A Survey of Medical Entity Recognition and Extraction Techniques in Biomedical Text Mining

Abstract

This survey explores the pivotal role of Medical Entity Recognition (MER) and related extraction techniques in enhancing biomedical informatics through the integration of deep learning and natural language processing (NLP) methodologies. MER facilitates the transformation of unstructured clinical narratives into structured data, thereby improving data retrieval processes and patient care outcomes. The survey highlights the advancements in named entity recognition (NER) and relation extraction (RE), emphasizing the superiority of deep learning models, such as transformer-based architectures, in handling complex biomedical texts. Despite these advancements, challenges persist, including data scarcity, model generalization, and ethical considerations. The survey underscores the necessity for continued research into innovative learning techniques and the integration of multimodal data to address these limitations. Furthermore, the survey discusses the implications of MER in enhancing clinical decision support systems, advancing drug discovery, and promoting personalized medicine. It calls for the development of robust ethical frameworks to mitigate biases and ensure the equitable application of MER technologies. The conclusion emphasizes the need for ongoing innovation and collaboration to overcome existing barriers, thereby enhancing the capabilities of MER systems and contributing to improved healthcare outcomes.

1 Introduction

1.1 Significance of Medical Entity Recognition

Medical entity recognition (MER) is essential in biomedical text mining, transforming unstructured clinical narratives into structured data, thereby enhancing data retrieval in healthcare [1]. The surge in biomedical literature complicates information extraction from extensive unstructured texts, a challenge MER addresses by extracting structured information from clinical notes, including patient histories and the relationships between medical conditions and medications, ultimately improving patient care and clinical outcomes [2].

MER's primary application lies in electronic health records (EHRs), where it identifies adverse drug events (ADEs) and supports clinical decision-making by correlating patient information with scientific literature [3]. Furthermore, MER extracts social determinants of health (SDOH) from clinical narratives, which significantly influence health outcomes, thereby facilitating advancements in healthcare systems through improved diagnostic tools and treatment methodologies [4].

Additionally, MER aids in predicting and managing chronic diseases, such as diabetes, which are leading causes of morbidity and mortality globally. Medical entity linking (MEL) ensures accurate normalization and standardization of various mentions of medical entities within clinical texts, crucial for effective data analysis and healthcare delivery [5]. MER also plays a role in preserving traditional medical knowledge, as evidenced by its application in extracting information from traditional Uyghur medical texts, highlighting its cultural and clinical significance [6].

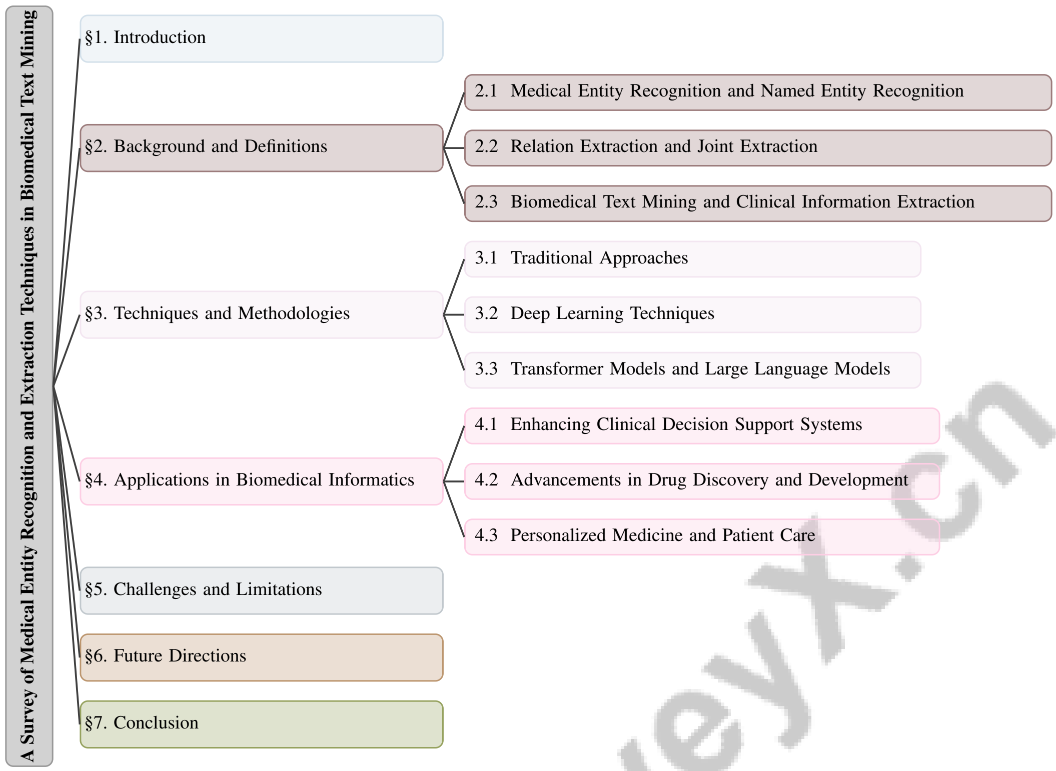

Figure 1: chapter structure

The impact of MER extends to enhancing biomedical text mining systems, improving the precision and relevance of extracted information, thereby acting as a cornerstone of biomedical informatics and a catalyst for advancing healthcare systems through improved clinical information extraction and decision-making processes [2].

1.2 Role of Deep Learning and NLP

Deep learning and natural language processing (NLP) have transformed clinical information extraction, significantly improving the processing and analysis of complex biomedical texts. The integration of advanced deep learning models with multimodal data from EHRs shows promise in predicting chronic disease risks, indicating the potential of these technologies to enhance healthcare outcomes [7]. Developing intelligent informatics tools and algorithms is crucial for automating the processing of large text volumes, facilitating valuable knowledge extraction from clinical data [8].

Deep learning approaches have established superiority over traditional methods regarding accuracy and efficiency, particularly in named entity recognition (NER), with transformer-based models enhancing medical information extraction from clinical narratives [9]. The BERT language model exemplifies this advancement, improving feature extraction capabilities in NLP tasks [1].

NLP techniques are widely used in EHRs to automatically extract ADE-related information through relation extraction methods [10]. The synergy between structured knowledge bases and advanced language models is illustrated by integrating the Unified Medical Language System (UMLS) with generative pre-trained transformer (GPT) models, enhancing clinical information extraction [11].

Generative clinical large language models (LLMs), such as GatorTronGPT, streamline clinical NLP tasks through a unified text-to-text learning architecture, improving clinical information extraction [12]. Despite advancements, challenges remain, including the complexity of biomedical language and the need for high-quality annotated datasets, underscoring the necessity for ongoing research and innovation [5]. The benchmark for instruction tuning effectiveness further highlights the role of refined methodologies in enhancing general LLM performance on biomedical NLP tasks, indicating the continuous evolution of these technologies [6].

The application of machine learning in healthcare systems focuses on enhancing diagnostic accuracy and treatment efficacy. As deep learning and NLP methodologies progress, they promise to further improve clinical information extraction, paving the way for more sophisticated healthcare solutions [4].

1.3 Structure of the Survey

This survey is organized to provide a comprehensive overview of advancements in medical information extraction, emphasizing deep learning techniques, named entity recognition (NER), and relation extraction (RE) [13]. The structure guides the reader through biomedical text mining stages, from data pre-processing to information retrieval, entity recognition, relation extraction, and section detection, each characterized by unique methodologies and challenges [14].

We begin with an introduction that highlights the significance of medical entity recognition and the transformative role of deep learning and NLP in clinical information extraction. This section explores foundational concepts and terminologies related to biomedical information extraction, focusing on critical tasks such as medical entity recognition and the joint extraction of entities and relationships, while illustrating how advanced NLP techniques convert unstructured medical texts into structured, actionable information, addressing challenges in the medical domain and enhancing decision-making for healthcare professionals [15, 14, 16].

The core of the survey delves into techniques and methodologies employed in medical entity recognition and extraction, encompassing traditional approaches and modern deep learning techniques, with a focus on transformer models and large language models. These sections highlight the nuances of NER and RE methods, providing a structured categorization of existing research [17].

