

Review

Health Care Language Models and Their Fine-Tuning for Information Extraction: Scoping Review

Miguel Nunes¹, BSc; Joao Bone², MSc; Joao C Ferreira^{1,3,4}, PhD; Luis B Elvas^{1,3,4}, MSc

¹ISTAR, Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal

²Select Data, Anaheim, CA, United States

³Department of Logistics, Molde, University College, Molde, Norway

⁴INOV Inesc Inovação, Instituto de Novas Tecnologias, Lisbon, Portugal

Corresponding Author:

Luis B Elvas, MSc

Department of Logistics, Molde, University College

Britvegen 2, Noruega

Molde, 6410

Norway

Phone: 47 969152334

Email: luis.m.elvas@himolde.no

Abstract

Background: In response to the intricate language, specialized terminology outside everyday life, and the frequent presence of abbreviations and acronyms inherent in health care text data, domain adaptation techniques have emerged as crucial to transformer-based models. This refinement in the knowledge of the language models (LMs) allows for a better understanding of the medical textual data, which results in an improvement in medical downstream tasks, such as information extraction (IE). We have identified a gap in the literature regarding health care LMs. Therefore, this study presents a scoping literature review investigating domain adaptation methods for transformers in health care, differentiating between English and non-English languages, focusing on Portuguese. Most specifically, we investigated the development of health care LMs, with the aim of comparing Portuguese with other more developed languages to guide the path of a non-English-language with fewer resources.

Objective: This study aimed to research health care IE models, regardless of language, to understand the efficacy of transformers and what are the medical entities most commonly extracted.

Methods: This scoping review was conducted using the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) methodology on Scopus and Web of Science Core Collection databases. Only studies that mentioned the creation of health care LMs or health care IE models were included, while large language models (LLMs) were excluded. The latest were not included since we wanted to research LMs and not LLMs, which are architecturally different and have distinct purposes.

Results: Our search query retrieved 137 studies, 60 of which met the inclusion criteria, and none of them were systematic literature reviews. English and Chinese are the languages with the most health care LMs developed. These languages already have disease-specific LMs, while others only have general-health care LMs. European Portuguese does not have any public health care LM and should take examples from other languages to develop, first, general-health care LMs and then, in an advanced phase, disease-specific LMs. Regarding IE models, transformers were the most commonly used method, and named entity recognition was the most popular topic, with only a few studies mentioning Assertion Status or addressing medical lexical problems. The most extracted entities were diagnosis, posology, and symptoms.

Conclusions: The findings indicate that domain adaptation is beneficial, achieving better results in downstream tasks. Our analysis allowed us to understand that the use of transformers is more developed for the English and Chinese languages. European Portuguese lacks relevant studies and should draw examples from other non-English languages to develop these models and drive progress in AI. Health care professionals could benefit from highlighting medically relevant information and optimizing the reading of the textual data, or this information could be used to create patient medical timelines, allowing for profiling.

(*JMIR Med Inform* 2024;12:e60164) doi: [10.2196/60164](https://doi.org/10.2196/60164)

KEYWORDS

language model; information extraction; healthcare; PRISMA-ScR; scoping literature review; transformers; natural language processing; European Portuguese

Introduction

The health care sector generates a vast amount of structured and unstructured data, including images from medical exams, text written in electronic medical records (EMRs) or Electronic Health Records (EHRs), and structured data from relational databases that store patient and admission information, as well as all the data collected during a patient's hospitalization [1]. Approximately 30% of the world's data volume is generated by the health care sector, and projections indicate that by 2025, the compound annual growth rate of data for health care will reach 36% [2].

Medical texts present several challenges due to the use of unfamiliar context-specific terminologies that differ from everyday language. In addition, physicians often use abbreviations and acronyms to save time and space. However, the same abbreviation can have different meanings, adding an additional layer of complexity when trying to understand the content of medical texts [3]. All these characteristics pose challenges when attempting to apply artificial intelligence (AI) techniques to interpret the text.

In the field of natural language processing (NLP), the introduction of transformers [4] has revolutionized the field, achieving state-of-the-art performance for numerous NLP tasks [5]. Their general architecture comprises an encoder, which receives the input and builds a representation of it, and a decoder that uses the encoder's representation along with other inputs to generate a target sequence. The introduction of the self-attention mechanism further revolutionized NLP by allowing the model to weigh the importance of different words in a sentence regardless of their position. This enables better handling of long-range dependencies compared with traditional deep learning (DL) architectures like recurrent neural networks (RNNs) and long short-term memory Networks [6]. In the context of medical text, transformers excel in interpreting and extracting medically relevant information by effectively handling context and meaning, even in complex and specialized language.

Transformers can be trained as language models (LMs) on raw text in a self-supervised manner, enabling them to develop a statistical understanding of the text they were trained on [7]. However, the benefits of this approach are only fully realized when fine-tuning a downstream task.

Another important concept is called domain adaptation, which stands for the process of adapting or adjusting something to be suitable within a different domain or context. In the field of machine learning (ML), domain adaptation is used to align the disparity between domains so that the trained model can generalize into the domain of interest [8]. For transformers, domain adaptation involves continuing the pretraining of an LM with text data from a different domain than the one it was originally trained on [9]. This approach allows for leveraging the learning capabilities of general-scope LMs and refining

them for specific contexts. For example, if we consider a general-scope LM, one that was trained using textual data from various domains, and continue its pretraining with health care-specific textual data, it will help the LM to refine its understanding of the health care data, leading to improvements when fine-tuning the LM for downstream tasks related to health care. To explore this further, we can take a health care LM who was trained using EMRs from a hospital and continue its pretraining using only text from patients with a specific disease. It will allow the LM to adjust its weights and become more precise when interpreting texts related to that particular disease.

An example of domain adaptation is the BioBERT model [10], which resulted from the continuation of the pretraining of the Bidirectional Encoder Representations from Transformers (BERT) [10] model on biomedical text. The BioBERT model outperformed its predecessor in biomedical named entity recognition (NER), relation extraction, and question-answering tasks. Alzheimer's Disease-BERT [11] and CancerBERT [12] are 2 examples of applying domain adaptation to a more restricted domain. Both models outperformed their respective baselines on downstream tasks related to their respective diseases. Summing up, performing domain adaptation for the health care sector appears inevitable to improve results, for example, for information extraction (IE) models, where a better understanding of medical terminologies and lexicon would make it easier to identify and extract information [13].

