

Named Entity Recognition in Electronic Health Records: A Methodological Review

María C. Durango¹, Ever A. Torres-Silva¹, Andrés Orozco-Duque^{1,2}

¹Grupo de Investigación e Innovación Biomédica, Instituto Tecnológico Metropolitano, Antioquia, Colombia

²Facultad de Ingenierías, Universidad de Medellín, Antioquia, Colombia

Objectives: A substantial portion of the data contained in Electronic Health Records (EHR) is unstructured, often appearing as free text. This format restricts its potential utility in clinical decision-making. Named entity recognition (NER) methods address the challenge of extracting pertinent information from unstructured text. The aim of this study was to outline the current NER methods and trace their evolution from 2011 to 2022. **Methods:** We conducted a methodological literature review of NER methods, with a focus on distinguishing the classification models, the types of tagging systems, and the languages employed in various corpora. **Results:** Several methods have been documented for automatically extracting relevant information from EHRs using natural language processing techniques such as NER and relation extraction (RE). These methods can automatically extract concepts, events, attributes, and other data, as well as the relationships between them. Most NER studies conducted thus far have utilized corpora in English or Chinese. Additionally, the bidirectional encoder representation from transformers using the BIO tagging system architecture is the most frequently reported classification scheme. We discovered a limited number of papers on the implementation of NER or RE tasks in EHRs within a specific clinical domain. **Conclusions:** EHRs play a pivotal role in gathering clinical information and could serve as the primary source for automated clinical decision support systems. However, the creation of new corpora from EHRs in specific clinical domains is essential to facilitate the swift development of NER and RE models applied to EHRs for use in clinical practice.

Keywords: Clinical Decision Support System, Electronic Health Records, Deep Learning, Natural Language Processing, Supervised Machine Learning

Submitted: April 21, 2023

Revised: July 29, 2023

Accepted: September 3, 2023

Corresponding Author

Andrés Orozco-Duque

Grupo de Investigación e Innovación Biomédica, Instituto Tecnológico Metropolitano, 050034, Calle 73 #76A-354, Medellín, Antioquia, Colombia. Tel: +573006828421, E-mail: andresorozco4302@correo.itm.edu.co (<https://orcid.org/0000-0001-8582-8015>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2023 The Korean Society of Medical Informatics

I. Introduction

An Electronic Health Record (EHR) is a digital repository of a patient's medical information. Research indicates that approximately 80% of the data within EHRs are in an unstructured format, meaning they are contained in free-text documents encoded in expressive and natural human language typically used for documenting clinical proceedings [1,2]. These data can be extracted through named entity recognition (NER) or relation extraction (RE) methods, which are crucial components of natural language processing (NLP). These tasks involve identifying, extracting, associating, and classifying clinical terms such as diseases, symptoms, treatments, tests, medications, procedures, and body parts, there-

by enabling the recognition of a range of clinical concepts [3]. The identification of concepts in medical texts is a critical aspect of clinical decision support systems, which are designed to assist healthcare personnel in making data-driven decisions that enhance the quality of healthcare services.

Several methods exist for extracting clinical information from EHRs, which can be categorized into two main types: rule/dictionary-based and machine learning-based. The former relies heavily on syntactic and semantic analyses, utilizing regular expressions or medical terms to match patterns within the EHR text. The latter can be further divided into traditional machine learning methods, deep learning methods, and graphical models [4]. Traditional machine learning methods encompass fully connected neural networks, support vector machines, decision trees, random forests, and other classifiers. These methods necessitate feature extraction steps, which are typically based on word embeddings [5]. Deep learning methods, for their part, consist of models based on convolutional and recurrent neural networks. Unlike traditional methods, these do not require a feature extraction step, but they do necessitate a substantial volume of data for training [6]. Finally, graphical models employ graphs to represent problems and utilize information from immediate neighbors. These models, which include hidden Markov models and conditional random fields, generally require prior feature extraction steps [7].

In the present study, we reviewed the existing NER and RE methods employed in the processing of EHRs. Additionally, we examined the trends in this field over the past decade. We made comparisons among the studies based on the classification method, which could be rule-based, traditional machine learning, graphical models, or deep learning. We also considered the type of corpus, whether private or public, the language used, and the tagging system.

This study encompasses manuscripts published from 2011 to 2022. Consequently, this review does not incorporate recent advancements published in 2023 that utilize large language models (LLMs) such as GPT-4 (Generative Pre-trained Transformer 4). However, some of these publications will be referenced in the discussion section. It remains imperative to review existing methodologies for NER in medical records for several reasons. First, medical records frequently contain domain-specific language, abbreviations, and acronyms. The majority of LLMs are trained on general-purpose corpora and may not effectively manage these specific challenges. Second, a review of existing methods allows for a comparison of performance, strengths, and limitations in future studies or applications that employ LLMs. Lastly,

such a review aids in identifying gaps in data availability and underscores potential avenues for future research and dataset creation.

II. Methods

In this review, we adhered as closely as possible to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (<https://prisma-statement.org/>). Our research was conducted across four databases: Science Direct, PubMed, IEEE, and the Biblioteca Virtual en Salud (VHL; <https://bvsalud.org/en/>). We began by identifying the keywords to construct the search string. These included terms such as “text mining,” “data mining,” “natural language processing,” “electronic health record,” and “named entity recognition.” During this initial phase (search patterns), we also incorporated synonyms or acronyms for each keyword. For example, for text mining, we included “text data mining,” “text analytics,” “text analysis,” and “text clustering.” For natural language processing, we included NLP. For electronic health records, we included “electronic medical records,” “EHR,” “EMR,” and “medical records.” For named entity recognition, we included “NER,” “named entity recognition,” and “classification,” and “NERC.” We then combined these keywords to formulate the queries, which are displayed in Appendix A.

The inclusion criteria were limited to papers published within the timeframe of January 2011 to December 2022. Furthermore, we focused solely on original articles. We utilized Rayyan (<https://rayyan.ai>), a free platform, to oversee the literature review process and to identify and eliminate any duplicate articles.

Subsequently, we screened titles and abstracts to exclude articles based on the subsequent criteria (utilizing the labels provided in parentheses):

- The study does not use EHRs (No EHRs).
- The study reports the application of NLP but does not mention any NER method (No NER).
- The study is unrelated to NLP (No NLP).
- The study is not reported as an original research paper, for example review articles or conference proceedings (Not Original).
- The paper is written in a language other than English (Language).

In cases where there was a discrepancy concerning any article, we ascertained whether the publication adhered to the exclusion criteria, using the information provided in the title and abstract.

Next, we screened the full texts. During full-text screening, we manually extracted the following information to describe the articles:

- Clinical domain: We identified the types of healthcare services utilized for the extraction of EHRs.
- Corpus language: We identified the language used in the development of the NER model.
- Corpus availability: We ascertained whether the study utilized private or public corpora. Additionally, we determined if the corpora were sourced from any NLP challenges.
- Tagging system: We explored the various tagging systems employed for the identification of tokens within entities that consist of multiple words.
- NLP approach: We classified the NER models into various types or approaches. These include rule-based, traditional machine learning, deep learning, and graphical models.

III. Results

Figure 1 illustrates the procedure of our methodological review. Initially, we pinpointed 588 articles, from which 152 were selected for inclusion in this study. It is important to note that a single article may be attributed to multiple reasons for exclusion. We discovered numerous articles that utilized NLP approaches, but in certain instances, they did not implement NER methods. Furthermore, we identified several articles that included medical text, but this was obtained either from social media or through web scraping.