We also examine the practical applications of advanced text mining techniques in biomedical informatics, showcasing their contributions to improving clinical decision support systems through enhanced semantic relation extraction, accelerating drug discovery via Literature Based Discovery (LBD) methodologies, and promoting personalized medicine by effectively structuring information from unstructured medical documents. These innovations streamline crucial medical knowledge retrieval and reduce the time needed to uncover novel associations in extensive scientific literature, leading to improved patient outcomes and minimized medical errors [18, 14, 19].

The survey addresses challenges and limitations in medical entity recognition and extraction, including data scarcity, annotation costs, model generalization, and ethical considerations. These discussions frame the current landscape of clinical NER and RE, while intentionally excluding unrelated NLP tasks and techniques not specifically applicable to clinical texts [20].

We conclude by outlining critical future research directions aimed at enhancing information extraction. These include advancing innovative learning techniques, integrating multimodal and domain-specific data to improve the accuracy and relevance of extracted information, and developing comprehensive strategies for ensuring ethical compliance and mitigating bias in automated systems. Emphasizing the importance of cross-sentence relation extraction and the application of deep learning models in specialized domains such as medicine, we highlight how NLP can significantly enhance question answering and knowledge discovery [21, 22, 16, 23, 24]. The conclusion synthesizes key insights and underscores the importance of ongoing research and innovation to advance biomedical informatics and improve healthcare outcomes.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Medical Entity Recognition and Named Entity Recognition

Medical Entity Recognition (MER) and Named Entity Recognition (NER) are fundamental to biomedical text mining, each playing a distinct role in extracting information from medical literature. MER targets the identification and association of medical concepts like diseases, treatments, and biological processes within unstructured texts, converting them into structured formats crucial for clinical diagnostics and drug development. The complexity of MER stems from the specialized terminologies and abbreviations common in biomedical literature, necessitating advanced methodologies for effective management [5].

NER, on the other hand, focuses on detecting and classifying entities into categories such as proteins, genes, diseases, and treatments, which is essential for interaction extraction. This task is complicated by challenges like polysemy and unique biochemical naming conventions, requiring sophisticated techniques for precise recognition [10]. NER is pivotal in the broader information extraction process, enabling the retrieval of structured information from vast amounts of unstructured text data [25].

A notable advancement in MER involves integrating relation extraction (RE), which enhances the understanding of complex biomedical texts by simultaneously detecting and classifying entities and their interrelations. This is particularly significant in clinical settings, where precise extraction of clinical entities and their attributes can significantly influence healthcare delivery and research outcomes [3]. The extraction of adverse drug events (ADEs) from unstructured clinical texts exemplifies MER's specialized nature, highlighting the limitations of traditional methods and the need for advanced extraction techniques [26].

While MER and NER are integral to information extraction, MER is more specialized, focusing on medical entities and their integration into clinical workflows, which is crucial for supporting healthcare delivery and advancing biomedical informatics [17]. The challenges of accurately recognizing biomedical entities, including entity boundaries and classification, underscore the complexity and necessity of these tasks in medical informatics [5]. Efforts to refine general large language models (LLMs) for specific biomedical tasks like NER and RE illustrate ongoing advancements in these methodologies [6]. The integration of NER and relation classification (RC) as key components of MER remains a research focus, as evidenced by the fine-tuning of models like BERT for joint entity and relation extraction tasks [1].

2.2 Relation Extraction and Joint Extraction

Relation extraction (RE) is crucial in biomedical text mining, focusing on identifying and classifying semantic relationships between biomedical entities within unstructured scientific or clinical texts [27]. This process is essential for constructing structured data formats from complex biomedical information, facilitating applications like drug discovery and disease pathway analysis. Traditional RE approaches often target binary relations represented as triplets; however, the intricate nature of biomedical texts requires advanced methodologies capable of handling document-level and cross-sentence relationships [28].

A significant challenge in RE is the error propagation that occurs when NER and RE are performed separately, potentially leading to inaccuracies in the final extraction results [3]. The presence of noisy training data due to distant supervision and the long-tail distribution of fact triples further complicates the extraction process, necessitating robust techniques to accurately capture relationships between biomedical entities [29].

Joint extraction models, which perform NER and RE simultaneously, offer a promising solution by reducing error propagation and enhancing the accuracy of extracted relationships [1]. These models are particularly advantageous in the biomedical domain, where the complexity of relationships between entities demands a comprehensive extraction approach [30]. Nonetheless, challenges remain, including the development of lightweight and efficient models capable of processing the unstructured and complex nature of clinical texts [12].

Advancements in RE and joint extraction methodologies are vital for harnessing the full potential of biomedical data, enhancing the extraction of valuable insights from complex medical texts, and improving clinical decision support systems while contributing to the creation of comprehensive medical knowledge graphs. As the field progresses, addressing challenges related to multi-modal, cross-lingual, and document-level relation extraction will be crucial for continued technological advancement [28].

2.3 Biomedical Text Mining and Clinical Information Extraction

Biomedical text mining and clinical information extraction are pivotal for converting unstructured clinical data into structured semantic representations, significantly enhancing the precision and efficiency of healthcare delivery and research [9]. These processes are crucial for managing the exponential growth of biomedical literature, which often contains complex language and specialized terminologies that can overwhelm healthcare professionals and hinder effective decision-making

AI-generated, for reference only.

[8]. By leveraging advanced analytical methods, including machine learning and deep learning techniques, these technologies facilitate the extraction of valuable clinical information, underscoring their importance in modern healthcare applications [4].

A critical aspect of biomedical text mining is the normalization of biomedical entities, which addresses the variation problem in entity mentions and improves clinical information extraction. This task ensures consistency in the identification and classification of biomedical entities across diverse datasets, contributing to the generation of synthetic electronic health records (EHRs) that enhance clinical information extraction [2]. Moreover, integrating multimodal data types enhances predictive accuracy for chronic diseases, showcasing the potential of these technologies to advance precision medicine [7].

The extraction of structured information from clinical narratives is vital for constructing coherent patient care timelines, thereby enhancing clinical decision support systems by improving the precision of case-based retrieval under time constraints faced by clinicians. This capability is particularly important in oncology, where detailed accounts of medical problems and drug events are extracted from clinical notes to inform patient management and treatment strategies [9]. Additionally, understanding interactions between species, such as those involving the gut microbiome, highlights the diverse applications of text mining techniques in biomedical research, supporting the transition towards personalized patient care [4].

As these technologies continue to evolve, addressing challenges such as data scarcity and the integration of multimodal data will be essential for advancing biomedical text mining and clinical information extraction. These advancements are critical for improving healthcare outcomes and driving innovation in medical research and practice, ultimately supporting the development of personalized medicine and enhancing patient care [4].

3 Techniques and Methodologies

Category	Feature	Method
Traditional Approaches	Pattern Recognition	PBKE[31], BIEP[32]
Deep Learning Techniques	Task-Specific Enhancement	GTP[12], Ad-MTL[33]
	Model Combination Strategies	HDL-CRE[34]
	Knowledge Integration	BiOnt[27], KRC[35]
	Relation Extraction Optimization	BiTT[30]
Transformer Models and Large Language Models	Attention Mechanisms	AMIL[29], FAM[1]

Table 1: This table provides a comprehensive summary of various methodologies employed in medical entity recognition and extraction. It categorizes the techniques into traditional approaches, deep learning techniques, and transformer models, highlighting specific methods and their corresponding features. The table serves as a comparative analysis of the evolution and adoption of advanced methods in biomedical text mining.

In the realm of medical entity recognition and extraction, the evolution of techniques has been marked by a transition from traditional methodologies to more advanced approaches. This shift reflects the growing complexity of biomedical texts and the need for more effective tools to navigate their intricacies. Table 3 offers an organized overview of the methods and features associated with different approaches to medical entity recognition and extraction, illustrating the progression from traditional to modern deep learning and transformer-based techniques. The subsequent subsection will delve into the traditional approaches that laid the groundwork for current practices, examining their methodologies, strengths, and limitations, as well as their role in shaping the landscape of biomedical text mining.

3.1 Traditional Approaches

Traditional methodologies for medical entity recognition and extraction have been instrumental in the early development of biomedical text mining, utilizing a range of rule-based and statistical techniques. Rule-based systems, which depend on manually crafted rules and lexicons, were initially effective for extracting specific entity types such as genes and diseases. These systems often required extensive domain-specific knowledge, making them laborious to construct and challenging to adapt to new contexts [31]. Despite their precision in certain cases, rule-based methods lack flexibility and

AI-generated, for reference only.

scalability, particularly when faced with the variability and complexity inherent in biomedical texts [15].

