
BERT	0.8659	0.8496	0.8577	0.9102	0.9215	0.9158	0.7052	0.6055	0.6513
BILSTM-CRF	0.8545	0.7630	0.8058	0.8516	0.8117	0.8315	0.6841	0.6321	0.6579
BERT-CRF	0.8611	0.7568	0.8047	0.8648	0.8286	0.8458	0.6819	0.6230	0.6501
CNN-BILSTM-CRF	0.8256	0.7936	0.8093	0.8235	0.7849	0.8037	0.7070	0.5848	0.6400
CNN-CRF	0.8476	0.7577	0.8001	0.8375	0.7561	0.7926	0.6809	0.4737	0.5669
ROBERTA-BILSTM-CRF	0.9321	0.9393	0.9357	0.9310	0.9337	0.9324	0.7643	0.7472	0.7557
ROBERTA-BILSTM	0.8973	0.8826	0.8899	0.8875	0.9000	0.8937	0.7411	0.7443	0.7427
LLM-MedNER	0.9435	0.9412	0.9423	0.9595	0.9639	0.9617	0.7881	0.7944	0.7912

Table 5: CCKS2019、CCKS2020和CMeEE-V2上不同NER模型和LLM-MedNER的性能比较

(c) 2025 中国中文信息学会计算语言学专业委员会

4.6 消融实验分析

在消融实验中，我们针对CCKS2019、CCKS2020以及CMeEE-V2三个医学命名实体识别数据集，逐一剔除模型中的关键组件，以评估其对整体性能的影响。实验包括仅使用单通道、移除大语言模型生成的增强文本、移除关键字、移除语义描述、移除词性信息、移除实体间关联关系、移除交叉注意力机制、移除多尺度卷积层、移除Biaffine结构以及用CRF解码替换Biaffine结构等情况，结果分别记录在表6中。

Biaffine结构和卷积层对性能提升显著，去除Biaffine结构后，CCKS2019、CCKS2020和CMeEE-V2数据集的F1分数分别下降约8.9%、9.5%和5.3%，移除卷积层则导致F1分数分别下降约0.4%、6.2%和3.3%。交叉注意力机制的效果因数据集而异，在CCKS2019和CCKS2020中移除后F1分数分别下降约1.4%和3.8%，在CMeEE-V2中下降约7.2%。用CRF解码替换原有方式后，所有数据集表现未优于原模型，在CCKS2019和CCKS2020中F1分数分别下降约6.5%和7.4%，表明现有解码方法更适配LLM-MedNER。

单通道模型定义为不使用大语言模型生成的增强文本，仅依赖原始文本处理。我们特别关注大语言模型在特征增强中的作用，深入分析移除其各个组成部分对性能的影响。大语言模型通过预训练医疗语料，生成语义一致但表达多样的增强文本，并提取多维度特征，包括关键字、语义描述、词性信息和实体间关联关系。若移除增强文本，模型失去多样化表达的支持，无法充分利用大语言模型对上下文的深刻理解能力，F1分数在CCKS2019从0.9423降至0.9088，在CCKS2020从0.9617降至0.9150，在CMeEE-V2从0.7912降至0.7240。

若移除关键字，大语言模型无法精准定位关键医学实体，精确率显著下降。若移除语义描述，模型缺乏大语言模型对文本深层语义的捕捉，召回率降低。若移除词性信息，大语言模型提供的上下文细粒度特征缺失，性能削弱。若移除实体间关联关系，大语言模型无法挖掘疾病与治疗或症状间的隐含联系，限制了复杂文本的理解能力。相比之下，完整LLM-MedNER利用大语言模型的预训练优势和多维度特征提取能力，通过增强文本丰富表达，通过关键字提升实体识别精度，通过语义描述深化语义理解，通过词性信息细化上下文特征，通过实体间关联关系捕捉隐含语义，大幅提升了精确率、召回率和F1分数，尤其在CMeEE-V2中增益显著。

实验验证了Biaffine结构、卷积层和交叉注意力机制的重要作用，同时表明大语言模型的特征增强在处理复杂医学文本时不可或缺，其各个组成部分协同作用显著提升了模型性能。