1. Classification Models

Figure 2 shows a timeline of the evolution of the NER approaches applied to EHRs, with this evolution being grounded in classification models. For instance, until 2016, rule-based methods, traditional machine learning methods, and graphical models were predominantly utilized. The support vector machine was the most frequently used traditional machine learning algorithm in the review [8-18]. In terms of graphical models, the conditional random field (CRF) was the most prevalent, primarily employed to address a label sequencing issue through NER tagging. Furthermore, the CRF model was commonly used in hybrid models or as an additional layer in the output of other models. Specifically, CRF was integrated with rule-based approaches [9,16,19-23], deep learning methods [4,24,25], and the conditional Markov model [26].

The first papers to report on NER models, based on deep learning and applied to EHRs, were published in 2015 [27]. By 2019, bidirectional long short-term memory (BiLSTM) had become the dominant architecture [7,28-45]. That same year, three studies were published that utilized the bidirectional encoder representations from transformers (BERT) architecture [28,43,46]. By 2021 [47-54], the BERT architecture and its variants had emerged as the primary NER model applied to EHRs, a trend that continues to this day [4,24,25,55-71]. However, this self-attention mechanism was initially introduced in 2017 [72].

Over the years, researchers have adapted various versions of the BERT model for use with EHRs. One such adaptation is BioBERT, a language representation model specifically pre-trained for the biomedical domain. This model utilizes

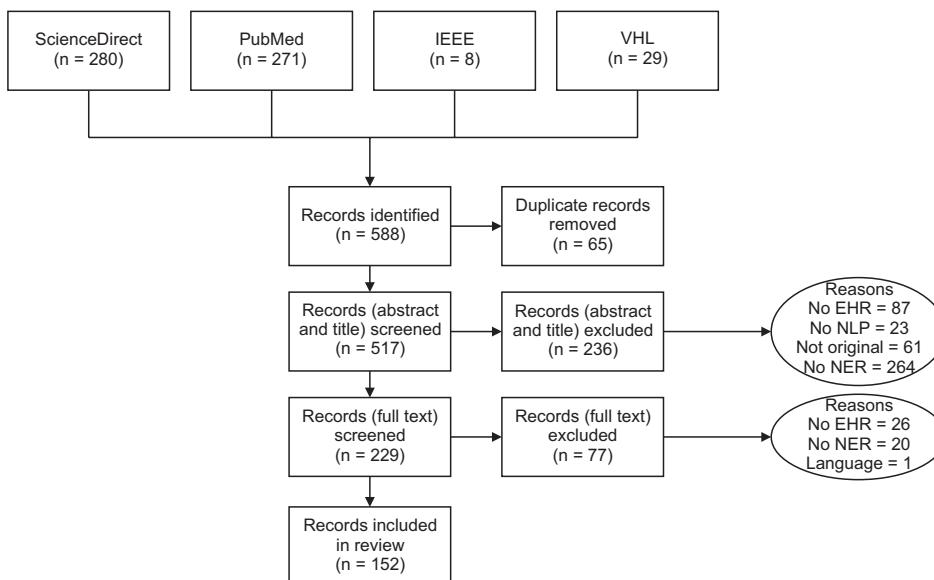


Figure 1. Flow diagram of the methodological review process. EHR: Electronic Health Record, NLP: natural language processing, NER: named entity recognition.

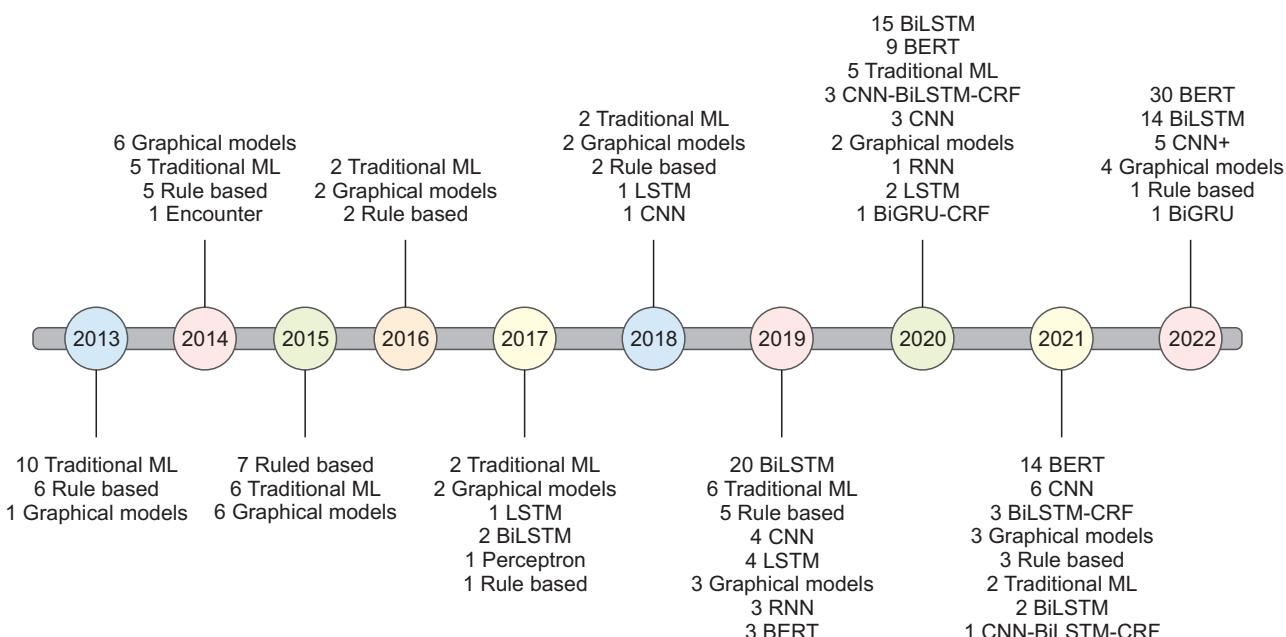


Figure 2. Timeline of named entity recognition models. ML: machine learning, LSTM: long short-term memory, BiLSTM: bidirectional long short-term memory, CNN: convolutional neural network, CRF: conditional random field, RNN: recurrent neural network, BiGRU: bidirectional gated recurrent unit, BERT: bidirectional encoder representations from transformers.

the original BERT code and has been pre-trained using PubMed abstracts and PubMed Central full-text articles [73]. BioBERT has demonstrated good performance in biomedical NER [3,64,65]. Another example is BioClinicalBERT, which was initialized using BioBERT weights and was additionally pre-trained on the Medical Information Mart for Intensive Care (MIMIC-III) datasets [74]. MIMIC-III represents the largest freely available resource of hospital data. In addition to BioBERT and BioClinicalBERT, BlueBERT [75] is another BERT variant used for EHRs. This model was pre-trained using PubMed abstracts and clinical notes, with the aim of improving the capture of language features in the biomedical and clinical domains. This could potentially lead to enhanced performance [62]. Beyond BiLSTM and BERT, several other notable deep learning models have been explored, including convolutional neural networks (CNNs) [31,44,76-83], and the hybrid CNN-BiLSTM-CRF model [84-86]. These alternative approaches have been applied in various contexts and have been demonstrated to be particularly effective for Chinese corpora [87]. As of 2022, BERT-based models are leading the field in NER applications within electronic health records. Notably, BlueBERT has emerged as a prominent solution, while BioClinicalBERT and BioBERT have also gained popularity.

2. Tagging Systems

Figure 3 details the types of tags used in the NER studies

included in this review. Such tagging systems help to represent the position of tokens within entities. The BIO system, an acronym for Beginning, Inside, and Outside, is the most commonly employed tagging system [55]. To illustrate, in the BIO format, “B” signifies that the word marks the beginning of the entity, “I” denotes that the word is within the entity but not at the start, and “O” indicates that the word does not belong to an entity. In this context, the NER task is centered on token classification, using data that have been labeled through sequence models functioning within a multiple-input, multiple-output system. Tagging systems also prove beneficial in identifying informative labels and understanding their meaning in related contexts.