Statistical methods, such as Conditional Random Fields (CRF) and Support Vector Machines (SVM), have been widely employed for tasks like named entity recognition (NER) and relation extraction (RE). These methods leverage linguistic features and statistical patterns to identify entities and their relationships within texts. For example, CRF models have been used to capture the contextual dependencies in sequence labeling tasks, while SVMs have been applied to classify relationships between biomedical entities. However, these approaches often rely on hand-crafted features, which can struggle with the noise and complexity of unstructured biomedical texts [15].

The BIEP framework exemplifies a systematic approach that integrates NER and RE to extract biological relationships from unstructured text, specifically targeting protein-protein interactions [32]. Despite their foundational role, traditional methods face limitations in handling the intricate and non-linear nature of biomedical data. They often assume linear relationships and can be less effective in dealing with noisy and unstructured clinical narratives [15].

As illustrated in Figure 2, traditional approaches in biomedical text mining can be categorized into rule-based systems and statistical methods, with the figure highlighting their limitations and the comparative advantages of machine learning techniques. Comparative analyses have shown that machine learning methods, particularly those based on neural networks, generally outperform traditional approaches for NER tasks. This is due to their ability to learn from data without relying on extensive feature engineering [36]. As the field progresses, the limitations of traditional methods in scalability, flexibility, and accuracy have paved the way for the adoption of more advanced methodologies, such as deep learning and neural networks, which offer improved performance with fewer feature dependencies [23]. These advancements highlight the ongoing evolution of techniques in medical entity recognition and extraction, as researchers continue to seek more robust and efficient solutions [37].

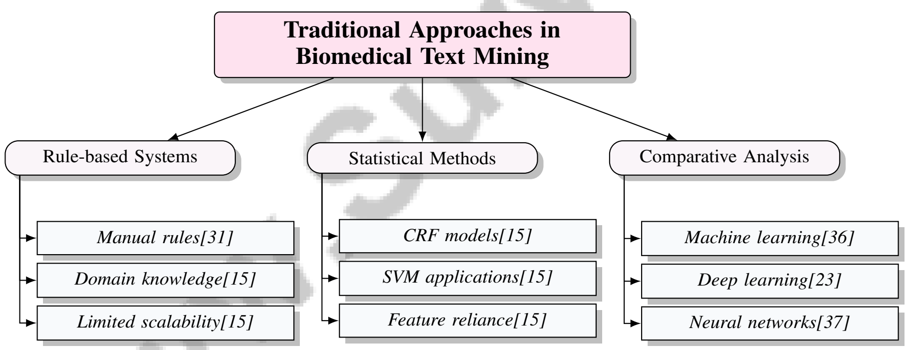

Figure 2: This figure illustrates the traditional approaches in biomedical text mining, categorizing them into rule-based systems and statistical methods, and highlighting their limitations and the comparative advantages of machine learning techniques.

3.2 Deep Learning Techniques

Method Name	Integration Techniques	Model Architectures	Application Domains
Ad-MTL[33]	Adversarial Training	Bi-GRU	Clinical Nlp
FAM[1]	Joint Learning Frameworks	Bert Language Model	Clinical Nlp Tasks
BiOnt[27]	Multi-channel Architecture	Lstm Networks	Clinical Nlp Tasks
AMIL[29]	Multi-instance Learning	Relationship Embedding Architecture	Biomedical Relation Extraction
GTP[12]	Multi-task Learning	Generative Llm	Clinical Nlp Tasks
BiTT[30]	-	-	Medical Relation Extraction
KRC[35]	Adversarial Training	Bert	Clinical Nlp
HDL-CRE[34]	Hybrid Deep Learning	Bert And Others	Semantic Search

Table 2: Overview of deep learning methods applied to medical entity recognition, highlighting integration techniques, model architectures, and application domains. The table lists various methods, including Ad-MTL, FAM, BiOnt, and others, detailing their specific approaches and target areas within clinical NLP and biomedical relation extraction.

Deep learning techniques have revolutionized medical entity recognition (MER), offering sophisticated frameworks that significantly enhance the accuracy and efficiency of extracting entities and their relationships from complex biomedical texts. These techniques address the intricacies of biomedical language, characterized by specialized terminologies and complex entity interrelations. The integration of multi-task learning with adversarial training has been shown to improve relation extraction by sharing representations across tasks while minimizing contamination from task-specific features, thereby enhancing overall MER system performance [33].

Joint extraction models, which simultaneously perform named entity recognition and relation extraction within a unified framework, represent a notable advancement in deep learning for MER. Such models mitigate error propagation and enhance the accuracy of extracted information, as demonstrated by the focused attention model (FAM) that integrates the BERT language model to improve feature extraction for NER and relation classification tasks [1]. The use of modern deep learning architectures, including BioGPT and BioBART, alongside general-domain models like GPT-2 and T5, exemplifies the adaptability and effectiveness of these techniques in handling complex biomedical texts [38].

The application of bidirectional Long Short-Term Memory (LSTM) networks, as implemented in systems like BiOnt, further enhances the extraction of relations between biomedical entities by integrating multiple biomedical ontologies [27]. This integration of ontological knowledge with deep learning models facilitates the accurate identification of biomedical relationships, crucial for applications such as drug discovery and disease pathway analysis. Moreover, the introduction of abstractified multi-instance learning (AMIL) enhances the denoising process during training by grouping entities by their corresponding UMLS semantic types [29].

State-of-the-art models like BERT, Flan-T5, Llama3, and GPT-4 have demonstrated significant improvements in MER tasks, showcasing superior performance across various applications [9]. The instruction-tuned models have shown remarkable enhancements in performance on biomedical tasks compared to their foundational counterparts, underscoring the transformative impact of deep learning on medical entity recognition [6]. Additionally, the GERNNERMED++ model highlights the potential of transfer learning in enhancing MER tasks across different languages and domains [26].

The integration of large language models, such as GatorTronGPT, into clinical NLP tasks exemplifies the modern deep learning approach to MER, streamlining the process of extracting and understanding clinical information [12]. Innovative tagging schemes like BiTT, which organize medical relation triples into binary trees, contribute to advancements in relation extraction by converting these structures into word-level tags [30]. Furthermore, the benchmark introducing self-questioning prompting (SQP) enhances model performance by generating informative questions and answers relevant to clinical scenarios, thereby improving task-specific performance [25].

As deep learning techniques continue to evolve, they promise to further advance biomedical informatics by improving the accuracy and efficiency of clinical information extraction, ultimately contributing to enhanced healthcare outcomes. The ongoing development of lightweight and efficient models tailored for clinical tasks, alongside the integration of knowledge representation and machine reading comprehension, underscores the potential of deep learning to revolutionize medical entity recognition [35]. Current research has led to significant improvements in diagnostic accuracy and the development of personalized treatment plans, showcasing the potential of machine learning in healthcare [4].

As shown in **??**, in exploring the diverse techniques and methodologies within the realm of deep learning, it is essential to consider the illustrative examples provided in Figure **??**. The first example, "Binary Relation vs. N-Ary Relation," highlights the distinction between binary and n-ary relations, where the former is depicted as a straightforward connection between two nodes, while the latter involves multiple lines connecting the same nodes to signify various relationships. This visualization underscores the complexity and flexibility of n-ary relations in capturing intricate data relationships. The second example, "Convolution Layer and Vector Representation in Sentence Analysis," offers insight into the application of convolutional layers in processing textual data. Here, a convolution layer is represented as a series of blocks corresponding to words within a sentence, with a vector representation that encapsulates the sentence's features. The dimensions of this vector (d$_c = 3), (l = 3), and(d =$

$5) are crucial for understanding how deep learning models analyze and interpret linguistic structures. Together, these ex$
$39]. Additionally, Table 2 provides a comprehensive summary of the deep learning techniques employed in medical entity r$

3.3 Transformer Models and Large Language Models

Transformer models and large language models (LLMs) have significantly advanced the field of medical entity recognition and relation extraction by providing sophisticated frameworks that enhance semantic understanding and contextual embedding. The BERT model, a prominent transformer-based architecture, leverages attention mechanisms to capture complex interrelations within biomedical texts, thereby improving the accuracy of entity extraction and relational mapping. This is achieved through its dynamic range attention mechanism in the focused attention model (FAM), which allows the model to concentrate on task-specific contexts, thereby enhancing representation capabilities [1].