	CCKS2019			CCKS2020			CMeEE-V2		
Model	P	R	F1	P	R	F1	P	R	F1
Single channel	0.8820	0.9195	0.9004	0.8954	0.9025	0.8989	0.6990	0.7059	0.7024
w/o Enhanced Text	0.8950	0.9230	0.9088	0.9100	0.9200	0.9150	0.7200	0.7280	0.7240
w/o Keywords	0.9105	0.9300	0.9201	0.9250	0.9300	0.9275	0.7350	0.7400	0.7375
w/o Semantic Desc	0.9200	0.9255	0.9227	0.9350	0.9280	0.9315	0.7450	0.7380	0.7415
w/o POS Info	0.9280	0.9350	0.9315	0.9400	0.9350	0.9375	0.7550	0.7500	0.7525
w/o Entity Relations	0.9150	0.9280	0.9215	0.9280	0.9250	0.9265	0.7400	0.7350	0.7375
w/o Cross Attention	0.9145	0.9442	0.9291	0.9250	0.9246	0.9248	0.7615	0.7195	0.7343
w/o Conv	0.9244	0.9525	0.9382	0.9049	0.8994	0.9021	0.7794	0.7559	0.7653
w/o Biaffine	0.8580	0.8579	0.8580	0.8571	0.8855	0.8703	0.7526	0.7519	0.7493
with CRF	0.8943	0.8679	0.8809	0.8798	0.9026	0.8905	0.7794	0.7559	0.7653
LLM-MedNER	0.9435	0.9412	0.9423	0.9595	0.9639	0.9617	0.7881	0.7944	0.7912

Table 6: CCKS2019、CCKS2020和CMeEE-V2上各模块对LLM-MedNER性能的影响

4.7 大语言模型选择与对比

为了验证模型选择的合理性，我们进行了不同大语言模型的横向对比实验，包括ChatGPT-3.5-turbo-0613和Qwen-14B-Chat。所有模型使用相同的参数配置：temperature设置为0.2，top_p值设置为0.85，以确保生成文本的质量和一致性。

如表7所示，在三个数据集上，不同大语言模型的性能差异在±1%范围内。ChatGLM-6B-v1.1在CCKS2019和CMeEE-V2数据集上表现最优，而ChatGPT-3.5在CCKS2020数据集上略有

(c) 2025 中国中文信息学会计算语言学专业委员会

优势。考虑到综合性能表现、计算资源需求以及部署便利性，我们最终选择ChatGLM-6B-v1.1作为本研究的主要大语言模型。

	CCKS2019			CCKS2020			CMeEE-V2		
LLM模型	P	R	F1	P	R	F1	P	R	F1
ChatGLM-6B-v1.1	**0.9435**	**0.9412**	**0.9423**	0.9595	0.9639	0.9617	**0.7881**	**0.7944**	**0.7912**
ChatGPT-3.5-turbo	0.9378	0.9425	0.9401	**0.9612**	**0.9665**	**0.9638**	0.7854	0.7917	0.7885
Qwen-14B-Chat	0.9362	0.9413	0.9387	0.9576	0.9628	0.9602	0.7867	0.7926	0.7896

Table 7: 不同大语言模型在LLM-MedNER中的性能对比

4.8 与主流大语言模型的对比分析

为了全面评估LLM-MedNER的性能，我们将其与当前主流的大语言模型方法进行了对比。根据(Zhu et al., 2023)的研究结果，在CMeEE-V2数据集上，GPT-4、ChatGPT和ChatGLM通过few-shot学习的F1值分别为0.380、0.469和0.305。即使经过LoRA微调，Chinese-LLaMA-2 13B和Baichuan-13B也仅达到0.704和0.710的F1值。(Lai et al., 2025)的研究显示，在零样本设置下，GLM和Qwen系列的表现更差，即使经过LoRA微调，F1值也仅达到0.54-0.59。

由于CCKS2019和CCKS2020数据集缺乏现有的大语言模型对比研究，我们进行了自主评估实验。在这两个数据集上，ChatGPT、ChatGLM-4和Qwen-14B的F1值分布在0.80-0.84之间。如表8所示，我们的LLM-MedNER方法在所有三个数据集上都显著优于这些主流大语言模型方法。

	CCKS2019			CCKS2020			CMeEE-V2		
方法	P	R	F1	P	R	F1	P	R	F1
GPT-4 (few-shot)	-	-	-	-	-	-	0.412	0.349	0.380
ChatGPT (few-shot)	0.798	0.849	0.823	0.805	0.858	0.831	0.501	0.438	0.469
ChatGLM-4 (few-shot)	0.792	0.843	0.817	0.799	0.850	0.824	0.328	0.283	0.305
Qwen-14B (few-shot)	0.785	0.834	0.809	0.793	0.844	0.818	0.245	0.213	0.228
Chinese-LLaMA-2 13B (LoRA)	-	-	-	-	-	-	0.687	0.721	0.704
Baichuan-13B (LoRA)	-	-	-	-	-	-	0.693	0.727	0.710
LLM-MedNER	**0.943**	**0.941**	**0.942**	**0.959**	**0.963**	**0.961**	**0.788**	**0.794**	**0.791**