Only a handful of studies have employed the BIESO or BIOES format. In this context, “B” stands for “begin,” “I” for “inside,” “E” for “end,” “S” for “single,” and “O” for “outside” or not an entity. An instance of the BIOES format is documented in a study where the researchers combined the attention mechanism with a deep learning methodology to suggest an enhanced clinical NER method for Chinese EHRs [3]. For this purpose, they identified five categories of entities: anatomical part, symptom, description, independent symptom, drug, and operation. In a similar vein, another tagging system, known as BILOU, has been employed in recent studies, including [64]. In this system, “B” signifies “begin,” “I” stands for “inside,” “L” for “last,” “O” for “outside,” and “U” for “unit.”

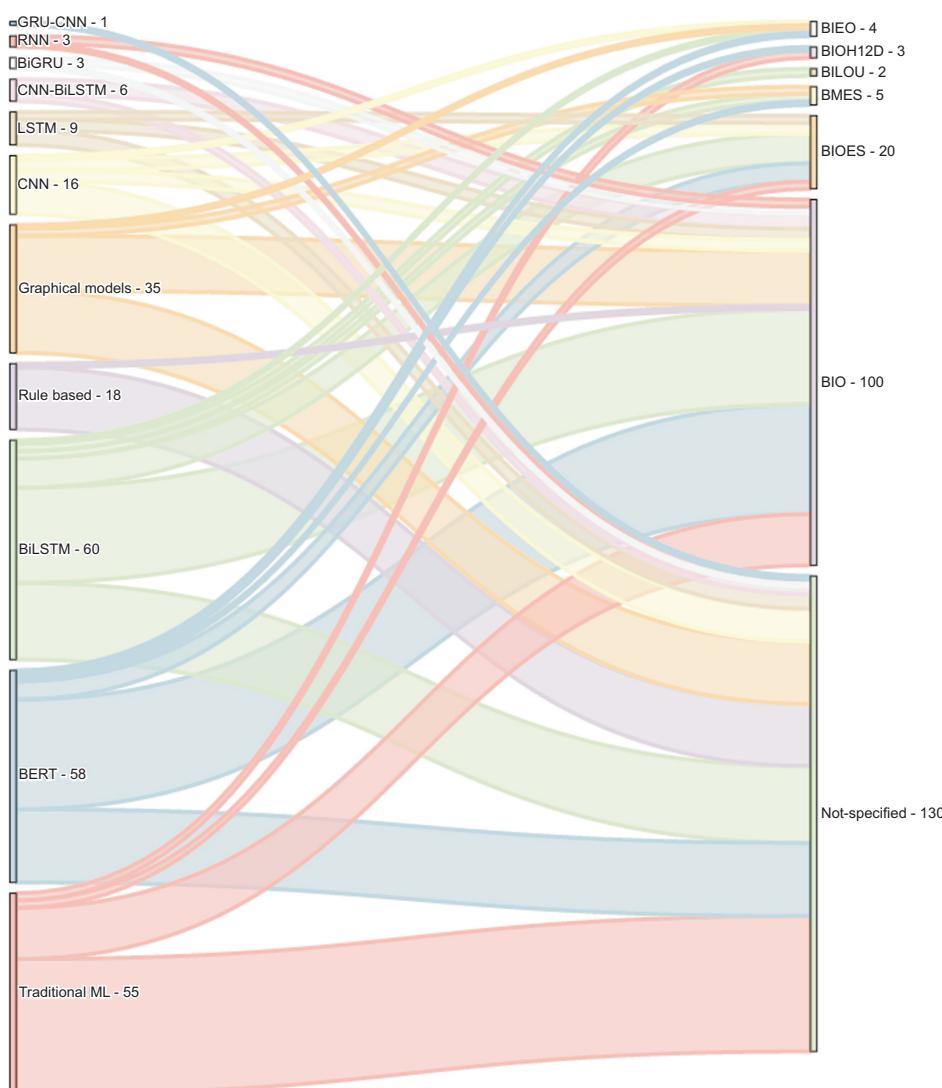


Figure 3. Named entity recognition approaches and types of tagging. GRU: gated recurrent unit, BiGRU: bidirectional gated recurrent unit, CNN: convolutional neural network, RNN: recurrent neural network, LSTM: long short-term memory, BiLSTM: bidirectional long short-term memory, ML: machine learning.

We observed that most of the articles did not specify the tagging system, a detail that is crucial for reproducing results, particularly with deep learning classifiers. This omission is more frequently seen in rule-based and traditional machine learning classifiers, as their goal is to classify each token on an individual basis. Conversely, graphical models and deep learning methods utilize context for token classification. Therefore, when an entity is defined by two or more tokens, it becomes necessary to specify the type of tagging system used.

Figure 4 presents a graph that categorizes the articles based on language, classification methods, and tagging systems. Deep learning models were predominantly used for the NER task, accounting for 58.86% of the methods, followed by traditional machine learning methods at 20.75%, graphical models at 13.20%, and rule-based approaches at 6.79%. In terms of languages, English was the most common,

representing 51.69% of the articles, followed by Chinese at 32.45%, Spanish and Swedish each at 4.15%, and Italian at 2.26%. We observed that 49.05% of the articles did not specify the tagging system, while 37.73% used BIO, and 7.54% used BIOES. It is noteworthy that deep learning approaches were adopted in 40 papers with Chinese corpora and 40 with English corpora.

3. Comparison between Shared-task Corpora and Private Corpora

In this section, we analyzed the corpora reported in the literature, which were either extracted from shared tasks or are publicly accessible (refer to Table 1). The most frequently encountered shared-task dataset in our review was the 2012 i2b2 challenge, which emphasized the extraction of concepts and relations from clinical texts [8,10,23,50,88,89]. This challenge targeted the following elements: (1) Clinically relevant

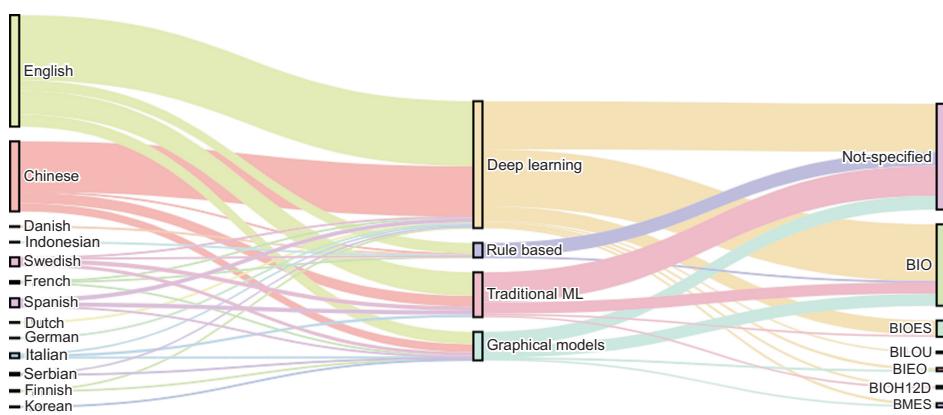


Figure 4. Corpus languages, types of models, and named entity recognition targets. ML: machine learning.

Table 1. Challenges in natural language processing

Challenge	Articles	Description
i2b2	31	A collaborative effort focused on automating the extraction of information from clinical narratives. Annually, I2b2 provides access to a collection of de-identified patient records that have been manually annotated by medical professionals.
CCKS	22	It comes from the China Conference on Knowledge Graph and Semantic Computing, which was founded in 2016 by the Chinese Information Processing Society. It provides task competition on NER and event extraction within Chinese EHRs.
n2c2	21	An outgrowth of the i2b2 challenge, and its datasets are under the stewardship of the Harvard Medical School Department of Biomedical Informatics. The challenge has multiple shared tasks, such as NER and RE within clinical notes. The datasets are hand-annotated by healthcare experts.
eHealth-KD	3	The eHealth Knowledge Discovery Challenge proposes tasks that involve the identification of semantic entities and relations between them in Spanish health documents.