The integration of multi-head attention mechanisms, as demonstrated in models like CoEx-Bert, enhances the extraction process by enabling the model to simultaneously focus on multiple aspects of the input, thereby improving the capture of intricate relationships within the data. Furthermore, the combination of Conditional Random Fields (CRF) with multi-head selection mechanisms in joint extraction models offers a robust solution to the challenges of biomedical text mining [29]. These innovations underscore the transformative impact of transformer models in capturing long-range dependencies, significantly improving the accuracy of relation extraction.

Large language models, including GPT-3.5, GPT-4, and Bard, have demonstrated significant potential in processing clinical texts by effectively integrating both local and structural information. These models excel in various clinical natural language processing tasks, such as span identification, token-level sequence classification, and relation extraction, even when not specifically trained for the clinical domain. Recent studies indicate that models like GPT-3 outperform existing benchmarks in few-shot clinical information extraction, achieving state-of-the-art results across multiple tasks. The unified architecture of these models, exemplified by GatorTronGPT, showcases their ability to handle diverse clinical NLP challenges through prompt tuning, further solidifying their role as versatile tools in the analysis of clinical data. [40, 12]. The use of compact language models, developed through Knowledge Distillation and continual learning techniques, exemplifies the adaptability and efficiency of LLMs in enhancing biomedical entity recognition. Additionally, the incorporation of structured knowledge bases, such as the Unified Medical Language System (UMLS), with GPT models highlights a novel approach to improve extraction accuracy by enhancing entity recognition.

The ongoing development of transformer models and LLMs, alongside innovations like the introduction of a multiple label strategy (MLS) that captures label correlations without the complexity of traditional CRF methods, underscores their potential to revolutionize medical entity recognition and relation extraction in the biomedical domain. As these models continue to evolve, they promise to provide more accurate and efficient solutions for clinical information extraction, ultimately advancing biomedical informatics and improving healthcare outcomes. The integration of multiple tasks across diverse biomedical and clinical datasets enables a thorough assessment of task interdependencies, which not only enhances the predictive capabilities of models but also facilitates automated knowledge extraction. This is particularly evident in relation extraction tasks, where understanding semantic relationships—such as drug-drug interactions and protein-protein interactions—requires leveraging knowledge from interconnected entities. By employing a multi-task learning (MTL) framework, models can simultaneously learn from various tasks, resulting in improved performance metrics across the board, as demonstrated by advancements in structured self-attentive networks and retrieval-augmented techniques. Such comprehensive approaches ultimately contribute to more efficient and accurate information extraction in the rapidly evolving biomedical literature landscape. [33, 41, 42, 43]

4 Applications in Biomedical Informatics

In biomedical informatics, advanced medical entity recognition (MER) technologies are integral to enhancing healthcare processes, particularly within clinical decision support systems (CDSS). MER facilitates the extraction and integration of crucial clinical information from unstructured texts, thereby improving decision-making and patient safety. Figure 3 illustrates the hierarchical structure of applications in biomedical informatics, emphasizing the contributions of MER not only to CDSS but also to advancements in drug discovery and personalized medicine. This diagram outlines key

Feature	Traditional Approaches	Deep Learning Techniques	Transformer Models and Large Language Models
Methodology	Rule-based, Statistical	Multi-task, Adversarial	Attention, Multi-head
Strengths	Precision, Domain-specific	Accuracy, Efficiency	Semantic Understanding
Limitations	Scalability, Flexibility	Complexity, Training Data	Training Complexity

Table 3: This table provides a comparative analysis of methodologies, strengths, and limitations across traditional approaches, deep learning techniques, and transformer models in the context of medical entity recognition and extraction. It highlights the evolution of methods from rule-based and statistical systems to advanced deep learning and transformer-based frameworks, showcasing their respective advantages and challenges in handling complex biomedical texts.

technological tools, frameworks, and evaluation methods that facilitate these advancements, thereby providing a comprehensive view of how MER optimizes clinical workflows and enhances the delivery of high-quality patient care. This subsection explores these facets, underscoring the pivotal role of MER in transforming healthcare practices.

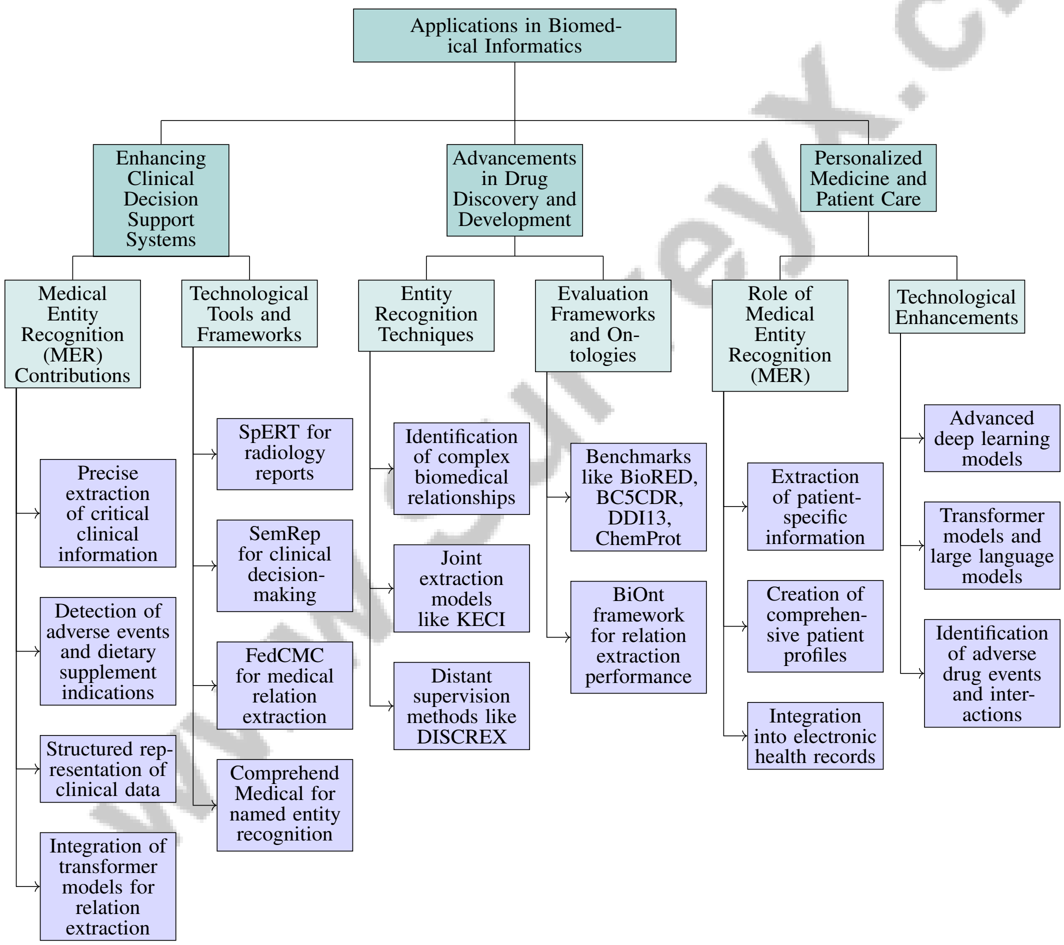

Figure 3: This figure depicts the hierarchical structure of applications in biomedical informatics, highlighting the contributions of medical entity recognition (MER) to enhancing clinical decision support systems, advancements in drug discovery and development, and the role of MER in personalized medicine and patient care. The diagram outlines key technological tools, frameworks, and evaluation methods that facilitate these advancements.

4.1 Enhancing Clinical Decision Support Systems

Medical entity recognition (MER) significantly enhances clinical decision support systems (CDSS) by enabling precise extraction of critical clinical information from unstructured texts. Advanced deep learning models effectively detect adverse events and dietary supplement indications within clinical notes, thus improving pharmacovigilance and patient safety [44, 45]. Structured representation of clinical data, as exemplified by models like SpERT, links accurate anatomical information to radiological findings, enhancing the usability of radiology reports critical for clinical decision support [46, 47].