The European Portuguese (PT-PT) language does not generate the same amount of data as the English language, resulting in limitations in the literature and the published models. A study published in 2023 by the Ernst & Young Audit highlights the following areas where AI can play a relevant role in Portugal's health care; disease diagnoses, precision medicine, remote monitoring and prevention, data management and hospital efficiency, and health policies [14]. Recently, a project was launched in Portugal, funded by the European Union, with the aim of creating PT-PT NLP solutions for the health care sector. Under this scope, the objective is to create PT-PT medical LM and IE models to automatically identify medically relevant entities.

Therefore, in this study, we aim to present a scoping literature review (SLR), in which we will begin by exploring the creation of health care LMs through domain adaptation and analyze their results. In addition, we aim to focus on the geographical domain to understand the current state-of-the-art for the Portuguese language and compare it to other, potentially more developed, languages to identify further steps. We also want to explore IE models in the health care sector, regardless of their data language, to understand the most commonly extracted medical entities and the methods used in doing so. Despite the literature being rich in studies focused on health care large language models (LLMs), there is a lack of studies that evaluate the current state-of-the-art of health care LMs not only in English but also in other less-resourced languages. This will enable us

to grasp how the community is using the capabilities of transformers and whether the advantages of using them are indeed present in the health care domain. In addition, researchers will have 1 study about health care LMs that could guide their path and help them understand how the literature has developed in their respective languages. Finally, we will present the corresponding discussion and the conclusion drawn from the SLR.

Methods

Overview

To complete our goal, we have conducted an SLR to gain a better understanding of the research conducted in the application of health care-domain LMs and the development of IE models within the health care domain. In the first stage, our study encompasses health care-domain LMs in various languages, with a focus on the Portuguese language. In the second stage, we searched for studies related to IE models to evaluate the methods most frequently used. In terms of methodology, we followed the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) [15] to ensure a systematic and transparent approach in conducting and reporting our scoping review.

Search Strategy and Inclusion Criteria

This SLR was conducted in November 2023 and focused exclusively on studies and reviews published in journals within the last 5 years (2019-2023) that were written in English or Portuguese. It was not an arbitrary data range since limiting the search to the last 5 years ensures that the review includes the most recent and relevant studies, reflecting the latest advancements, technologies, and methodologies in the field. Since BERT [10], one of the most popular transformer architectures, and LM were launched at the end of 2018, we searched for studies from 2019 onward. The primary databases used for this review were Scopus [16] and the Web of Science Core Collection (WOSCC) [17] since both databases are renowned for indexing a wide array of peer-reviewed journals across multidisciplinary fields [18,19]. While acknowledging that additional databases might offer further insights, the significant overlap with these resources ensures that relevant studies are unlikely to be missed.

The criteria were defined to include studies focused on continuing the pretraining of LMs to achieve health care-domain LMs or studies focused on creating IE models within the health care field. Therefore, we formulated a query that includes the training or fine-tuning LMs or IE Models within the context of health care or similar, using EMRs or EHRs as data.

Since there is a significant semantic similarity between LMs and LLMs, we decided to exclude the second from the search query because it has a different purpose from the aim of our study. LLMs are typically composed of more than 7B parameters and are suited for text generation. LMs are models that are not, by themselves, suited to perform any downstream NLP task, needing to be readjusted or fine-tuned with labeled data to be able to perform downstream tasks.

Our final query is as follows: (“Language Model” OR “Masked Language Model” OR “Information Extraction” OR “Content Extraction”) AND (“EHR” OR “EMR” OR “Electronic Health Record” OR “Electronic Medical Record”) AND (“Fine-Tuned” OR “Fine-tuning” OR “Training” OR “Trained”) AND (“Healthcare” OR “Health Care” OR “Clinical” OR “Medical”) AND NOT (“Large Language Model” OR “LLM”).

According to our objectives, a study was considered valid if it documented a continuation of the pretraining of an LM to create a health care LM or if it focused on the creation of health care IE models.

Study Selection

To minimize the risk of bias in the study selection, the process was conducted independently by 3 researchers. A total of 2 researchers were responsible for reading and judging the studies according to the inclusion criteria, while the third researcher was involved in cases of disagreement.

Data Charting and Synthesis

A data-charting form was jointly developed by two reviewers to extract relevant information from the selected studies systematically. The form included variables such as study title, year of publication, language focus (English or non-English), domain adaptation techniques for Transformer-based models, healthcare-specific information extraction tasks, evaluation metrics used, and the specific health-related entities being extracted. Both reviewers independently charted the data to ensure comprehensive coverage of healthcare language models in English and non-English languages, with particular attention to languages other than English (referred to as non-English). Discrepancies in the extracted data were discussed and resolved through consensus. As the review progressed, the data-charting form was iteratively updated to capture emerging themes, especially regarding the disparity between language resources and technological development for healthcare information extraction across different languages.

Results

The query retrieved 137 papers, with the vast majority of these studies being retrieved from Scopus, adding up to 90 when compared with the 47 studies WOSCC has yielded. The PRISMA-ScR methodology was then followed, as seen in Figure 1. Since we included studies from sources beyond the 2 selected databases, we adhered to the updated PRISMA-ScR guideline [20]. In the following subsections, we explained the decision to include studies by other methods.

The first step was to identify and remove duplicated papers, resulting in 101 studies. Following a screening of titles and abstracts, 10 records were deemed out of scope, and 1 could not be retrieved, leaving us with 90 fully reviewed studies.

After screening all the papers that matched our criteria, we realized that 30 of them did not meet our inclusion criteria. Some studies referred to the fine-tuning of pretrained LMs for tasks unrelated to IE, or they lacked relevant information to contribute to this study, or even though we excluded them from our search query, they mentioned the use of LLMs.

As we were focused on the Portuguese language, our study also emphasized the geographical domain, with an aim to comprehend the medical data language used in health care LMs. **Table 1** resumes the studies focused on the pretraining of LMs separated by the language of their data.

From the reading of **Table 1**, we can understand that English is the main language, which can be explained by the much higher availability of English data and the overwhelming presence and applicability of this language throughout the world. However, we notice that in the Chinese language, there are studies attempting to fill the gap of being non-English, creating in-domain LMs aware of their benefits. We also found studies in Brazilian Portuguese (PT-BR), Spanish, and PT-PT, and we

acknowledge that there might be other studies in different languages, even though they did not match our search query criteria.