NER: named entity recognition, EHR: Electronic Health Record, RE: relation extraction.

concepts, which include problems, tests, treatments, and clinical departments, as well as events such as admissions or transfers between departments that are pertinent to the patient's clinical timeline. (2) Temporal expressions that denote dates, times, durations, or frequencies within the clinical text. (3) Temporal relations between clinical events and temporal expressions. The best F1-score in this challenge was attained by a hybrid NLP system that merged a rule-based method with a machine learning approach, achieving an F1-score of 0.876 in the extraction of temporal expressions [90]. This underscores that the application of sophisticated NLP techniques can significantly enhance the identification and extraction of information from clinical data.

Another popular competition in the context of NLP tasks was the CCKS2017 challenge [6,35,91-94]. This challenge incorporated a dataset of 1,596 manually labeled medical records. The primary task involved the extraction of various entity types, including disease, anatomy, symptom, check, and treatment. The most successful results were obtained through the use of a straightforward CNN attention mecha-

nism, which achieved an F1-score of 90.34% [6].

The n2c2 challenge was designed to extract adverse drug events (ADEs) from a vast quantity of unstructured clinical records [50,54,78,84,85,94]. The annotations typically encompassed a variety of entity types, such as the drug, its strength, dosage, duration, frequency, form, route of administration, the reason for its prescription, and any ADEs. The data for the 2018 n2c2 challenge were derived from discharge summaries in the MIMIC-III database. One of the most notable results that year was reported in a study where a deep learning-based approach using BiLSTM-CRF was developed, resulting in an F1-score of 92% [84].

The eHealth-KD was an NLP challenge designed to model human language within Spanish EHRs. This challenge included NER and RE tasks within a general health domain [31,42]. The datasets utilized in this challenge identified the following entity types: concept, action, reference, and negation. Furthermore, the relation types included: part-of, property-of, same-as, subject, and target. A hybrid model that combined BiLSTM-CRF and CNN was applied to this

corpus, achieving an F1-score of 80.30% [31].

In regard to private corpora, this study found that only approximately 7% of the evaluated papers utilized their own unique private datasets. These datasets necessitate exclusive licenses or permissions from external sources for data access. The use of private corpora is subject to stringent confidentiality requirements; thus, they are not freely accessible [95]. Nevertheless, the employment of private corpora facilitates more thorough and contextually pertinent examinations

of medical texts. This contributes to the advancement of sophisticated healthcare applications and enhances patient outcomes. Table 2 presents a selection of studies that utilized private corpora and reported F1-scores exceeding 0.75. Most private corpora were customized to cater to specific domains, such as oncology or pathology. The types of notes included in these corpora encompassed pathology reports, admission notes, and medical notes from the intensive care unit.

Table 2. Summary of the papers utilizing a private corpus and attaining an F1-score greater than 0.75

NLP scheme	Description	Entities	Performance
BiLSTM-CRF [32]	Swedish general medical corpora and Spanish discharge reports from clinical units	In Spanish: diseases and drugs. In Swedish: body parts, disorders, and findings.	Micro-average F1-score of 0.75 for Spanish and 0.76 for Swedish
cTAKES with rule-and-dictionary-based approaches [95]	English progress notes from oncology centers	“Lung cancer,” “non-small cell lung cancer,” and “recurrence”	“Lung cancer” and “non-small cell lung cancer” achieved F1-scores of 0.828–0.947 and 0.86–0.93, respectively.
Bag of words and bag of bi-characters with a dictionary-based approach [43]	English progress notes from Michigan Pain Consultants	Relief, injections, drugs, surgery, and polarity	The best precision mean was found for polarity (97%).
Ruled-based and dictionary-matching approach [2]	English collection of clinical notes	Patients, encounters, findings, diagnoses, procedures, medications, and diagnostic tests	Recall of 94%, precision of 99%, and F-measure of 96%
ML approach (SVM with a linear kernel) and rigid rule-based approach [79]	English internal and external pathology notes	Specimen accession number within the report, received location, dates, and tissue block identifier	F1-score of 0.9014 for external reports, 0.9154 for internal reports, and 0.9708 for dates
Fine-tuned feature combined with CNN and BERT [78]	English collection of discharge letters from the intensive care unit	Drug names, routes of administration, frequencies, dosage, strength, form, and duration	Micro-average F1-score of 0.944
Convolutional neural networks + word embedding strategy using sub-word feature and “Bloom” embeddings [81]	English progress notes and medication list data from EHR	Drug names, routes, frequencies, dosage, strength, duration, adverse drug events, adherence, and current medications	The overall performance of the NER system was shown by an F1-score of 0.955. Higher performance was found for medication entities (drugs, names, routes, and frequencies).
BERT model with fine tuning in cancer domain [69]	Data obtained from 21,291 breast cancer patients (between 2001 to 2018)	Breast cancer phenotypes: hormone receptor type, hormone receptor status, tumor size, tumor site, cancer grade, histological type, tumor laterality, and cancer stage	The best model had macro-average F1-scores of 0.876 for an exact entity match.

NLP: natural language processing, BiLSTM: bidirectional long short-term memory, CRF: conditional random field, ML: machine learning, SVM: support vector machine, CNN: convolutional neural network, BERT: bidirectional encoder representations from transformers, EHR: Electronic Health Record, NER: named entity recognition.

Table 3. Summary of the papers reporting on RE

NLP scheme	Corpus	Relations	Performance
Ensemble deep learning methods (BiLSTM-CRF and transformers) [100]	n2c2/adverse drug events (ADEs) challenge	Strength-Drug, Dosage-Drug, Duration-Drug, Frequency-Drug, Form-Drug, Route-Drug, Reason-Drug, and ADE-Drug	The F1-score was 0.9458. The Reason-Drug relation had the highest improvement in the dataset.
BiLSTM encoder layer with segment attention layer and tensor-based approach in clinical texts [46]	2010 i2b2/VA challenge on problems, treatments, and test entities	Treatment was administered for a medical problem. Treatment worsened a medical problem. Treatment was not administered because of a medical problem. A test was conducted to investigate a medical problem. A medical problem indicates a medical problem.	The best performance was found for “test several problems,” with an F1-score of 0.833.
Dictionary-based approach, SVM, random forest, logistic regression, and CNN [85].	Chinese clinical records	Treatment improved or cured a medical problem. Treatment had no effect on a medical problem. Treatment worsened a medical problem. No relation was observed between a treatment and a medical problem.	Logistic regression showed the best performance, with an F1-score of 87.12%.

RE: relation extraction, BiLSTM: bidirectional long-short term memory, CRF: conditional random field, SVM: support vector machine, CNN: convolutional neural network.

Creating new corpora is a labor-intensive and resource-heavy task; however, it becomes essential when there is a need to develop more specialized applications [96]. The findings of this review indicate that the use of so-called public corpora is the dominant trend in publications related to NER.

4. Relation Extraction

RE is an active area of research in numerous specialized clinical fields. In this review, only three of the selected articles addressed the issue of RE, as shown in Table 3. Within the realm of medical records, we identified only two public corpora that included relation labeling: the 2010 i2b2 and the 2018 n2c2. Additionally, we discovered one article that utilized a private corpus for RE in the Chinese language [52]. It is evident that RE has not been as extensively explored as NER. Furthermore, the types of relations typically need to be defined based on the clinical domain or the entity type, adding a layer of complexity to this task that surpasses that of NER [96].

IV. Discussion

In this review, we found that the state-of-the-art has progressed from rule-based and traditional machine learning methods to deep-learning models over the past decade. This shift is due to the fact that the latter can comprehend the

context and enhance performance beyond what the former can achieve.