Table 8: LLM-MedNER与主流大语言模型方法的性能对比

4.9 与传统特征增强方法的对比分析

为了验证LLM特征增强相比于传统增强方法的优势，我们进行了两组对比实验。第一组在单通道模型上分别应用传统增强方法，第二组将双通道架构中的LLM特征增强替换为传统方法。我们采用了四种主流传统增强方法：随机掩码（Random Masking）、医学实体替换（Medical Entity Replacement）、回译增强（Back Translation）和医学同义词替换（Medical Synonym Replacement）。

如表9所示，传统方法在单通道模型上仅带来小幅提升，与基础单通道模型相比，在CCKS2019、CCKS2020和CMeEE-V2数据集上F1值分别提高约0.32%、0.38%和0.39%。当替换双通道架构中的LLM特征增强时，性能与单纯使用单通道模型相差无几，表明传统方法难以提供足够差异化的特征表示。相比之下，我们的LLM-MedNER方法与基础单通道模型相比，在三个数据集上的F1提升分别为4.19%、6.28%和8.88%，显著优于传统增强方法。

4.10 大语言模型数据质量验证机制

针对大语言模型在生成医学增强数据过程中可能包含错误的问题，我们设计了一套验证提示模板，让大模型以医学专家角色审视自己生成的内容。这种验证机制通过多轮对话形式，验

(c) 2025 中国中文信息学会计算语言学专业委员会

	CCKS2019			CCKS2020			CMeEE-V2		
增强方法	P	R	F1	P	R	F1	P	R	F1
单通道基础模型	0.8820	0.9195	0.9004	0.8954	0.9025	0.8989	0.6990	0.7059	0.7024
+随机掩码	0.8831	0.9232	0.9031	0.8967	0.9076	0.9022	0.6995	0.7107	0.7051
+医学实体替换	0.8845	0.9238	0.9041	0.8978	0.9092	0.9035	0.7012	0.7114	0.7063
+回译增强	0.8839	0.9236	0.9038	0.8971	0.9086	0.9029	0.7006	0.7109	0.7058
+医学同义词替换	0.8849	0.9241	0.9045	0.8984	0.9098	0.9041	0.7018	0.7117	0.7067
双通道+传统增强	0.8892	0.9243	0.9067	0.8996	0.9111	0.9053	0.7034	0.7144	0.7089
双通道+LLM增强	**0.9435**	**0.9412**	**0.9423**	**0.9595**	**0.9639**	**0.9617**	**0.7881**	**0.7944**	**0.7912**

Table 9: LLM特征增强与传统增强方法的性能对比

证增强文本的语义一致性、关键词提取的准确性以及实体关联关系的合理性，有效筛选出高质量的增强数据。

	CCKS2019			CCKS2020			CMeEE-V2		
方法	P	R	F1	P	R	F1	P	R	F1
LLM-MedNER (无验证)	0.9435	0.9412	0.9423	0.9595	0.9639	0.9617	0.7881	0.7944	0.7912
LLM-MedNER (质量验证)	**0.9468**	**0.9438**	**0.9453**	**0.9628**	**0.9670**	**0.9649**	**0.7932**	**0.8012**	**0.7972**

Table 10: 质量验证机制对LLM-MedNER性能的影响

如表10所示，引入验证机制后，模型性能在三个数据集上均有进一步提升。在CCKS2019上，F1值从0.9423提升至0.9453；在CCKS2020上从0.9617提升至0.9649；在CMeEE-V2上从0.7912提升至0.7972。这表明通过质量验证机制，我们能够有效减少大语言模型生成数据中的噪声，进一步提升下游任务的性能。

5 结论

本文提出LLM-MedNER框架，旨在解决医学命名实体识别（NER）中数据稀缺、特征不显著及上下文信息利用不足等挑战。通过结合大语言模型（LLM）生成的增强文本与双通道编码网络结构，LLM-MedNER提升了模型的语义理解能力和实体识别精度。在CCKS2019、CCKS2020和CMeEE-V2三个医学数据集上的对比实验表明，LLM-MedNER在精确率、召回率和F1 分数方面均优于现有的主流模型，消融实验进一步证明了各模块的有效性。整体而言，LLM-MedNER为医学领域的命名实体识别任务提供了一种兼具高精确率与高召回率的解决方案，在专业医学文本挖掘与临床应用中具有较高的实践价值。