Changing the view for the health care IE studies, **Figure 2** resumes the distribution of studies by topic.

From the reading of **Figure 2**, NER appears as the main topic on the IE, with only 2 studies performing Assertion Status and 3 studies focused on solving medical lexical problems.

To provide a more in-depth review of each study, we present the subsequent 3 subsections where we differentiate between non-Portuguese health care LMs, Portuguese health care LMs, and health care IE models.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) workflow diagram. WOSCC: Web of Science Core Collection.

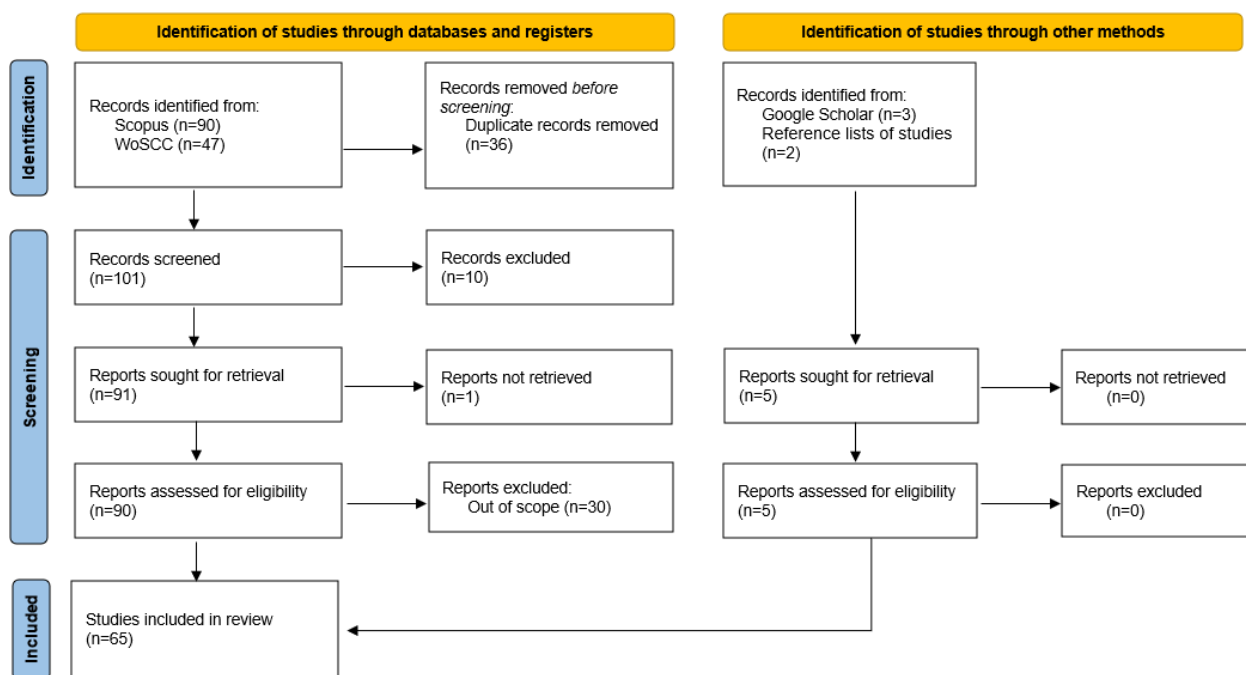
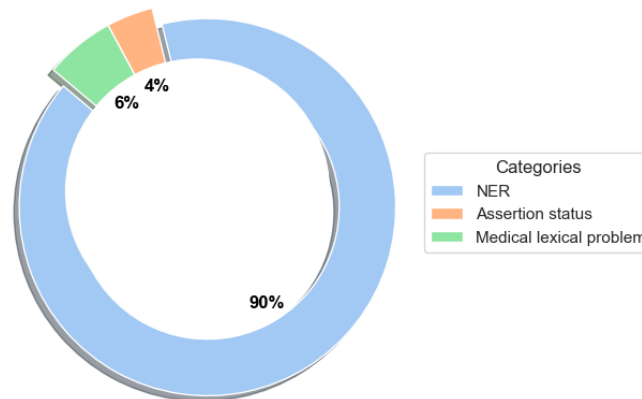


Table 1. Studies for pretraining language models (LMs) were reviewed by their data language.

Medical data language	Reference	Studies, n
English	[11,12,21-25]	7
Chinese	[26-29]	4
Brazilian Portuguese	[30,31]	2
Spanish	[32]	1
European Portuguese	[33]	1

Figure 2. Distribution of health care information extraction (IE) studies by topic. NER: named entity recognition.



Non-Portuguese Health Care Language Models

A study by Zhou et al [12] introduces CancerBERT, a domain-specific LM, that resulted from continuing the pretraining of the BlueBERT model [34] with a cancer corpus, resulting in various checkpoints of CancerBERT. The evaluation was performed for the cancer phenotyping NER task, with the results showing that the CancerBERT model pretrained with the cancer corpus outperformed the checkpoint using the original BERT [10] vocabulary.

A similar approach was conducted in a study by Mao et al [11], where the objective was to predict the risk of disease progression from Mild Cognitive Impairment to Alzheimer disease. A BERT model specifically tailored for Alzheimer disease (ie, AD-BERT) was pretrained with clinical notes, and its comparison with other models in experiments showed the benefits of domain adaptation.

Within the same scope, the identification of fall incidents from EHRs is discussed in the study by Fu et al [21]. A context-aware LM, BERT-based, was trained and integrated into a hybrid architecture along with post hoc heuristic rules. The performance of the BERT-based model was compared with DL methods, and the conclusions highlighted that the BERT model achieved superior results in identifying fall events.

In a study by Wang et al [26], a Chinese medical text corpus was used to pretrain BERT and obtain the Chinese BERT model. The results were aligned with previous studies, and domain adaptation demonstrated better results than traditional DL models and other pretrained LMs.

Studies by Roitero et al [22] and Agnikula Kshatriya et al [23] once again mention the pretraining of BERT models on a medical corpus, achieving comparable or better performance than state-of-the-art models. In a study by Zhang et al [24], an unsupervised adversarial domain adaptation framework with a pretrained LM for clinical event sequences is presented. Another example can be found in a study by Chen et al [25], where a contextual LM is used in combination with rule-based preprocessing methods to develop a model for *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) multilabel classification. The results demonstrate superiority over state-of-the-art models. Studies by Wen G et al and Wen C et al [27,28] refer to the training of a

domain-specific pretrained LM on unlabeled medical data, with the evaluation being made through NER.