The review is limited by the current lack of a universally accepted standard method for assessing the quality of NER models. For example, while the F1-score is the most common metric, some papers do not clearly specify whether they are reporting the macro-average or the micro-average F1-score. Some papers even report different metrics altogether. Similarly, some papers do not clearly state the type of tagging system they employ. We recommend that researchers, when reporting NER results, provide explicit details about the corpus, the tagging system, and the performance metrics used in their study.

In terms of tagging systems, the BIO scheme is most frequently reported, particularly in studies employing deep-learning models. The introduction of more complex tagging schemes augments the number of classes that the model is required to predict, which could potentially impact its performance.

We have noted that recent progress in this field is largely dependent on publicly accessible corpora and datasets associated with challenges. Consequently, there is a conspicuous absence of research involving actual EHRs, and a substantial gap in thorough external validation, both with respect to fresh data and real-world applications. Therefore, the creation of new corpora is essential to facilitate the swift development of NER and RE models applicable to EHRs for use

in clinical practice, and to validate the outcomes in various datasets. Furthermore, initiatives should be undertaken to convert private corpora into public ones.

Most of the existing corpora utilize EHRs written in English. Additionally, we have observed a swift expansion of corpora in the Chinese language, primarily employing deep-learning models. The creation of new models in various languages presents a challenge for the global implementation of NER in clinical practice. Likewise, when dealing with languages other than English, only a few corpora are freely accessible, which underscores the importance of developing custom datasets. This is crucial to guarantee the relevance and effectiveness of NLP models in a variety of linguistic contexts.

In the realm of clinical domains, there is a scarcity of studies linked to specific medical specialties. A mere 8.13% of the studies concentrated on neoplasms, while 6.5% focused on cardiovascular diseases, 1.62% on factors influencing health status, and a scant 0.813% on mental and behavioral disorders. Most studies targeting specific domains utilized private corpora, underscoring the role of these resources in supplementing the use of public datasets. This heavy reliance on private corpora emphasizes the necessity for researchers to forge partnerships or collaborations with healthcare institutions or data providers. This allows secure access to these invaluable resources, while maintaining adherence to ethical and legal considerations to protect sensitive patient information.

As of December 2022, the most advanced models in clinical NER are those based on BERT, which undergo a fine-tuning or training stage using a domain-specific corpus. For example, models such as BioBERT, BioClinicalBERT, and BlueBERT have demonstrated superior performance in this field. In 2023, ChatGPT gained recognition for leading a revolution in the field of NLP, with notable performance on generic text corpora. However, a study conducted by Hu et al. revealed that the performance of ChatGPT, for the NER task defined in the 2010 i2b2 challenge, was inferior to that of the BioClinicalBERT model [97]. The latter model underwent a fine-tuning stage using a specific corpus. This finding aligns with the study conducted by Li et al. [98], which discovered that ChatGPT and GPT-4 encountered difficulties in areas requiring domain-specific knowledge. Specifically, they utilized financial textual datasets. Furthermore, Lai et al. [99] proposed that, for the time being, it is more practical to employ task-specific models for domain-specific tasks rather than using ChatGPT. Nevertheless, future research should be conducted to investigate the potential use of new develop-

ments in LLM for classifying named entities in EHRs.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was funded by the Instituto Tecnológico Metropolitano through the project (No. P20242). Also, the project received funds from the Agencia de Educación Superior de Medellín (Sapiencia).

ORCID

María C. Durango (<https://orcid.org/0000-0002-2779-2461>)

Ever A. Torres-Silva (<https://orcid.org/0000-0002-6302-6131>)

Andrés Orozco-Duque (<https://orcid.org/0000-0001-8582-8015>)

References

1. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;309(13):1351-2. <https://doi.org/10.1001/jama.2013.393>
2. Yehia E, Boshnak H, AbdelGaber S, Abdo A, Elzafaly DS. Ontology-based clinical information extraction from physician's free-text notes. *J Biomed Inform* 2019;98:103276. <https://doi.org/10.1016/j.jbi.2019.103276>
3. ElDin HG, AbdulRazek M, Abdelshafi M, Sahlol AT. Med-Flair: medical named entity recognition for diseases and medications based on Flair embedding. *Procedia Comput Sci* 2021;189:67-75. <https://doi.org/10.1016/j.procs.2021.05.078>
4. Kaplar A, Stosovic M, Kaplar A, Brkovic V, Naumovic R, Kovacevic A. Evaluation of clinical named entity recognition methods for Serbian electronic health records. *Int J Med Inform* 2022;164:104805. <https://doi.org/10.1016/j.ijmedinf.2022.104805>
5. Neuraz A, Looten V, Rance B, Daniel N, Garcelon N, Llanos LC, et al. Do you need embeddings trained on a massive specialized corpus for your clinical natural language processing task? *Stud Health Technol Inform* 2019;264:1558-9. <https://doi.org/10.3233/SHTI190533>
6. Kong J, Zhang L, Jiang M, Liu T. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition. *J Biomed Inform* 2021;116:103737. <https://doi.org/10.1016/j.jbi.2021.103737>

7. Xiong Y, Wang Z, Jiang D, Wang X, Chen Q, Xu H, et al. A fine-grained Chinese word segmentation and part-of-speech tagging corpus for clinical text. *BMC Med Inform Decis Mak* 2019;19(Suppl 2):66. <https://doi.org/10.1186/s12911-019-0770-7>
8. Jindal P, Roth D. Extraction of events and temporal expressions from clinical narratives. *J Biomed Inform* 2013;46 Suppl:S13-9. <https://doi.org/10.1016/j.jbi.2013.08.010>
9. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *J Am Med Inform Assoc* 2011;18(5):601-6. <https://doi.org/10.1136/amiajnl-2011-000163>
10. Nikfarjam A, Emadzadeh E, Gonzalez G. Towards generating a patient's timeline: extracting temporal relationships from clinical notes. *J Biomed Inform* 2013;46 Suppl(0):S40-7. <https://doi.org/10.1016/j.jbi.2013.11.001>
11. Casillas A, Perez A, Oronoz M, Gojenola K, Santiso S. Learning to extract adverse drug reaction events from electronic health records in Spanish. *Expert Syst Appl* 2016;61(1):235-45. <https://doi.org/10.1016/j.eswa.2016.05.034>
12. Henriksson A, Kvist M, Dalianis H, Duneld M. Identifying adverse drug event information in clinical notes with distributional semantic representations of context. *J Biomed Inform* 2015;57:333-49. <https://doi.org/10.1016/j.jbi.2015.08.013>
13. de Bruijn B, Cherry C, Kiritchenko S, Martin J, Zhu X. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. *J Am Med Inform Assoc* 2011;18(5):557-62. <https://doi.org/10.1136/amiajnl-2011-000150>
14. Cormack J, Nath C, Milward D, Raja K, Jonnalagadda SR. Agile text mining for the 2014 i2b2/UTHealth Cardiac risk factors challenge. *J Biomed Inform* 2015;58 Suppl(0):S120-7. <https://doi.org/10.1016/j.jbi.2015.06.030>
15. Grouin C, Grabar N, Hamon T, Rosset S, Tannier X, Zweigenbaum P. Eventual situations for timeline extraction from clinical reports. *J Am Med Inform Assoc* 2013;20(5):820-7. <https://doi.org/10.1136/amiajnl-2013-001627>
16. Lei J, Tang B, Lu X, Gao K, Jiang M, Xu H. A comprehensive study of named entity recognition in Chinese clinical text. *J Am Med Inform Assoc* 2014;21(5):808-14. <https://doi.org/10.1136/amiajnl-2013-002381>
17. Dligach D, Bethard S, Becker L, Miller T, Savova GK. Discovering body site and severity modifiers in clinical texts. *J Am Med Inform Assoc* 2014;21(3):448-54. <https://doi.org/10.1136/amiajnl-2013-001766>
18. Jung K, LePendu P, Iyer S, Bauer-Mehren A, Percha B, Shah NH. Functional evaluation of out-of-the-box text-mining tools for data-mining tasks. *J Am Med Inform Assoc* 2015;22(1):121-31. <https://doi.org/10.1136/amiajnl-2014-002902>
19. Yang H, Garibaldi JM. Automatic detection of protected health information from clinic narratives. *J Biomed Inform* 2015;58(Suppl):S30-8. <https://doi.org/10.1016/j.jbi.2015.06.015>
20. Lee W, Kim K, Lee EY, Choi J. Conditional random fields for clinical named entity recognition: a comparative study using Korean clinical texts. *Comput Biol Med* 2018;101:7-14. <https://doi.org/10.1016/j.compbio.2018.07.019>
21. Wang H, Zhang W, Zeng Q, Li Z, Feng K, Liu L. Extracting important information from Chinese Operation Notes with natural language processing methods. *J Biomed Inform* 2014;48:130-6. <https://doi.org/10.1016/j.jbi.2013.12.017>
22. Grouin C, Neveol A. De-identification of clinical notes in French: towards a protocol for reference corpus development. *J Biomed Inform* 2014;50:151-61. <https://doi.org/10.1016/j.jbi.2013.12.014>
23. Kovacevic A, Dehghan A, Filannino M, Keane JA, Nenadic G. Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. *J Am Med Inform Assoc* 2013;20(5):859-66. <https://doi.org/10.1136/amiajnl-2013-001625>
24. Xiong Y, Peng H, Xiang Y, Wong KC, Chen Q, Yan J, et al. Leveraging Multi-source knowledge for Chinese clinical named entity recognition via relational graph convolutional network. *J Biomed Inform* 2022;128:104035. <https://doi.org/10.1016/j.jbi.2022.104035>
25. Jiale N, Gao D, Sun Y, Li X, Shen X, Li M, et al. Surgical procedure long terms recognition from Chinese literature incorporating structural feature. *Heliyon* 2022;8(11):e11291. <https://doi.org/10.1016/j.heliyon.2022.e11291>
26. Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. *Artif Intell Med* 2016;66:29-39. <https://doi.org/10.1016/j.artmed.2015.09.007>
27. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging [Internet]. Ithaca (NY): arXiv.org; 2015 [cited at 2023 Sep 30]. Available from: <https://arxiv.org/abs/1508.01991>.
28. Zhang X, Zhang Y, Zhang Q, Ren Y, Qiu T, Ma J, et al. Extracting comprehensive clinical information for breast cancer using deep learning methods. *Int J Med Inform* 2019;132:103985. <https://doi.org/10.1016/j.ijmedinf.2019.103985>