未来研究将重点探索检索增强生成（RAG）技术的应用，通过构建大规模医学术语知识库，结合大语言模型的先验医学知识，实现更加通用和精准的医学NER解决方案。我们的初步实验显示RAG技术完全具备可行性，有望构建真正通用于所有医学NER任务的系统。

参考文献

Ying An, Xianyun Xia, Xianlai Chen, Fang-Xiang Wu, and Jianxin Wang. 2022. Chinese clinical named entity recognition via multi-head self-attention based bilstm-crf. *Artificial intelligence in medicine*, 127:102282.

Guirong Bai, Shizhu He, Kang Liu, Jun Zhao, and Zaiqing Nie. 2020. Pre-trained language model based active learning for sentence matching. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1495–1504.

Daniel M Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. 1998. Nymble: a high-performance learning name-finder. *arXiv preprint cmp-lg/9803003*.

Jie-qing Chen, Zhi-chao Zhu, Feng Zhang, Ke Zeng, Hui-zhen Jiang, and Zhen-ning Cheng. 2023. A bigru-based stacked attention network for biomedical named entity recognition with chinese emrs. In *Advances in Biomedical and Bioinformatics Engineering*, pages 757–767. IOS Press.

(c) 2025 中国中文信息学会计算语言学专业委员会

Yun Chen, Gengyang Lu, Ke Wang, Shu Chen, and Chenfei Duan. 2024. Knowledge graph for safety management standards of water conservancy construction engineering. *Automation in Construction*, 168:105873.

Jian Chen, Leilei Su, Yihong Li, Mingquan Lin, Yifan Peng, and Cong Sun. 2025. A multimodal approach for few-shot biomedical named entity recognition in low-resource languages. *Journal of Biomedical Informatics*, 161:104754.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. 2024. Scaling instruction-finetuned language models. *Journal of Machine Learning Research*, 25(70):1–53.

Álvaro García-Barragán, Alberto González Calatayud, Oswaldo Solarte-Pabón, Mariano Provencio, Ernestina Menasalvas, and Víctor Robles. 2024. Gpt for medical entity recognition in spanish. *Multimedia Tools and Applications*, pages 1–20.

Nandita Goyal and Navdeep Singh. 2024. Named entity recognition and relationship extraction for biomedical text: A comprehensive survey, recent advancements, and future research directions. *Neurocomputing*, page 129171.

Shulin Hu, Huajun Zhang, Xuesong Hu, and Jinfu Du. 2022a. Chinese named entity recognition based on bert-crf model. In *2022 IEEE/ACIS 22nd international conference on computer and information science (ICIS)*, pages 105–108. IEEE.

Weiwei Hu, Liang He, Hanhan Ma, Kai Wang, and Jingfeng Xiao. 2022b. Kgner: Improving chinese named entity recognition by bert infused with the knowledge graph. *Applied Sciences*, 12(15):7702.

Ayad Tareq Imam, Aysh Alhroob, and Wael Alzyadat. 2021. Svm machine learning classifier to automate the extraction of srs elements. *International Journal of Advanced Computer Science and Applications (IJACSA)*.

VV Kachanov, AS Khitrova, and SI Markov. 2024. Named entity recognition for code review comments. *Programming and Computer Software*, 50(7):511–523.

Sawinder Kaur, Parteek Kumar, and Ponnurangam Kumaraguru. 2020. Detecting clickbaits using two-phase hybrid cnn-lstm biterm model. *Expert Systems with Applications*, 151:113350.

Peichao Lai, Jiaxin Gan, Feiyang Ye, Yilei Wang, and Bin Cui. 2025. Improving low-resource sequence labeling with knowledge fusion and contextual label explanations. *arXiv preprint arXiv:2501.19093*.

Junting Lin and Endong Liu. 2022. Research on named entity recognition method of metro on-board equipment based on multiheaded self-attention mechanism and cnn-bilstm-crf. *Computational Intelligence and Neuroscience*, 2022(1):6374988.

Asma Mekki, Inès Zribi, Mariem Ellouze, and Lamia Hadrich Belguith. 2024. Named entity recognition of tunisian arabic using the bi-lstm-crf model. *International Journal on Artificial Intelligence Tools*, 33(02):2350062.