In the process of reviewing all the papers, we observed references to papers that aligned with our requirements despite not being explicitly included within our search query criteria. It is the case of studies by Zhang et al [29] and Carrino et al [32]. Carrino et al [32] present a large-scale biomedical Spanish LM, where the models were pretrained from scratch, using a RoBERTa [35] base model, and then fine-tuned on 3 clinical NER tasks. The comparison between general-domain and other available Spanish clinical models revealed the superiority of the models presented in the paper. Zhang et al [29] share a similar scenario with BERT being pretrained on Chinese biomedical corpora, and MC-BERT, an in-domain LM, was developed. The results are consistent with previous studies, with MC-BERT outperforming BERT-based models in all evaluated tasks.

Portuguese Health Care Language Models

Our search query did not retrieve any studies for the Portuguese language. To address this scarcity of studies and since it is one of the objectives of this research, we carried out a broader search on Google Scholar [36] to include studies that mentioned the creation of Portuguese health care LMs.

The PT-BR language has already presented several studies, with BioBERTpt [30] being one example. The authors used clinical notes and biomedical abstracts to pretrain 3 BERT-based checkpoints that were fine-tuned for the NER task to assess their performance. The results align with others, showing that the in-domain models achieved better performance. Another example is the study by Schneider et al [31], where several clinical BERT-based checkpoints were developed resulting from the continuation of the pretraining of BERTimbau [35], mBERT [11], and all 3 BioBERTpt checkpoints on 150,000 clinical narratives from cardiology ambulatory. The results of fine-tuning for NER align with previous studies, demonstrating that the in-domain models outperformed general LMs.

For the PT-PT language, we found the literature to be scarcer, with only 1 study being found that mentioned the continuation of the pretraining of an LM to achieve health care-domain LM. Coutinho and Martins [33] propose a BERT-based model for assigning *ICD-10* codes to causes of death by using BERTimbau and continuing its pretraining with death certificates. The

evaluation was made through NER, with all the checkpoints involved being fine-tuned for the classification task, and the findings indicated that transformer models could produce promising outcomes for health care tasks involving the analysis of relatively short documents.

Health Care IE Models

To better organize this section, we decided to categorize the studies by topic. Therefore, the first subsection presents NER studies where the authors attempted to automatically identify and extract medical information. The second subsection contains Assertion Status models, where entities are classified according to their status (present or absent), and finally, the third subsection presents studies that attempt to solve medical lexical problems.

NER

Zhou et al [37] evaluated the performance of CancerBERT along with ML models for the breast cancer phenotype extraction task, with the results proving that CancerBERT has superior learning ability and generalizability for this task. Rahman et al [38] refer to the use of BERT to identify the presence of a diagnosis in EHRs. With BERT's ability to understand the context of text and based on conditions presented in EHRs, a pipeline was successfully designed to identify EHRs with the presence of a diagnosis, reducing the manual note review load. Crema et al [39], use an Italian biomedical BERT model, fine-tuning it for the NER task with the entities of interest, including diagnoses, symptoms, drugs, and medical assessments, achieving an F_1 -score of nearly 0.85 values. Entity-BERT was introduced in the study by Lu et al [40], a DL-based model for entity IE that is capable of recognizing entities such as medical terminologies, disease names, or drug information. Zhang et al [41] propose the combination of data augmentation and domain information using the Adapter Transformer Encoder Model for Clinical Event Detection. It uses the BioBERT model to generate word-level features, addressing the issue of many obscure professional terms in EMRs leading to poor recognition performance of the model. The results were reported to be superior to those of other existing models. A multilingual transformer was fine-tuned in a study by Kim et al [42], where researchers successfully extracted alcohol-related information from unstructured clinical notes with an extraction accuracy of 0.84 as measured by the macro F_1 -score. Kormilitzin et al [43], initially pretrained a model on the task of predicting the next word and subsequently fine-tuned it for the NER task, extracting various categories of drugs and achieving performance with an F_1 -score above 0.95 values. Solarte-Pabón et al [44] evaluate the fine-tuning of several pretrained LMs for the NER task, aiming to identify breast cancer concepts in the Spanish language. The results show that BERT-based and RoBERTa-based LMs exhibit competitive performance on this task. Liu et al [45] propose the use of BERT-BiLSTM-CRF for the NER task of rheumatoid arthritis vocabulary and then MC-BERT for the entity extraction task, with results showing F_1 -scores above 90%. Wang et al [46] compare the use of 4 pretrained transformer-based LMs fine-tuned for the NER task with a baseline regular expression model in order to extract ophthalmic examination components, demonstrating that

transformers achieve superior results. In the study by Singh et al [47], a pretrained transformer-based LM was fine-tuned with cardiac magnetic resonance imaging annotations to effectively extract measurements from clinical reports, and it achieved high extraction performance without requiring heuristics or expert annotations.

Several studies focus on extracting information about family history, such as studies by Kim et al [48], He et al [49], Silva et al [50], Dai et al [51], and Zhan et al [52]. They use ML methods, incorporating rule-based approaches, multitask-based artificial neural networks (ANN), attention-based neural networks, and even combinations such as convolutional neural networks (CNNs) BiLSTM and BERT. The goal was to automatically extract entities such as people's names, residence, birth date, or death date, and in some cases, there is an additional subtask related to relation extraction, which involves identifying relationships between family members. Overall, the results have proven to be satisfactory, particularly in the NER task.

CNNs are highly popular methods in the scientific community for extracting clinical information and studies by Yang et al [53], Santus et al [54], Mahajan and Rana [55], and Landlosi et al [56] primarily used them, often supplemented with rule-based approaches or feature optimization in some cases. The use of these methods lies in extracting clinical information from EHRs, tasks that could be time-consuming if done manually. Within the broader category of neural networks, RNNs are also a method used for IE in which the authors of studies [57-66] all use RNNs, with BiLSTM-CRF (Bidirectional Long Short-Term Memory - Conditional Random Field) being a very popular network among these studies. The main topics extracted include terms related to specific diseases, drug names with associated attributes (dosage, frequency, duration, route, and condition), adverse drug events, the presence of a diagnosis, or even important information in medical image reports, with the results globally proven to be promising.