- 2019.103985
29. Xu J, Li Z, Wei Q, Wu Y, Xiang Y, Lee HJ, et al. Applying a deep learning-based sequence labeling approach to detect attributes of medical concepts in clinical text. *BMC Med Inform Decis Mak* 2019;19(Suppl 5):236. <https://doi.org/10.1186/s12911-019-0937-2>
 30. Li PL, Yuan ZM, Tu WN, Yu K, Lu DX. Medical knowledge extraction and analysis from electronic medical records using deep learning. *Chin Med Sci J* 2019;34(2):133-9. <https://doi.org/10.24920/003589>
 31. Suarez-Paniagua V, Rivera Zavala RM, Segura-Bedmar I, Martinez P. A two-stage deep learning approach for extracting entities and relationships from medical texts. *J Biomed Inform* 2019;99:103285. <https://doi.org/10.1016/j.jbi.2019.103285>
 32. Weegar R, Perez A, Casillas A, Oronoz M. Recent advances in Swedish and Spanish medical entity recognition in clinical texts using deep neural approaches. *BMC Med Inform Decis Mak* 2019;19(Suppl 7):274. <https://doi.org/10.1186/s12911-019-0981-y>
 33. Ji B, Liu R, Li S, Yu J, Wu Q, Tan Y, et al. A hybrid approach for named entity recognition in Chinese electronic medical record. *BMC Med Inform Decis Mak* 2019;19(Suppl 2):64. <https://doi.org/10.1186/s12911-019-0767-2>
 34. Wang Q, Zhou Y, Ruan T, Gao D, Xia Y, He P. Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition. *J Biomed Inform* 2019;92:103133. <https://doi.org/10.1016/j.jbi.2019.103133>
 35. Yin M, Mou C, Xiong K, Ren J. Chinese clinical named entity recognition with radical-level feature and self-attention mechanism. *J Biomed Inform* 2019;98:103289. <https://doi.org/10.1016/j.jbi.2019.103289>
 36. Alex B, Grover C, Tobin R, Sudlow C, Mair G, Whiteley W. Text mining brain imaging reports. *J Biomed Semantics* 2019;10(Suppl 1):23. <http://dx.doi.org/10.1186/s13326-019-0211-7>
 37. Shi X, Yi Y, Xiong Y, Tang B, Chen Q, Wang X, et al. Extracting entities with attributes in clinical text via joint deep learning. *J Am Med Inform Assoc* 2019;26(12):1584-91. <https://doi.org/10.1093/jamia/ocz158>
 38. Wang Y, Ananiadou S, Tsujii J. Improving clinical named entity recognition in Chinese using the graphical and phonetic feature. *BMC Med Inform Decis Mak* 2019;19(Suppl 7):273. <https://doi.org/10.1186/s12911-019-0980-z>
 39. Casillas A, Ezeiza N, Goenaga I, Perez A, Soto X. Measuring the effect of different types of unsupervised word representations on Medical Named Entity Recognition. *Int J Med Inform* 2019;129:100-6. <https://doi.org/10.1016/j.ijmedinf.2019.05.022>
 40. Chen Y, Zhou C, Li T, Wu H, Zhao X, Ye K, et al. Named entity recognition from Chinese adverse drug event reports with lexical feature based BiLSTM-CRF and tri-training. *J Biomed Inform* 2019;96:103252. <https://doi.org/10.1016/j.jbi.2019.103252>
 41. Liu X, Zhou Y, Wang Z. Recognition and extraction of named entities in online medical diagnosis data based on a deep neural network. *J Vis Commun Image Represent* 2019;60:1-15. <https://doi.org/10.1016/j.jvcir.2019.02.001>
 42. Li Z, Yang J, Gou X, Qi X. Recurrent neural networks with segment attention and entity description for relation extraction from clinical texts. *Artif Intell Med* 2019;97:9-18. <https://doi.org/10.1016/j.artmed.2019.04.003>
 43. Juckett DA, Kasten EP, Davis FN, Gostine M. Concept detection using text exemplars aligned with a specialized ontology. *Data Knowl Eng* 2019;119:22-35. <https://doi.org/10.1016/j.datak.2018.11.002>
 44. Su J, Hu J, Jiang J, Xie J, Yang Y, He B, et al. Extraction of risk factors for cardiovascular diseases from Chinese electronic medical records. *Comput Methods Programs Biomed* 2019;172:1-10. <https://doi.org/10.1016/j.cmpb.2019.01.007>
 45. Li L, Zhao J, Hou L, Zhai Y, Shi J, Cui F. An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records. *BMC Med Inform Decis Mak* 2019;19(Suppl 5):235. <https://doi.org/10.1186/s12911-019-0933-6>
 46. Dong X, Chowdhury S, Qian L, Li X, Guan Y, Yang J, et al. Deep learning for named entity recognition on Chinese electronic medical records: Combining deep transfer learning with multitask bi-directional LSTM RNN. *PLoS One* 2019;14(5):e0216046. <https://doi.org/10.1371/journal.pone.0216046>
 47. Lin CH, Hsu KC, Liang CK, Lee TH, Liou CW, Lee JD, et al. A disease-specific language representation model for cerebrovascular disease research. *Comput Methods Programs Biomed* 2021;211:106446. <https://doi.org/10.1016/j.cmpb.2021.106446>
 48. Murugadoss K, Rajasekharan A, Malin B, Agarwal V, Bade S, Anderson JR, et al. Building a best-in-class automated de-identification tool for electronic health records through ensemble learning. *Patterns (N Y)* 2021;2(6):100255.