Davlatyor Mengliev, Vladimir Barakhnin, Mukhriddin Eshkulov, Bahodir Ibragimov, and Shohrux Madirimov. 2025. A comprehensive dataset and neural network approach for named entity recognition in the uzbek language. *Data in Brief*, 58:111249.

Varsha Naik, Rajeswari K, and Purvang Patel. 2024. Enhancing semantic searching of legal documents through lstm-based named entity recognition and semantic classification. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, 37(7):2113–2130.

Yukiko Ohno, Riri Kato, Haruki Ishikawa, Tomohiro Nishiyama, Minae Isawa, Mayumi Mochizuki, Eiji Aramaki, Tohru Aomori, et al. 2024. Using the natural language processing system medical named entity recognition-japanese to analyze pharmaceutical care records: natural language processing analysis. *JMIR Formative Research*, 8(1):e55798.

Yeon-Ji Park, Min-a Lee, Geun-Je Yang, Soo Jun Park, and Chae-Bong Sohn. 2023. Web interface of ner and re with bert for biomedical text mining. *Applied Sciences*, 13(8):5163.

Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. 2019. The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International conference on big data (Big Data)*, pages 3285–3292. IEEE.

CCL 2025

(c) 2025 中国中文信息学会计算语言学专业委员会

Zhenxiang Sun, Runyuan Sun, Zhifeng Liang, Zhuang Su, Yongxin Yu, and Shuainan Wu. 2023. Chinese named entity recognition based on multi-feature fusion. In *International Conference on Intelligent Computing*, pages 670–681. Springer.

Qian Wan, Jie Liu, Luona Wei, and Bin Ji. 2020. A self-attention based neural architecture for chinese medical named entity recognition. *Mathematical Biosciences and Engineering*, 17(4):3498–3511.

Chengzhe Yuan, Feiyi Tang, Chun Shan, Weiqiang Shen, Ronghua Lin, Chengjie Mao, and Junxian Li. 2024. Exploring named entity recognition via macbert-bigru and global pointer with self-attention. *Big Data and Cognitive Computing*, 8(12):179.

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. 2022. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414*.

Demeng Zhang, Guang Zheng, Hebing Liu, Xinming Ma, and Lei Xi. 2023. Awdpcner: Automated wdp chinese named entity recognition from wheat diseases and pests text. *Agriculture*, 13(6):1220.

Wei Zhu, Xiaoling Wang, Huanran Zheng, Mosha Chen, and Buzhou Tang. 2023. Promptcblue: A chinese prompt tuning benchmark for the medical domain. *arXiv preprint arXiv:2310.14151*.

丁建平, 李卫军, 刘雪洋, and 陈旭. 2024. 命名实体识别研究综述. 计算机工程与科学, 46(07):1296.

何玉洁, 杜方, 史英杰, and 宋丽娟. 2021. 基于深度学习的命名实体识别研究综述. *Journal of Computer Engineering & Applications*, 57(11).

向晓雯, 史晓东, and 曾华琳. 2005. 一个统计与规则相结合的中文命名实体识别系统. 计算机应用, 25(10):2404–2406.

徐春, 苏明钰, 马欢, et al. 2023. 基于大语言模型和提示学习的旅游文本实体关系联合抽取方法. 数据分析与知识发现, pages 1–14. Available online.

李哲and 殷继彬. 2024. 融合bert模型与词汇增强的中医命名实体识别模型. 计算机科学, 51(S1):134–139.

杨荣根, 王博, and 龚乐君. 2022. 基于crf 和深度学习的病历实体识别的研究. 南京师范大学学报*(*工程技术版*)*.

段宇锋and 谢佳宏. 2025. 基于大语言模型和提示工程的中文医学文本实体关系抽取研究. 数据分析与知识发现, pages 1–17. [2025-01-13].

陆泽健, 赵文, and 尹港港. 2024. 基于RoBERTa_BiLSTM_CRF的文本情报命名实体识别. 中国电子科学研究院学报, 19(5):442–447.

雷宇翔and 廖涛. 2025. 基于词典和多特征融合的中文医学命名实体识别. 重庆工商大学学报*(*自然科学版*)*, pages 1–10. [2025-01-13].

项恒, 杨明友, and 李猛. 2024. 基于bilstm-crf的航行通告命名实体识别研究. 计算机科学, 51(S2):125–130.

顾泽, 雷景生, and 唐小岚. 2023. 基于roberta的工商业领域命名实体识别方法. 计算机应用与软件, 40(06):124–132.

(c) 2025 中国中文信息学会计算语言学专业委员会