Studies [67-69] use ML methods, with the first focusing on automatically classifying the outcomes of specific tasks related to the clinical conditions of stroke survivors, the second aiming to extract useful information in abdominopelvic radiology reports, and the third one focused on extracting travel history mentions from clinical documents. In Malmasi et al [70], the use of different methods to extract low-prevalence concepts is discussed, specifically in the case of insulin rejection by patients with attempts at both sentence-level and token-level approaches using ML and DL methods, but the results showed that it is challenging to automatically identify low-prevalence concepts. Similar proposals have been presented in studies [71-79] using spaCy's [80] pipeline for IE, contextual embeddings such as embeddings from language models (ELMo) [81] and BERT, position-attention mechanisms, knowledge graph embeddings, word segmentation models, or even NLP models developed using Java for extracting medical information, for example, extracting details related to drugs, drug attributes, or diagnoses.

In Lee and Uppal [82], a web-based summarization and visualization tool is introduced for extracting salient information from clinical and biomedical text, featuring sentence ranking by relevance and facilitating early medical risk detection in

clinical settings. Chen et al [83] aimed to create a model to extract concept embeddings from EHRs for disease pattern retrieval and subsequent classification tasks.

Assertion Status

Sykes et al [84] address the issue of negation and non-negation of clinical terms in EHRs. It is an Assertion Status case, in which the text can be characterized by cases where diseases are stated to be absent or only hypothesized. In this study, they propose various methods to address this issue, including rule-based, ML, or DL approaches, and all proposals yielded good results in a test set, achieving an F_1 -score of more than 0.95. In Chaturvedi et al [85], a corpus annotated with mentions of pain was developed, considering the presence or absence of pain. It is another example of an Assertion Status problem aimed at facilitating further studies using the corpus to better understand how pain is mentioned in clinical notes.

Medical Lexical Problems

From a different perspective, there have been studies focusing on medical lexical problems. Newman-Griffis et al [86] discuss the presence of ambiguous words and attempts to normalize medical concepts to standardize vocabularies, while the study by Jaber et al [87] addresses the problem of the frequent use of abbreviations by proposing a method, by fine-tuning a pretrained LM, to successfully disambiguate clinical abbreviations. Lee et al [88] propose a typographical error correction model that considers context, based on a masked LM, to address the issue of typographical errors in real-world medical data. They conclude that typographical errors in unstructured text negatively impact the performance of NLP tasks, and their method is robust and applicable in real-world environments.

Discussion

Principal Findings

Continuing the pretraining of LMs to develop health care LMs has proven beneficial. The most common method to evaluate this approach is by fine-tuning both the baseline and the in-domain LM on downstream NLP tasks and comparing the results.

In IE models, NER is the most popular topic aimed at automatically identifying and extracting medically relevant information. Transformers are the preferred technology for this purpose, with fine-tuning of medical LMs consistently achieving superior results.

To conclude our SLR, we engaged in a deeper discussion divided into health care LMs and health care IE models.

Health Care Language Models

On a global scale, we have identified numerous studies that continued the pretraining of LMs to develop domain-specific LM, specifically medical LMs. In general, the findings across almost all of these studies substantiate the advantages of in-domain training before undertaking any other downstream

tasks. The favorite evaluation task is NER, with almost every study mentioning the fine-tuning of LMs for the NER task.

As shown in Table 1, English and Chinese are the languages with the most studies and published models due to the available resources in terms of data and hardware power. The level of domain adaptation for these languages is more advanced, with dedicated health care LMs developed for specific diseases such as Alzheimer Disease-BERT [11] and CancerBERT [12], which represent very focused domains. These studies offer advantages by achieving better performance in extracting specific concepts from textual data related to these diseases compared with general health care LMs.

For non-English languages, the process is not so developed, which can be considered as expected since they have their known limitations, such as the scarcity of data and resources available. Nevertheless, there have been concerted efforts to create general health care LMs, underscoring the community's recognition of the use of these models. The Portuguese language fits this context, and despite initial strides that have already been taken, there exists ample room for improvement, particularly in the context of PT-PT where the only published study is [33], yet, to the best of our knowledge, the model is not publicly available.

Non-English languages, particularly Portuguese, should draw inspiration from advancements and results in medical domain adaptation studies. Despite limited resources and available data, efforts should first focus on creating general medical LMs. In a subsequent phase, efforts should be directed toward narrowing down to specific diseases while performing domain adaptation. This approach ensures that knowledge previously acquired by the LMs is refined within the medical domain and then adapted to smaller medical domains without losing the previously acquired knowledge completely. This initiative aims to foster the development of AI technologies in Portuguese, thereby promoting health care equality and access in languages with fewer resources. These models can be further fine-tuned for medical NLP tasks, such as IE, aimed at automatically identifying or highlighting specific information or structuring medical information extracted from textual data for ML analysis to aid health care professionals.

Health Care IE Models

Several methods have been used to create health care IE models. The most common method is the use of transformers, followed by the application of other DL and ML methods (Table 2). As previously discussed, the most popular topic was NER, where authors attempted to identify and extract medically relevant information.

The results indicate that the most successful approach involves using pretrained LMs fine-tuned for IE tasks, benefiting from the contextual understanding of the text to achieve better results. The most commonly identified entities were diagnoses or diseases and drugs, along with specific phenotypes related to certain diseases.

Table 2. Number of studies used per method.

Methods	Articles
Transformers	16
Other DL ^a	15
RNN ^b	10
Other ML ^c	9
CNN ^d	5
Rule-based	4

^aDL: deep learning.

^bRNN: recurrent neural network.

^cML: machine learning.

^dCNN: convolutional neural network.

It is also relevant to mention that in our query, 2 studies were focused on Assertion Status. This task involves classification at the sentence level aimed at assessing an entity based on its presence or absence in the text. Examples of absence include negation or hypothesizing medical information. From another perspective, to address the problems presented by medical text, we also found 2 studies that propose solutions to disambiguate the multitude of abbreviations present in medical text and 1 study that presents a typo correction model. Both solutions aimed to improve text quality and seek to correct issues in the text that are considered inevitable by health care professionals. These 5 studies could also be seen as an improvement to NER results. The ones focused on correcting the text could be viewed as a preprocessing step that would enhance the understanding of the medical text, while the Assertion Status studies could help ascertain whether an identified entity is present or absent in a patient's condition. When compared with the distribution of NER, these 2 topics lack development, as together, they account for only 10% of the health care IE studies found. The community would benefit from more studies using different technologies and identifying new challenges to be solved.