- <https://doi.org/10.1016/j.patter.2021.100255>
49. Harnoune A, Rhanoui M, Mikram M, Yousfi S, Elkaimbillah Z, El Asri B. BERT based clinical knowledge extraction for biomedical knowledge graph construction and analysis. *Comput Methods Programs Biomed Update* 2021;1:100042. <https://doi.org/10.1016/j.cmpbup.2021.100042>
50. Narayanan S, Achan P, Rangan PV, Rajan SP. Unified concept and assertion detection using contextual multi-task learning in a clinical decision support system. *J Biomed Inform* 2021;122:103898. <https://doi.org/10.1016/j.jbi.2021.103898>
51. Thieu T, Maldonado JC, Ho PS, Ding M, Marr A, Brandt D, et al. A comprehensive study of mobility functioning information in clinical notes: entity hierarchy, corpus annotation, and sequence labeling. *Int J Med Inform* 2021;147:104351. <https://doi.org/10.1016/j.ijmedinf.2020.104351>
52. Liu X, Liu Y, Wu H, Guan Q. A tag based joint extraction model for Chinese medical text. *Comput Biol Chem* 2021;93:107508. <https://doi.org/10.1016/j.compbiolchem.2021.107508>
53. Puccetti G, Chiarello F, Fantoni G. A simple and fast method for Named Entity context extraction from patients. *Expert Syst Appl* 2021;184:115570. <https://doi.org/10.1016/j.eswa.2021.115570>
54. El-allaly ED, Sarrouti M, En-Nahnabi N, El Alaoui SO. MTTLADE: a multi-task transfer learning-based method for adverse drug events extraction. *Inf Process Manag* 2021;58(3):102473. <https://doi.org/10.1016/j.ipm.2020.102473>
55. Uronen L, Salanterä S, Hakala K, Hartiala J, Moen H. Combining supervised and unsupervised named entity recognition to detect psychosocial risk factors in occupational health checks. *Int J Med Inform* 2022;160:104695. <https://doi.org/10.1016/j.ijmedinf.2022.104695>
56. Xiong Y, Peng W, Chen Q, Huang Z, Tang B. A unified machine reading comprehension framework for cohort selection. *IEEE J Biomed Health Inform* 2022;26(1):379-87. <https://doi.org/10.1109/jbhi.2021.3095478>
57. Zhang B, Liu K, Wang H, Li M, Pan J. Chinese named-entity recognition via self-attention mechanism and position-aware influence propagation embedding. *Data Knowl Eng* 2022;139:101983. <https://doi.org/10.1016/j.datak.2022.101983>
58. Landolsi MY, Romdhane L Ben, Hlaoua L. Medical named entity recognition using surrounding sequences matching. *Procedia Comput Sci* 2022;207:674-83. <https://doi.org/10.1016/j.procs.2022.09.122>
59. Shi J, Sun M, Sun Z, Li M, Gu Y, Zhang W. Multi-level semantic fusion network for Chinese medical named entity recognition. *J Biomed Inform* 2022;133:104144. <https://doi.org/10.1016/j.jbi.2022.104144>
60. Gerardin C, Wajsburd P, Vaillant P, Bellamine A, Carrat F, Tannier X. Multilabel classification of medical concepts for patient clinical profile identification. *Artif Intell Med* 2022;128:102311. <https://doi.org/10.1016/j.artmed.2022.102311>
61. Madan S, Julius Zimmer F, Balabin H, Schaaf S, Frohlich H, Fluck J, et al. Deep learning-based detection of psychiatric attributes from German mental health records. *Int J Med Inform* 2022;161:104724. <https://doi.org/10.1016/j.ijmedinf.2022.104724>
62. Narayanan S, Madhuri SS, Ramesh MV, Rangan PV, Rajan SP. Deep contextual multi-task feature fusion for enhanced concept, negation and speculation detection from clinical notes. *Informatics Med Unlocked* 2022;34:101109. <https://doi.org/10.1016/j.imu.2022.101109>
63. An Y, Xia X, Chen X, Wu FX, Wang J. Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF. *Artif Intell Med* 2022;127:102282. <https://doi.org/10.1016/j.artmed.2022.102282>
64. Wang SY, Huang J, Hwang H, Hu W, Tao S, Hernandez-Boussard T. Leveraging weak supervision to perform named entity recognition in electronic health records progress notes to identify the ophthalmology exam. *Int J Med Inform* 2022;167:104864. <https://doi.org/10.1016/j.ijmedinf.2022.104864>
65. Narayanan S, Mannam K, Achan P, Ramesh MV, Rangan PV, Rajan SP. A contextual multi-task neural approach to medication and adverse events identification from clinical text. *J Biomed Inform* 2022;125:103960. <https://doi.org/10.1016/j.jbi.2021.103960>
66. El-Allaly ED, Sarrouti M, En-Nahnabi N, Ouatik El Alaoui S. An attentive joint model with transformer-based weighted graph convolutional network for extracting adverse drug event relation. *J Biomed Inform* 2022;125:103968. <https://doi.org/10.1016/j.jbi.2021.103968>
67. Sun J, Liu Y, Cui J, He H. Deep learning-based methods for natural hazard named entity recognition. *Sci Rep* 2022;12(1):4598. <https://doi.org/10.1038/s41598-022-08667-2>
68. Fang A, Hu J, Zhao W, Feng M, Fu J, Feng S, et al. Extracting clinical named entity for pituitary adenomas from Chinese electronic medical records. *BMC Med Inform Decis Mak* 2022;22(1):72. <https://doi.org/10.1186/s12911-022-01810-z>