The extraction of medication and temporal relations through advanced models underscores the potential for automation in CDSS, facilitating timely decision-making [48]. The integration of transformer models into clinical relation extraction tasks improves understanding of complex clinical relationships, thereby enhancing patient care outcomes [49]. Tools like SemRep illustrate the impact of MER in clinical decision-making, while federated learning approaches, such as FedCMC, have shown superior performance in medical relation extraction tasks, enhancing efficiency and accuracy [50, 51].

Recent advancements in natural language processing (NLP) pipelines for adverse drug reaction (ADR) detection highlight the scalability and real-world applicability of these technologies in pharmacovigilance, achieving state-of-the-art accuracy across multiple datasets [52]. Continuous evaluation of tools like Comprehend Medical in performing named entity recognition and relationship extraction in clinical texts further demonstrates the evolving capabilities of these systems in supporting clinical decision-making [53].

The RAMIE framework improves CDSS by efficiently extracting dietary supplement information from clinical narratives [41], while the CoEx-Bert model enhances extraction accuracy and efficiency [54]. The CACER benchmark aims to enhance clinical decision support applications through fine-grained annotations that clarify relationships in oncology care [9].

Integrating large language models (LLMs) into CDSS exemplifies the potential of these technologies. The LLM-IE package bridges advanced LLM technology with practical biomedical NLP applications, significantly improving information extraction processes [55]. Labrak et al. demonstrate that LLMs effectively capture medical knowledge and perform well in zero- and few-shot scenarios, indicating their potential role in enhancing CDSS [56].

The incorporation of advanced MER technologies, including the Unified Medical Language System (UMLS) and state-of-the-art deep learning models like Generative Pre-trained Transformers (GPT), into CDSS represents a significant advancement in healthcare delivery. These technologies enhance the precise extraction of medical concepts and relationships from electronic health records (EHRs), improving the retrieval of relevant medical literature and facilitating informed clinical decisions. By leveraging the contextual understanding capabilities of models like GPT alongside UMLS's structured knowledge, healthcare providers can achieve higher precision in identifying entities and relations, ultimately leading to improved patient outcomes and more effective healthcare services. Tools such as Comprehend Medical exemplify the scalability and accessibility of these technologies, enabling both developers and non-developers to utilize sophisticated named entity recognition and relationship extraction functionalities [11, 18, 53].

As illustrated in Figure 4, the key components enhancing clinical decision support systems (CDSS) are highlighted, focusing on Medical Entity Recognition (MER), Relation Extraction, and the integration of Large Language Models (LLMs). This figure underscores the role of advanced models and tools in improving the precision and efficiency of CDSS.

4.2 Advancements in Drug Discovery and Development

The integration of advanced entity recognition techniques in drug discovery and development has significantly improved the identification and understanding of complex biomedical relationships crucial for developing new therapeutic interventions. By leveraging datasets focused on chemical-protein and chemical-disease relations, researchers can extract meaningful entity relationships essential for drug discovery processes [57]. Joint extraction models, such as KECI, have demonstrated enhanced performance in extracting biomedical entities and their interrelations, facilitating the identification of potential drug targets and interactions [58].

Distant supervision methods like DISCREX have proven effective in cross-sentence relation extraction, increasing the yield of unique interactions while maintaining high accuracy, particularly beneficial in drug discovery where novel interactions can lead to innovative therapeutic strategies [22].

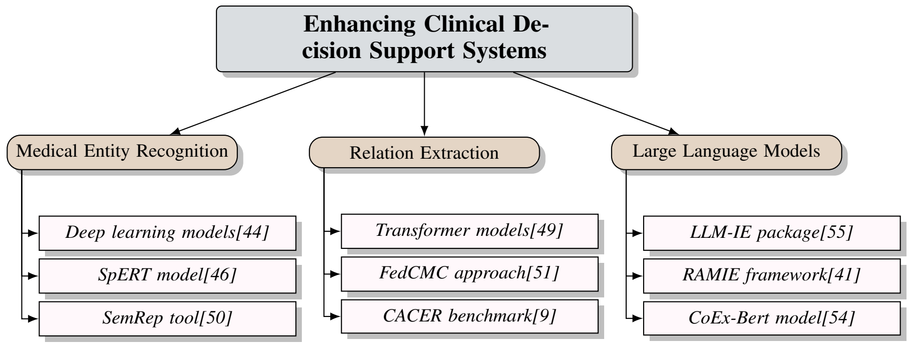

Figure 4: This figure illustrates the key components enhancing clinical decision support systems (CDSS), focusing on Medical Entity Recognition (MER), Relation Extraction, and the integration of Large Language Models (LLMs). It highlights the role of advanced models and tools in improving the precision and efficiency of CDSS.

Benchmarks like BioRED, BC5CDR, DDI13, and ChemProt provide a comprehensive evaluation framework that aids in advancing research in underexplored biomedical domains [59].

Entity recognition techniques also play a pivotal role in identifying beneficial drug combinations, as demonstrated by benchmarks designed to extract information about drug efficacy from scientific literature [60]. These benchmarks facilitate model comparison and advance relation extraction tasks critical for drug discovery [61]. The KRC framework has achieved competitive F1 scores on datasets like CDR and CHR, underscoring its effectiveness in drug discovery contexts [35].

Furthermore, integrating knowledge from biomedical ontologies, exemplified by the BiOnt framework, has led to significant improvements in relation extraction performance, achieving notable F-score enhancements across various datasets [27].

Advancements in entity recognition techniques have streamlined the drug discovery process and provided a robust foundation for developing effective and targeted therapeutic interventions. Techniques in biomedical informatics, particularly those related to Literature Based Discovery (LBD) and information extraction from electronic medical documents, are rapidly advancing. These methods leverage natural language processing and deep learning to extract meaningful insights from the vast and growing body of biomedical literature, facilitating novel associations between medical concepts and enhancing decision-making in precision oncology. They offer promising directions for future research and development, particularly in improving knowledge discovery efficiency and reducing medical errors associated with information overload [14, 62, 19].

4.3 Personalized Medicine and Patient Care

Medical entity recognition (MER) is crucial in advancing personalized medicine and patient care by enabling the extraction of detailed, patient-specific information from unstructured clinical texts. This capability is essential for tailoring healthcare treatments to individual profiles, enhancing the precision and effectiveness of medical interventions [2]. By accurately identifying and categorizing medical entities such as diseases, symptoms, and treatments, MER creates comprehensive patient profiles that inform personalized treatment plans [5].

Integrating MER into electronic health records (EHRs) allows for seamless extraction and normalization of patient data, crucial for developing personalized medicine strategies [4]. This process ensures healthcare providers can access and analyze relevant patient information efficiently, leading to informed decision-making and improved patient outcomes [7]. Additionally, extracting social determinants of health (SDOH) from clinical narratives highlights MER's broader impact on patient care by considering factors influencing health outcomes beyond traditional clinical parameters [3].

Advanced deep learning models enhance MER capabilities, allowing for the extraction of complex relationships between medical entities critical for personalized treatment approaches [1]. The use of transformer models and large language models (LLMs) in MER tasks has significantly improved the accuracy and efficiency of entity recognition, contributing to more precise personalized medicine solutions [12].

Moreover, MER facilitates the identification of potential adverse drug events (ADEs) and drug-drug interactions, vital for ensuring patient safety and optimizing treatment regimens [10]. By leveraging MER capabilities, healthcare providers can anticipate and mitigate potential risks associated with personalized treatment plans, thereby enhancing overall patient care quality [25].

As MER technologies evolve, their integration into clinical workflows promises to further advance personalized medicine and patient care, equipping healthcare professionals with tools to deliver tailored, patient-centric treatments that improve health outcomes and enhance patient satisfaction [4].

5 Challenges and Limitations

The development of medical entity recognition (MER) systems is significantly challenged by data scarcity and annotation costs, impacting the performance of machine learning models in the biomedical domain. This section delves into these challenges and their implications for clinical information extraction and MER advancements.

5.1 Data Scarcity and Annotation Costs

Data scarcity and high annotation costs are critical barriers in MER system development, especially in the biomedical field. Creating extensive labeled datasets for training complex models requires substantial resources and domain expertise, complicated by the informal and unstructured nature of electronic health records (EHRs) [2]. This scarcity of high-quality annotated datasets obstructs the training of machine learning models essential for accurate clinical information extraction [4].