Conclusions

Our SLR highlights the benefits of in-domain training for health care LMs and the effectiveness of transformers in IE tasks, addressing a research gap regarding the lack of studies on health care LMs. Transformers excel in NER, identifying diagnoses, diseases, drugs, and phenotypes. English and Chinese lead in research and LM development, while non-English languages such as Portuguese show promise but need exploration. Challenges include Assertion Status and text disambiguation, necessitating diverse methodologies and research in health care IE.

We have identified several health care-domain LMs, but there is a clear gap for non-English languages where the data and resources available are low. There is much to improve in those languages, with Portuguese being an example. The benefits of creating a medical-domain LM are already proven, and the health care sector could benefit greatly from a symbiosis with AI. Therefore, non-English languages should be motivated by the scarce studies already published and try to replicate them for their own language in order to fill this existing gap.

From another point of view, the use of transformers appears to be the better technique to automatically identify medical information. Despite the annotation process for any supervised learning task being very time-consuming, transformers achieve better results on fewer annotations, making their usage on new tasks relatively easier. This task also benefits from an in-domain medical LM. The entities most commonly extracted are diagnosis or disease, posology-related entities, symptoms, and phenotypes related to specific diseases.

Despite our belief that this was the right choice, we highlight the 2 databases that we searched, and we acknowledge that, despite our best efforts, there is always a possibility that not all relevant papers will be found when formulating a query. These are the limitations of our study. The chosen timeframe may also limit the availability of relevant studies, even though we believe it is the right timeframe to include studies that establish the current state-of-the-art with new technologies. While we focused on the Portuguese language, we acknowledge that our conclusions cannot be generalized to all non-English languages. However, other languages with similar characteristics in terms of available data and resources can certainly gain insights from this SLR.

Globally, the development and research in these topics for the English language are very advanced compared with non-English languages. In English, several studies have been presented that perform domain adaptation for smaller domains, such as specific-disease LMs, which have improved results in extracting medical information related to these diseases. The next steps should involve continuing the pretraining for different medical areas or diseases to ensure the most comprehensive coverage with LMs. In addition, fine-tuning the already available models to meet the specific requirements of health care professionals is essential.

Non-English languages are still performing domain adaptation for general domains, such as medical or biomedical fields, and should be motivated by these studies to overcome the barriers inherent in their respective language. In the next step, they should focus on performing domain adaptation, aiming to narrow down to specific medical areas or diseases. They should strive to replicate studies on Assertion Status or even those focused on resolving the frequent presence of abbreviations and typos

in the text. In non-English languages where there is a scarcity of available data, it would be beneficial to have open corpora, even if distributed under licenses that protect data privacy, to enable more researchers to develop models.

These types of studies are important to assess and guide the development of non-English languages trying to bridge the gaps and capitalize on the opportunities provided by these technologies to promote equity and improve access to health

care all over the world. The differences in the available data and resources are almost impossible to correct but at least should be minimized.

This effort aims to harness AI to enhance health care by developing advanced LMs tailored for non-English languages, thereby supporting health care professionals with decision-making tools that alleviate their workload and improve patient care indirectly.

Acknowledgments

MN's work was supported by the project Blockchain.PT, (PRR RE-C05-i01.02: AGENDAS/ALIANÇAS VERDES PARA A INOVAÇÃO EMPRESARIAL).

Authors' Contributions

MN, JB, and LBE performed conceptualization. MN and LBE conducted the investigation. JB, JCF, and LBE performed supervision. JB, LBE, and JCF conducted validation. MN wrote the original draft. JB, JCF, and LBE performed review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR checklist.

[\[PDF File \(Adobe PDF File\), 509 KB-Multimedia Appendix 1\]](#)

References

1. HealthTech. How to navigate structured and unstructured data as a healthcare organization. URL: <https://healthtechmagazine.net/article/2023/05/structured-vs-unstructured-data-in-healthcare-perfcon> [accessed 2024-06-14]
2. RBC Capital Markets. Navigating the changing face of healthcare episode. URL: <https://www.rbccm.com/en/gib/healthcare/story.page> [accessed 2023-10-30]
3. Tayefi M, Ngo P, Chomutare T, Dalianis H, Salvi E, Budrionis A, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. WIREs Computational Stats. 2021;13(6):e1549. [doi: [10.1002/wics.1549](https://doi.org/10.1002/wics.1549)]
4. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Computer Science Computation and Language. 2023. [doi: [10.5860/choice.189890](https://doi.org/10.5860/choice.189890)]
5. Chernyavskiy A, Ilvovsky D, Nakov P. Transformers: "The End of History" for NLP? Computer Science Computation and Language. 2021. [FREE Full text]
6. Rokon OF. RNN vs. LSTM vs. Transformers: unraveling the secrets of sequential data processing. Medium. 2023. URL: <https://tinyurl.com/432k5mn5> [accessed 2024-05-20]
7. How do Transformers work? - Hugging Face NLP Course. URL: <https://huggingface.co/learn/nlp-course/chapter1/4> [accessed 2023-12-06]
8. Farahani A, Voghoei S, Rasheed K, Arabnia HR. A brief review of domain adaptation. 2020. URL: <http://arxiv.org/abs/2010.03978> [accessed 2024-04-21]
9. Guo X, Yu H. On the domain adaptation and generalization of pretrained language models: a survey. 2022. URL: <http://arxiv.org/abs/2211.03154> [accessed 2024-04-21]
10. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. ACL Anthology. 2019. [doi: [10.5260/chara.21.2.8](https://doi.org/10.5260/chara.21.2.8)]
11. Mao C, Xu J, Rasmussen L, Li Y, Adekkanattu P, Pacheco J, et al. AD-BERT: using pre-trained language model to predict the progression from mild cognitive impairment to alzheimer's disease. J Biomed Inform. 2023;144:104442. [FREE Full text] [doi: [10.1016/j.jbi.2023.104442](https://doi.org/10.1016/j.jbi.2023.104442)] [Medline: [37429512](https://pubmed.ncbi.nlm.nih.gov/37429512/)]
12. Zhou S, Wang N, Wang L, Liu H, Zhang R. CancerBERT: a cancer domain-specific language model for extracting breast cancer phenotypes from electronic health records. J Am Med Inform Assoc. 2022;29(7):1208-1216. [FREE Full text] [doi: [10.1093/jamia/ocac040](https://doi.org/10.1093/jamia/ocac040)] [Medline: [35333345](https://pubmed.ncbi.nlm.nih.gov/35333345/)]
13. Wang B, Xie Q, Pei J, Chen Z, Tiwari P, Li Z, et al. Pre-trained language models in biomedical domain: a systematic survey. ACM Comput Surv. 2023;56(3):1-52. [doi: [10.1145/3611651](https://doi.org/10.1145/3611651)]
14. A Inteligência Artificial na Saúde, uma Breve Perspectiva. URL: https://www.ey.com/pt_pt/health/a-inteligencia-artificial-na-saude-uma-breve-perspectiva [accessed 2024-05-18]