69. Zhou S, Wang N, Wang L, Liu H, Zhang R. Cancer-BERT: a cancer domain-specific language model for extracting breast cancer phenotypes from electronic health records. *J Am Med Inform Assoc* 2022;29(7):1208-16. <https://doi.org/10.1093/jamia/ocac040>
70. Guo S, Yang W, Han L, Song X, Wang G. A multi-layer soft lattice based model for Chinese clinical named entity recognition. *BMC Med Inform Decis Mak* 2022;22(1):201. <https://doi.org/10.1186/s12911-022-01924-4>
71. Lee EB, Heo GE, Choi CM, Song M. MLM-based typographical error correction of unstructured medical texts for named entity recognition. *BMC Bioinformatics* 2022;23(1):486. <https://doi.org/10.1186/s12859-022-05035-9>
72. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30:5999-6009.
73. Xu Y, Wang Y, Liu T, Liu J, Fan Y, Qian Y, et al. Joint segmentation and named entity recognition using dual decomposition in Chinese discharge summaries. *J Am Med Inform Assoc* 2014;21(e1):e84-92. <https://doi.org/10.1136/amiajnl-2013-001806>
74. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;36(4):1234-40. <https://doi.org/10.48550/arXiv.1901.08746>
75. Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets [Internet]. Ithaca (NY): arXiv.org; 2019 [cited at 2023 Sep 30]. Available from: <https://arxiv.org/abs/1906.05474>.
76. Ji B, Li S, Yu J, Ma J, Tang J, Wu Q, et al. Research on Chinese medical named entity recognition based on collaborative cooperation of multiple neural network models. *J Biomed Inform* 2020;104:103395.
77. Sterckx L, Vandewiele G, Dehaene I, Janssens O, Ongevrae F, De Backere F, et al. Clinical information extraction for preterm birth risk prediction. *J Biomed Inform* 2020;110:103544. <https://doi.org/10.1016/j.jbi.2020.103395>
78. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: a transferable clinical natural language processing model for electronic health records. *Artif Intell Med* 2021;118:102086. <https://doi.org/10.1016/j.artmed.2021.102086>
79. Wei Q, Ji Z, Li Z, Du J, Wang J, Xu J, et al. A study of deep learning approaches for medication and adverse drug event extraction from clinical text. *J Am Med Inform Assoc* 2020;27(1):13-21. <https://doi.org/10.1093/jamia/ocz063>
80. Khan S, Shamsi JA. Health Quest: a generalized clinical decision support system with multi-label classification. *J King Saud Univ-Comput Inf Sci* 2021;33(1):45-53. <https://doi.org/10.1016/j.jksuci.2018.11.003>
81. Lin WC, Chen JS, Kaluzny J, Chen A, Chiang MF, Hribar MR. Extraction of active medications and adherence using natural language processing for glaucoma patients. *AMIA Annu Symp Proc* 2021;2021:773-82.
82. Dai HJ. Family member information extraction via neural sequence labeling models with different tag schemes. *BMC Med Inform Decis Mak* 2019;19(Suppl 10):257.
83. Chen L, Li Y, Chen W, Liu X, Yu Z, Zhang S. Utilizing soft constraints to enhance medical relation extraction from the history of present illness in electronic medical records. *J Biomed Inform* 2018;87:108-17.
84. Dai HJ, Su CH, Wu CS. Adverse drug event and medication extraction in electronic health records via a cascading architecture with different sequence labeling models and word embeddings. *J Am Med Inform Assoc* 2020;27(1):47-55. <https://doi.org/10.1093/jamia/ocz120>
85. Yang X, Bian J, Fang R, Bjarnadottir RI, Hogan WR, Wu Y. Identifying relations of medications with adverse drug events using recurrent convolutional neural networks and gradient boosting. *J Am Med Inform Assoc* 2020;27(1):65-72. <https://doi.org/10.1093/jamia/ocz144>
86. Li L, Xu W, Yu H. Character-level neural network model based on Nadam optimization and its application in clinical concept extraction. *Neurocomputing* 2020;414:182-90. <https://doi.org/10.1016/j.neucom.2020.07.027>
87. Zhang R, Zhao P, Guo W, Wang R, Lu W. Medical named entity recognition based on dilated convolutional neural network. *Cogn Robot* 2022;2:13-20. <https://doi.org/10.1016/j.cogr.2021.11.002>
88. Cheng Y, Anick P, Hong P, Xue N. Temporal relation discovery between events and temporal expressions identified in clinical narrative. *J Biomed Inform* 2013;46 Suppl:S48-53. <https://doi.org/10.1016/j.jbi.2013.09.010>
89. Liu Z, Yang M, Wang X, Chen Q, Tang B, Wang Z, et al. Entity recognition from clinical texts via recurrent neural network. *BMC Med Inform Decis Mak* 2017;17(Suppl 2):67. <https://doi.org/10.1186%2Fs12911-017-0468-7>
90. Kochmar E, Andersen O, Briscoe T. HOO 2012 error recognition and correction shared task: Cambridge University submission report. Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications; 2012 Jun 7; Montreal, Canada. p.

- 242-50. <https://doi.org/10.17863/CAM.9671>
91. Zhao S, Cai Z, Chen H, Wang Y, Liu F, Liu A. Adversarial training based lattice LSTM for Chinese clinical named entity recognition. *J Biomed Inform* 2019;99:103290. <https://doi.org/10.1016/j.jbi.2019.103290>
92. Yu G, Yang Y, Wang X, Zhen H, He G, Li Z, et al. Adversarial active learning for the identification of medical concepts and annotation inconsistency. *J Biomed Inform* 2020;108:103481. <https://doi.org/10.1016/j.jbi.2020.103481>
93. Wang C, Wang H, Zhuang H, Li W, Han S, Zhang H, et al. Chinese medical named entity recognition based on multi-granularity semantic dictionary and multimodal tree. *J Biomed Inform* 2020;111:103583. <https://doi.org/10.1016/j.jbi.2020.103583>
94. Chen X, Ouyang C, Liu Y, Bu Y. Improving the named entity recognition of Chinese electronic medical records by combining domain dictionary and rules. *Int J Environ Res Public Health* 2020;17(8):2687. <https://doi.org/10.3390/ijerph17082687>
95. Kersloot MG, Lau F, Abu-Hanna A, Arts DL, Cornet R. Automated SNOMED CT concept and attribute relationship detection through a web-based implementation of cTAKES. *J Biomed Semantics* 2019;10(1):14. <https://doi.org/10.1186/s13326-019-0207-3>
96. Christopoulou F, Tran TT, Sahu SK, Miwa M, Ananiadou S. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods. *J Am Med Inform Assoc* 2020;27(1):39-46.
97. Hu Y, Ameer I, Zuo W, Peng X, Zhou Y, Li Z, et al. Zero-shot clinical entity recognition using ChatGPT [Internet]. Ithaca (NY): arXiv.org; 2023 [cited at 2023 Sep 30]. Available from: <https://arxiv.org/abs/2303.16416>.
98. Li X, Zhu X, Ma Z, Liu X, Shas S. Are ChatGPT and GPT-4 general-purpose solvers for financial text analytics? An examination on several typical tasks [Internet]. Ithaca (NY): arXiv.org; 2023 [cited at 2023 Sep 30]. Available from: <https://arxiv.org/abs/2305.05862>.
99. Lai VD, Ngo NT, Veyseh AP, Man H, Dernoncourt F, Bui T, et al. ChatGPT beyond English: towards a comprehensive evaluation of large language models in multilingual learning [Internet]. Ithaca (NY): arXiv.org; 2023 [cited at 2023 Sep 30]. Available from: <https://arxiv.org/abs/2304.05613>.

Appendix A. Search equations

Database	Search term
PubMed	(“Data Mining”[Mesh] OR “Natural Language Processing”[Mesh]) AND (“Electronic Health Records”[Mesh]) AND (“text mining”[tiab] OR “Text data mining”[tiab] OR “Text analytics”[tiab] OR “Text analysis” OR “Text clustering”) AND (“Named Entity Recognition”[tiab] OR “NER”[tiab] OR “Named Entity Recognition and Classification”[tiab] OR “NERC”[tiab] OR “Named Entity Clustering”[tiab] OR “Clinical named entity recognition”[tiab])
VHL	(mh:(“Minería de Datos”) OR mh:(“Procesamiento de Lenguaje Natural”)) AND (mh: (“Registros Electrónicos de Salud”)) AND ((text mining) OR (minería de texto) OR (text data mining) OR (minería de datos de texto) OR (text analytics) OR (analítica de texto) OR (text clustering) OR (agrupación de texto)) AND ((named entity recognition) OR (Clinical named entity recognition) OR (reconocimiento de entidades nombradas) OR (NER) OR (named entity recognition and classification) OR (reconocimiento y clasificación de entidades nombradas) OR (NERC) OR (named entity clustering) OR (agrupación de entidades con nombre) OR (reconocimiento de entidades nombradas clínicas))
IEEE	(“Mesh Terms”:Data mining OR “Mesh Terms”:Natural Language Processing) AND (“Mesh Terms”:Electronic Health Records) AND (“Full Text & Metadata”:Text mining OR “Full Text & Metadata”:Text data mining OR “Full Text & Metadata”:Text analy* OR “Full Text & Metadata”:Text clustering) AND (“Full Text & Metadata”:named entity recognition OR “Full Text & Metadata”:NER OR “Full Text & Metadata”:Named entity recognition and classification OR “Full Text & Metadata”:NERC OR “Full Text & Metadata”:Named entity Clustering OR “Full Text & Metadata”:Clinical named entity recognition)
Science Direct	(“Text mining” OR “Natural Language Processing”) AND (“Electronic Health Records” OR “Electronic Medical Records”) AND (“Named entity recognition” OR “Named Entity recognition and classification” OR “Named Entity Clustering” OR “Clinical named entity recognition”)