The complexity of nested entity recognition and the limitations of current language models in identifying medical concepts further complicate annotation, leading to extraction errors. Additionally, the need for multiple task-specific models complicates training and deployment, revealing inefficiencies in existing clinical NLP approaches [12]. Current benchmarks often focus on a narrow range of tasks or rely on supervised learning, failing to capture the complexities of clinical language or the diverse capabilities of large language models (LLMs) [25]. This is particularly evident in biomedical tasks where general LLMs underperform due to insufficient task-specific adaptations [6].

Innovative strategies such as distant supervision and semi-supervised learning are being explored to reduce dependency on large-scale labeled datasets while maintaining annotation accuracy. Nonetheless, challenges like noisy knowledge from databases such as the CTD and unclear contextual relations remain significant hurdles [29]. Ongoing advancements aim to enhance MER systems' performance despite these constraints, striving for improved accuracy and efficiency in biomedical information extraction [30].

5.2 Model Generalization and Scalability

Model generalization and scalability pose substantial challenges in MER and relation extraction due to the complex and diverse nature of biomedical texts. The variability in medical language constrains models' ability to generalize across datasets and contexts, hindering effective information extraction [63]. Class imbalance and the generalization of models to underrepresented classes exacerbate these challenges, affecting relation extraction performance in clinical settings [64].

The reliance on specific architectures, such as BERT, limits models' adaptability to other domains, raising scalability concerns [65]. Although BERT-based models show promise in various NLP tasks, their dependence on architecture-specific features can hinder broader applicability, necessitating more flexible models for diverse biomedical data [64]. Scalability issues are compounded by the computational complexities of processing large biomedical data volumes. As data dimensionality increases, models face longer training times and resource constraints, limiting practical applications. Integrating external knowledge bases also presents scalability challenges when such resources are not readily available or comprehensive [64].

Innovative approaches are essential to enhance model robustness and scalability, ensuring MER and relation extraction systems significantly advance biomedical informatics and improve healthcare outcomes. Enhancing model generalization and scalability is crucial for overcoming current limitations and achieving more effective medical information extraction [63].

5.3 Ethical Considerations and Bias

Deploying MER models in healthcare requires careful consideration of ethical issues and potential biases. A major concern is the high error rates in large language models (LLMs), including hallucinations and biases, which can compromise their safe use in clinical settings [66]. Such errors may lead to incorrect information extraction, jeopardizing patient safety and the reliability of clinical decision support systems.

Bias in MER models often stems from training data inaccuracies, resulting in skewed outputs that do not accurately reflect real-world clinical diversity [67]. The lack of diverse datasets representing various patient demographics and medical conditions can yield models that perform well on certain populations but poorly on others. This potential for biased outputs underscores the need for strategies ensuring MER models are equitable and inclusive, providing accurate information across diverse clinical contexts.

Data privacy and model explainability are also critical ethical considerations in deploying MER systems [68]. Handling sensitive patient data requires stringent privacy measures to protect against unauthorized access and misuse. Ensuring AI models' explainability is vital for gaining healthcare professionals' trust, enabling them to understand and interpret model decisions, thereby facilitating informed clinical decision-making.

Ongoing research focuses on methodologies that enhance MER models' transparency and accountability, including bias mitigation techniques like adversarial training and fairness constraints to reduce biased training data impact and improve model robustness. Establishing ethical guidelines and regulatory frameworks is crucial for the responsible development and deployment of MER technologies in healthcare. Such frameworks will ensure MER systems are effectively integrated into clinical workflows, enhancing the extraction and structuring of critical medical information from unstructured clinical texts, such as electronic medical records (EMRs). This integration will improve patient care by minimizing medical errors and supporting informed decision-making, facilitating advancements in medical research through more efficient data utilization and analysis [69, 14, 70].

6 Future Directions

6.1 Advancements in Learning Techniques

The progression of medical entity recognition (MER) and relation extraction is poised for significant enhancement through innovative learning techniques aimed at improving robustness, efficiency, and scalability. Advances in attention mechanisms, particularly the focused attention model (FAM), promise to enhance the identification of complex relationships within biomedical texts, thereby refining entity recognition and relation extraction, especially in resource-limited settings [1]. The integration of the BiTT scheme with robust rule restrictions and advanced pre-trained encoders is anticipated to elevate accuracy and efficiency in relation extraction tasks [30]. Furthermore, embedding domain-specific knowledge into relation extraction techniques is expected to yield contextually relevant outputs across diverse biomedical scenarios [9].

The amalgamation of abstractified multi-instance learning (AMIL) with sophisticated bag aggregation methods aims to enhance the denoising process during training, facilitating the application of these architectures to general-domain datasets and thereby improving the generalizability of MER models across various biomedical tasks [29]. Moreover, expanding methodologies to encompass additional clinical domains and assessing the clinical validity of generated texts are crucial for broadening the applicability of MER systems [2]. Future research should focus on developing robust datasets and enhancing model interpretability while addressing ethical considerations in machine learning applications in healthcare. These advancements promise to significantly improve the accuracy and efficiency of clinical information extraction, thereby contributing to better healthcare outcomes and advancing biomedical informatics [6].

6.2 Integration of Multimodal and Domain-Specific Data

The integration of multimodal and domain-specific data holds substantial potential for enhancing MER and relation extraction models. By leveraging diverse data sources, including clinical narratives, imaging data, and genomic information, these models can achieve a comprehensive understanding of complex biomedical entities and their interrelations, thus improving the robustness of MER systems and producing more accurate outputs [71]. Future research should prioritize expanding annotated corpora and incorporating longitudinal data to strengthen the robustness of relation extraction models, particularly in chronic disease management and personalized medicine contexts [72]. Developing robust evaluation frameworks for NLP models is essential to address biases and ensure ethical compliance in healthcare applications [66].

Ongoing efforts to extend models to other languages and integrate multimodal data sources promise to enhance clinical NLP capabilities. This includes creating models capable of processing and analyzing data from various modalities, thereby improving clinical decision-making and patient care [73]. The introduction of refined approaches, such as combining translation, annotation projection, and transfer learning, exemplifies the potential for developing robust multilingual medical NER models, thereby expanding the applicability of these technologies across diverse linguistic and cultural contexts [26]. Moreover, expanding datasets to encompass more diverse languages and patient populations is vital to ensure that MER models are inclusive and representative of global healthcare needs [74]. As these advancements continue, the integration of multimodal and domain-specific data is poised to revolutionize MER and relation extraction, ultimately enhancing healthcare outcomes and advancing biomedical informatics.

6.3 Ethical Compliance and Bias Mitigation

Ensuring ethical compliance and mitigating bias in MER systems is crucial for their effective deployment in clinical settings. Future research should focus on strategies to address ethical concerns and biases, particularly as the reliance on large language models (LLMs) in healthcare applications grows. The potential biases in LLMs, arising from their training data, necessitate comprehensive frameworks to ensure fairness and equity in model outputs [25]. Incorporating additional background knowledge from unsupervised data sources to enhance the contextual understanding of MER models and reduce reliance on potentially biased training datasets is a promising direction. Utilizing ontology reasoners to infer relationships within the data could further enhance model performance and mitigate bias by providing a nuanced understanding of biomedical concepts [28].

Future research should also explore alternative prompting strategies for LLMs to minimize bias and improve ethical deployment in clinical settings. Expanding MER systems to include a wider range of clinical tasks will be crucial for their applicability and effectiveness across diverse healthcare scenarios [25]. Enhancing the transparency and interpretability of MER models is vital for building trust among healthcare professionals and ensuring that model decisions are understandable and justifiable. Improving the quality and diversity of training datasets, including the use of transfer and semi-supervised learning techniques, will be critical for developing robust and unbiased MER systems. Expanding datasets to include full-text biomedical articles and integrating advanced deep learning techniques with established methods will effectively address challenges associated with biomedical literature. This approach enhances the extraction of valuable insights from complex data, promoting comprehensive and equitable performance of models across diverse biomedical applications, thereby improving tasks such as named entity recognition, relation extraction, and section identification in various medical contexts [75, 76, 24, 62, 19].