15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8(5):336-341. [FREE Full text] [doi: [10.1016/j.jisu.2010.02.007](https://doi.org/10.1016/j.jisu.2010.02.007)] [Medline: [20171303](https://pubmed.ncbi.nlm.nih.gov/20171303/)]
16. Scopus - Document search. URL: <https://www.scopus.com/search/form.uri?display=basic#basic> [accessed 2023-05-15]
17. Web of science core collection. Clarivate. URL: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/webofscience-platform/web-of-science-core-collection/> [accessed 2024-01-26]
18. Pranckutė R. Web of science (WoS) and scopus: The titans of bibliographic information in today's academic world. *Publications*. 2021;9(1):12. [doi: [10.3390/publications9010012](https://doi.org/10.3390/publications9010012)]
19. Zhu J, Liu W. A tale of two databases: the use of web of science and scopus in academic papers. *Scientometrics*. 2020;123(1):321-335. [doi: [10.1007/s11192-020-03387-8](https://doi.org/10.1007/s11192-020-03387-8)]
20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. [FREE Full text] [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
21. Fu S, Thorsteinsdottir B, Zhang X, Lopes GS, Pagali SR, LeBrasseur NK, et al. A hybrid model to identify fall occurrence from electronic health records. *Int J Med Inf*. 2022;162:104736. [FREE Full text] [doi: [10.1016/j.ijmedinf.2022.104736](https://doi.org/10.1016/j.ijmedinf.2022.104736)] [Medline: [35316697](https://pubmed.ncbi.nlm.nih.gov/35316697/)]
22. Roitero K, Portelli B, Popescu MH, Mea VD. DiLBERT: cheap embeddings for disease related medical NLP. *IEEE Access*. 2021;9:159714-159723. [doi: [10.1109/access.2021.3131386](https://doi.org/10.1109/access.2021.3131386)]
23. Agnikula Kshatriya BS, Sagheb E, Wi CI, Yoon J, Seol HY, Juhn Y, et al. Identification of asthma control factor in clinical notes using a hybrid deep learning model. *BMC Med Inform Decis Mak*. 2021;21(Suppl 7):272. [FREE Full text] [doi: [10.1186/s12911-021-01633-4](https://doi.org/10.1186/s12911-021-01633-4)] [Medline: [34753481](https://pubmed.ncbi.nlm.nih.gov/34753481/)]
24. Zhang T, Chen M, Bui AAT. AdaDiag: adversarial domain adaptation of diagnostic prediction with clinical event sequences. *J Biomed Inform*. 2022;134:104168. [FREE Full text] [doi: [10.1016/j.jbi.2022.104168](https://doi.org/10.1016/j.jbi.2022.104168)] [Medline: [35987449](https://pubmed.ncbi.nlm.nih.gov/35987449/)]
25. Chen P-F, Chen K-C, Liao W-C, Lai F, He T-L, Lin S-C, et al. Automatic international classification of diseases coding system: deep contextualized language model with rule-based approaches. *JMIR Med Inform*. 2022;10(6):e37557. [FREE Full text] [doi: [10.2196/37557](https://doi.org/10.2196/37557)] [Medline: [35767353](https://pubmed.ncbi.nlm.nih.gov/35767353/)]
26. Wang J, Zhang G, Wang W, Zhang K, Sheng Y. Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. *J Cloud Comp*. 2021;10(1). [doi: [10.1186/s13677-020-00218-2](https://doi.org/10.1186/s13677-020-00218-2)]
27. Wen G, Chen H, Li H, Hu Y, Li Y, Wang C. Cross domains adversarial learning for Chinese named entity recognition for online medical consultation. *J Biomed Inform*. 2020;112:103608. [FREE Full text] [doi: [10.1016/j.jbi.2020.103608](https://doi.org/10.1016/j.jbi.2020.103608)] [Medline: [33132138](https://pubmed.ncbi.nlm.nih.gov/33132138/)]
28. Wen C, Chen T, Jia X, Zhu J. Medical named entity recognition from un-labelled medical records based on pre-trained language models and domain dictionary. *Data Intell*. 2021;3(3):402-417. [doi: [10.1162/dint_a_00105](https://doi.org/10.1162/dint_a_00105)]
29. Zhang N, Jia Q, Yin K, Dong L, Gao F, Hua N. Conceptualized representation learning for Chinese biomedical text mining. *Computer Science Computation and Language*. 2020. URL: <http://arxiv.org/abs/2008.10813> [accessed 2023-10-19]
30. Schneider ETR, de Souza JVA, Knafou J, Oliveira LES, Copara J, Gumiel YB, et al. BioBERTpt - a portuguese neural language model for clinical named entity recognition. Association for Computational Linguistics; 2020. Presented at: Proceedings of the 3rd Clinical Natural Language Processing Workshop; August 27, 2024:65-72; USA. [doi: [10.18653/v1/2020.clinicalnlp-1.7](https://doi.org/10.18653/v1/2020.clinicalnlp-1.7)]
31. Schneider ETR, Gumiel YB, De Souza JVA, Mie ML, Emanuel SEOL, De SRM, et al. CardioBERTpt: transformer-based models for cardiology language representation in portuguese. 2023. Presented at: Proceedings of the 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS); June 22-24, 2023:378-381; L'Aquila, Italy. [doi: [10.1109/CBMS58004.2023.00247](https://doi.org/10.1109/CBMS58004.2023.00247)]
32. Carrino CP, Llop J, Pàmies M, Gutiérrez-Fandiño A, Armengol-Estapé J, Silveira-Ocampo J, et al. Pre-trained biomedical language models for clinical NLP in Spanish. *ACL Anthology*. 2022:193-199. [doi: [10.18653/v1/2022.bionlp-1.19](https://doi.org/10.18653/v1/2022.bionlp-1.19)]
33. Coutinho I, Martins B. Transformer-based models for ICD-10 coding of death certificates with portuguese text. *J Biomed Inform*. 2022;136:104232. [FREE Full text] [doi: [10.1016/j.jbi.2022.104232](https://doi.org/10.1016/j.jbi.2022.104232)] [Medline: [36307020](https://pubmed.ncbi.nlm.nih.gov/36307020/)]
34. Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural Language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. *Computer Science Computation and Language*. 2019. [doi: [10.18653/v1/w19-5006](https://doi.org/10.18653/v1/w19-5006)]
35. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. *Annual Reviews*. 2019;21(2):8-10. [doi: [10.5260/chara.21.2.8](https://doi.org/10.5260/chara.21.2.8)]
36. Google Académico. URL: <https://scholar.google.com/> [accessed 2023-11-03]
37. Zhou S, Wang N, Wang L, Sun J, Blaes A, Liu H, et al. A cross-institutional evaluation on breast cancer phenotyping NLP algorithms on electronic health records. *Comput Struct Biotechnol J*. 2023;22:32-40. [FREE Full text] [doi: [10.1016/j.csbj.2023.08.018](https://doi.org/10.1016/j.csbj.2023.08.018)] [Medline: [37680211](https://pubmed.ncbi.nlm.nih.gov/37680211/)]
38. Rahman P, Ye C, Mittendorf KF, Lenoue-Newton M, Micheel C, Wolber J, et al. Accelerated curation of checkpoint inhibitor-induced colitis cases from electronic health records. *JAMIA Open*. 2023;6(1):ooad017. [FREE Full text] [doi: [10.1093/jamiaopen/ooad017](https://doi.org/10.1093/jamiaopen/ooad017)] [Medline: [37012912](https://pubmed.ncbi.nlm.nih.gov/37012912/)]