As the field progresses, developing ethical guidelines and regulatory frameworks will be essential for guiding the responsible development and deployment of MER technologies in healthcare. These initiatives will enhance the effectiveness of MER systems in improving healthcare outcomes while ensuring compliance with the highest ethical standards, fostering fairness, transparency, and accountability in their deployment. By integrating advanced Natural Language Processing techniques for extracting and analyzing clinical data, these systems can better identify relevant medical information, support informed decision-making, and address social determinants of health, ultimately leading to improved patient care and outcomes [53, 70, 77, 78, 14].

7 Conclusion

Medical entity recognition (MER) and extraction techniques are fundamental to the progression of biomedical informatics, significantly impacting healthcare delivery. The integration of deep learning and natural language processing (NLP) has notably refined the precision and efficiency of clinical data extraction. The development of benchmarks like CACER exemplifies advancements in extracting valuable medical information from clinical texts, which is essential for enhancing patient care. Despite these advancements, issues such as data quality and model interpretability persist, necessitating continuous research and innovation.

The potential of artificially generated training data to enhance NLP model performance, as highlighted by recent studies, underscores the necessity for further exploration in this domain to address data scarcity. Ensuring the reliability and effectiveness of these models in practical applications requires thorough evaluation and validation before their clinical integration.

The availability of public resources and the seamless integration of extraction and normalization processes are pivotal for achieving improved semantic representations. Progress in these areas is crucial for enhancing the accuracy and applicability of MER systems across various healthcare settings, ultimately advancing patient care and supporting medical research.

AI-generated, for reference only.

References

[1] Kui Xue, Yangming Zhou, Zhiyuan Ma, Tong Ruan, Huanhuan Zhang, and Ping He. Fine-tuning bert for joint entity and relation extraction in chinese medical text, 2019.

[2] Zixu Wang, Julia Ive, Sumithra Velupillai, and Lucia Specia. Is artificial data useful for biomedical natural language processing algorithms?, 2019.

[3] Xue Shi, Dehuan Jiang, Yuanhang Huang, Xiaolong Wang, Qingcai Chen, Jun Yan, and Buzhou Tang. Family history information extraction via deep joint learning. *BMC medical informatics and decision making*, 19:1–6, 2019.

[4] Shyni Sharaf and V. S. Anoop. An analysis on large language models in healthcare: A case study of biobert, 2023.

[5] Pir Noman Ahmad, Adnan Muhammad Shah, and KangYoon Lee. A review on electronic health record text-mining for biomedical name entity recognition in healthcare domain. In *Healthcare*, volume 11, page 1268. MDPI, 2023.

[6] Omid Rohanian, Mohammadmahdi Nouriborji, and David A. Clifton. Exploring the effectiveness of instruction tuning in biomedical language processing, 2023.

[7] Jun-En Ding, Phan Nguyen Minh Thao, Wen-Chih Peng, Jian-Zhe Wang, Chun-Cheng Chug, Min-Chen Hsieh, Yun-Chien Tseng, Ling Chen, Dongsheng Luo, Chi-Te Wang, Pei fu Chen, Feng Liu, and Fang-Ming Hung. Large language multimodal models for 5-year chronic disease cohort prediction using ehr data, 2024.

[8] Eva K Lee and Karan Uppal. Cerc: an interactive content extraction, recognition, and construction tool for clinical and biomedical text. *BMC medical informatics and decision making*, 20:1–14, 2020.

[9] Yujuan Fu, Giridhar Kaushik Ramachandran, Ahmad Halwani, Bridget T. McInnes, Fei Xia, Kevin Lybarger, Meliha Yetisgen, and Özlem Uzuner. Cacer: Clinical concept annotations for cancer events and relations, 2024.

[10] Fenia Christopoulou, Thy Thy Tran, Sunil Kumar Sahu, Makoto Miwa, and Sophia Ananiadou. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods. *Journal of the American Medical Informatics Association*, 27(1):39–46, 2020.

[11] Kriti Bhattarai, Inez Y. Oh, Zachary B. Abrams, and Albert M. Lai. Document-level clinical entity and relation extraction via knowledge base-guided generation, 2024.

[12] Cheng Peng, Xi Yang, Aokun Chen, Zehao Yu, Kaleb E Smith, Anthony B Costa, Mona G Flores, Jiang Bian, and Yonghui Wu. Generative large language models are all-purpose text analytics engines: Text-to-text learning is all your need, 2023.

[13] Udo Hahn and Michel Oleynik. Medical information extraction in the age of deep learning. *Yearbook of medical informatics*, 29(01):208–220, 2020.

[14] Mohamed Yassine Landolsi, Lobna Hlaoua, and Lotfi Ben Romdhane. Information extraction from electronic medical documents: state of the art and future research directions. *Knowledge and Information Systems*, 65(2):463–516, 2023.

[15] Surag Nair. A biomedical information extraction primer for nlp researchers, 2017.

[16] Parisa Naderi Golshan, HosseinAli Rahmani Dashti, Shahrzad Azizi, and Leila Safari. A study of recent contributions on information extraction, 2018.

[17] Nadeesha Perera, Matthias Dehmer, and Frank Emmert-Streib. Named entity recognition and relation detection for biomedical information extraction. *Frontiers in cell and developmental biology*, 8:673, 2020.

[18] Maristella Agosti, Giorgio Maria Di Nunzio, Stefano Marchesin, and Gianmaria Silvello. A relation extraction approach for clinical decision support, 2019.

[19] Balu Bhasuran, Gurusamy Murugesan, and Jeyakumar Natarajan. Literature based discovery (lbd): Towards hypothesis generation and knowledge discovery in biomedical text mining, 2023.

[20] Priyankar Bose, Sriram Srinivasan, William C Sleeman IV, Jatinder Palta, Rishabh Kapoor, and Preetam Ghosh. A survey on recent named entity recognition and relationship extraction techniques on clinical texts. *Applied Sciences*, 11(18):8319, 2021.

[21] Xiaoyan Zhao, Yang Deng, Min Yang, Lingzhi Wang, Rui Zhang, Hong Cheng, Wai Lam, Ying Shen, and Ruifeng Xu. A comprehensive survey on relation extraction: Recent advances and new frontiers, 2024.

[22] Chris Quirk and Hoifung Poon. Distant supervision for relation extraction beyond the sentence boundary, 2017.

[23] Qing Li, Lei Li, and Yu Li. Developing chatgpt for biology and medicine: A complete review of biomedical question answering, 2024.

[24] Sijia Zhou and Xin Li. Feature engineering vs. deep learning for paper section identification: Toward applications in chinese medical literature, 2024.

[25] Yuqing Wang, Yun Zhao, and Linda Petzold. Are large language models ready for healthcare? a comparative study on clinical language understanding, 2023.

[26] Johann Frei, Ludwig Frei-Stuber, and Frank Kramer. Gernermed++: Transfer learning in german medical nlp, 2022.

[27] Diana Sousa and Francisco M. Couto. Biont: Deep learning using multiple biomedical ontologies for relation extraction, 2020.

[28] Monika Jain, Kuldeep Singh, and Raghava Mutharaju. Reonto: A neuro-symbolic approach for biomedical relation extraction, 2023.

[29] William Hogan, Molly Huang, Yannis Katsis, Tyler Baldwin, Ho-Cheol Kim, Yoshiki Vazquez Baeza, Andrew Bartko, and Chun-Nan Hsu. Abstractified multi-instance learning (amil) for biomedical relation extraction, 2021.

[30] Xukun Luo, Weijie Liu, Meng Ma, and Ping Wang. A bidirectional tree tagging scheme for joint medical relation extraction, 2022.

[31] Julien Fauqueur, Ashok Thillaisundaram, and Theodosia Togia. Constructing large scale biomedical knowledge bases from scratch with rapid annotation of interpretable patterns, 2019.

[32] Jupinder Parmar, William Koehler, Martin Bringmann, Katharina Sophia Volz, and Berk Kapicioglu. Biomedical information extraction for disease gene prioritization, 2020.

[33] Shweta Yadav, Srivatsa Ramesh, Sriparna Saha, and Asif Ekbal. Relation extraction from biomedical and clinical text: Unified multitask learning framework. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(2):1105–1116, 2020.

[34] R. Priyadharshini, G. Jeyakodi, and P. Shanthi Bala. Comparative analysis of contextual relation extraction based on deep learning models, 2023.

[35] Jing Chen, Baotian Hu, Weihua Peng, Qingcai Chen, and Buzhou Tang. Biomedical relation extraction via knowledge-enhanced reading comprehension. *BMC bioinformatics*, 23(1):20, 2022.

[36] Sarah Schulz, Jurica Ševa, Samuel Rodriguez, Malte Ostendorff, and Georg Rehm. Named entities in medical case reports: Corpus and experiments, 2020.

[37] Oumaima El Khettari, Solen Quiniou, and Samuel Chaffron. Building a corpus for biomedical relation extraction of species mentions, 2023.

[38] Aviv Brokman and Ramakanth Kavuluru. How important is domain specificity in language models and instruction finetuning for biomedical relation extraction?, 2024.

[39] Shantanu Kumar. A survey of deep learning methods for relation extraction, 2017.

[40] Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. Large language models are few-shot clinical information extractors, 2022.

[41] Zaifu Zhan, Shuang Zhou, Mingchen Li, and Rui Zhang. Ramie: Retrieval-augmented multi-task information extraction with large language models on dietary supplements, 2024.

[42] Yifan Peng, Qingyu Chen, and Zhiyong Lu. An empirical study of multi-task learning on bert for biomedical text mining, 2020.

[43] Shweta Yadav, Srivatsa Ramesh, Sriparna Saha, and Asif Ekbal. Relation extraction from biomedical and clinical text: Unified multitask learning framework, 2020.

[44] Yadan Fan, Sicheng Zhou, Yifan Li, and Rui Zhang. Deep learning approaches for extracting adverse events and indications of dietary supplements from clinical text, 2020.

[45] Ying Xiong, Shuai Chen, Buzhou Tang, Qingcai Chen, Xiaolong Wang, Jun Yan, and Yi Zhou. Improving deep learning method for biomedical named entity recognition by using entity definition information. *BMC bioinformatics*, 22:1–13, 2021.

[46] Kevin Lybarger, Aashka Damani, Martin Gunn, Ozlem Uzuner, and Meliha Yetisgen. Extracting radiological findings with normalized anatomical information using a span-based bert relation extraction model, 2021.

[47] Saahil Jain, Ashwin Agrawal, Adriel Saporta, Steven QH Truong, Du Nguyen Duong, Tan Bui, Pierre Chambon, Yuhao Zhang, Matthew P. Lungren, Andrew Y. Ng, Curtis P. Langlotz, and Pranav Rajpurkar. Radgraph: Extracting clinical entities and relations from radiology reports, 2021.

[48] Hangyu Tu, Lifeng Han, and Goran Nenadic. Extraction of medication and temporal relation from clinical text using neural language models, 2023.

[49] Xi Yang, Zehao Yu, Yi Guo, Jiang Bian, and Yonghui Wu. Clinical relation extraction using transformer-based models, 2021.

[50] Halil Kilicoglu, Graciela Rosemblat, Marcelo Fiszman, and Dongwook Shin. Broad-coverage biomedical relation extraction with semrep. *BMC bioinformatics*, 21:1–28, 2020.

[51] Chunhui Du, Hao He, and Yaohui Jin. Contrast with major classifier vectors for federated medical relation extraction with heterogeneous label distribution, 2023.

[52] Hasham Ul Haq, Veysel Kocaman, and David Talby. Mining adverse drug reactions from unstructured mediums at scale, 2022.

[53] Parminder Bhatia, Busra Celikkaya, Mohammed Khalilia, and Selvan Senthivel. Comprehend medical: a named entity recognition and relationship extraction web service. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1844–1851. IEEE, 2019.

[54] Fan Lu, Quan Qi, and Huaibin Qin. Joint extraction of uyghur medicine knowledge with edge computing, 2024.

[55] Enshuo Hsu and Kirk Roberts. Llm-ie: A python package for generative information extraction with large language models, 2024.

[56] Yanis Labrak, Mickael Rouvier, and Richard Dufour. A zero-shot and few-shot study of instruction-finetuned large language models applied to clinical and biomedical tasks, 2024.

[57] Yongkang Li. An empirical study on relation extraction in the biomedical domain, 2021.

[58] Tuan Lai, Heng Ji, ChengXiang Zhai, and Quan Hung Tran. Joint biomedical entity and relation extraction with knowledge-enhanced collective inference, 2021.

[59] William Hogan and Jingbo Shang. Entangled relations: Leveraging nli and meta-analysis to enhance biomedical relation extraction, 2024.

[60] Aryeh Tiktinsky, Vijay Viswanathan, Danna Niezni, Dana Meron Azagury, Yosi Shamay, Hillel Taub-Tabib, Tom Hope, and Yoav Goldberg. A dataset for n-ary relation extraction of drug combinations, 2022.

[61] Maxime Delmas, Magdalena Wysocka, and André Freitas. Relation extraction in underexplored biomedical domains: A diversity-optimised sampling and synthetic data generation approach, 2023.

[62] Ting He, Kory Kreimeyer, Mimi Najjar, Jonathan Spiker, Maria Fatteh, Valsamo Anagnostou, and Taxiarchis Botsis. Ai-assisted knowledge discovery in biomedical literature to support decision-making in precision oncology, 2024.

[63] Vikas Yadav and Steven Bethard. A survey on recent advances in named entity recognition from deep learning models, 2019.

[64] Jue Wang and Wei Lu. Two are better than one: Joint entity and relation extraction with table-sequence encoders, 2020.

[65] Peng Su and K. Vijay-Shanker. Investigation of bert model on biomedical relation extraction based on revised fine-tuning mechanism, 2020.

[66] Reza Khanmohammadi, Mohammad M. Ghassemi, Kyle Verdecchia, Ahmed I. Ghanem, Luo Bing, Indrin J. Chetty, Hassan Bagher-Ebadian, Farzan Siddiqui, Mohamed Elshaikh, Benjamin Movsas, and Kundan Thind. An introduction to natural language processing techniques and framework for clinical implementation in radiation oncology, 2023.

[67] Yannis Papanikolaou, Marlene Staib, Justin Grace, and Francine Bennett. Slot filling for biomedical information extraction, 2022.

[68] Zabir Al Nazi and Wei Peng. Large language models in healthcare and medical domain: A review, 2024.

[69] Enwei Zhu, Qilin Sheng, Huanwan Yang, and Jinpeng Li. A unified framework of medical information annotation and extraction for chinese clinical text, 2022.

[70] Benjamin E. Nye, Jay DeYoung, Eric Lehman, Ani Nenkova, Iain J. Marshall, and Byron C. Wallace. Understanding clinical trial reports: Extracting medical entities and their relations, 2022.

[71] Knowledge-graph-enabled biomedic.

[72] Fei Li and Hong Yu. An investigation of single-domain and multidomain medication and adverse drug event relation extraction from electronic health record notes using advanced deep learning models. *Journal of the American Medical Informatics Association*, 26(7):646–654, 2019.

[73] Omid Rohanian, Mohammadmahdi Nouriborji, Hannah Jauncey, Samaneh Kouchaki, ISARIC Clinical Characterisation Group, Lei Clifton, Laura Merson, and David A. Clifton. Lightweight transformers for clinical natural language processing, 2023.

[74] Amelie Wührl and Roman Klinger. Recovering patient journeys: A corpus of biomedical entities and relations on twitter (bear), 2022.

[75] Hasham Ul Haq, Veysel Kocaman, and David Talby. Deeper clinical document understanding using relation extraction. *arXiv preprint arXiv:2112.13259*, 2021.

[76] Mehmet Efruz Karabulut and K. Vijay-Shanker. Sectioning of biomedical abstracts: A sequence of sequence classification task, 2022.

[77] Xavier Tannier, Perceval Wajsbürt, Alice Calliger, Basile Dura, Alexandre Mouchet, Martin Hilka, and Romain Bey. Development and validation of a natural language processing algorithm to pseudonymize documents in the context of a clinical data warehouse, 2023.

[78] Kevin Lybarger, Nicholas J Dobbins, Ritche Long, Angad Singh, Patrick Wedgeworth, Ozlem Ozuner, and Meliha Yetisgen. Leveraging natural language processing to augment structured social determinants of health data in the electronic health record, 2023.

AI-generated, for reference only.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.