39. Crema C, Buonocore TM, Fostinelli S, Parimbelli E, Verde F, Fundarò C, et al. Advancing italian biomedical information extraction with transformers-based models: methodological insights and multicenter practical application. *J Biomed Inform.* 2023;148:104557. [FREE Full text] [doi: [10.1016/j.jbi.2023.104557](https://doi.org/10.1016/j.jbi.2023.104557)] [Medline: [38012982](https://pubmed.ncbi.nlm.nih.gov/38012982/)]
40. Lu W, Jiang J, Shi Y, Zhong X, Gu J, Huangfu L, et al. Application of entity-BERT model based on neuroscience and brain-like cognition in electronic medical record entity recognition. *Front Neurosci.* 2023;17:1259652. [FREE Full text] [doi: [10.3389/fnins.2023.1259652](https://doi.org/10.3389/fnins.2023.1259652)] [Medline: [37799340](https://pubmed.ncbi.nlm.nih.gov/37799340/)]
41. Zhang Z, Liu D, Zhang M, Qin X. Combining data augmentation and domain information with TENER model for clinical event detection. *BMC Med Inform Decis Mak.* 2021;21(Suppl 9):261. [FREE Full text] [doi: [10.1186/s12911-021-01618-3](https://doi.org/10.1186/s12911-021-01618-3)] [Medline: [34789246](https://pubmed.ncbi.nlm.nih.gov/34789246/)]
42. Kim HK, Park Y, Park Y, Choi E, Kim S, You H, et al. Identifying alcohol-related information from unstructured bilingual clinical notes with multilingual transformers. *IEEE Access.* 2023;11:16066-16075. [doi: [10.1109/access.2023.3245523](https://doi.org/10.1109/access.2023.3245523)]
43. 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. [doi: [10.1016/j.artmed.2021.102086](https://doi.org/10.1016/j.artmed.2021.102086)] [Medline: [34412834](https://pubmed.ncbi.nlm.nih.gov/34412834/)]
44. Solarte-Pabón O, Montenegro O, García-Barragán A, Torrente M, Provencio M, Menasalvas E, et al. Transformers for extracting breast cancer information from Spanish clinical narratives. *Artif Intell Med.* 2023;143:102625. [FREE Full text] [doi: [10.1016/j.artmed.2023.102625](https://doi.org/10.1016/j.artmed.2023.102625)] [Medline: [37673566](https://pubmed.ncbi.nlm.nih.gov/37673566/)]
45. Liu F, Liu M, Li M, Xin Y, Gao D, Wu J, et al. Automatic knowledge extraction from chinese electronic medical records and rheumatoid arthritis knowledge graph construction. *Quant Imaging Med Surg.* 2023;13(6):3873-3890. [FREE Full text] [doi: [10.21037/qims-22-1158](https://doi.org/10.21037/qims-22-1158)] [Medline: [37284084](https://pubmed.ncbi.nlm.nih.gov/37284084/)]
46. 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. [FREE Full text] [doi: [10.1016/j.ijmedinf.2022.104864](https://doi.org/10.1016/j.ijmedinf.2022.104864)] [Medline: [36179600](https://pubmed.ncbi.nlm.nih.gov/36179600/)]
47. Singh P, Haimovich J, Reeder C, Khurshid S, Lau ES, Cunningham JW, et al. One clinician is all you need-cardiac magnetic resonance imaging measurement extraction: deep learning algorithm development. *JMIR Med Inform.* 2022;10(9):e38178. [FREE Full text] [doi: [10.2196/38178](https://doi.org/10.2196/38178)] [Medline: [35960155](https://pubmed.ncbi.nlm.nih.gov/35960155/)]
48. Kim Y, Heider PM, Lally IRH, Meystre SM. A hybrid model for family history information identification and relation extraction: development and evaluation of an end-to-end information extraction system. *JMIR Med Inform.* 2021;9(4):e22797. [FREE Full text] [doi: [10.2196/22797](https://doi.org/10.2196/22797)] [Medline: [33885370](https://pubmed.ncbi.nlm.nih.gov/33885370/)]
49. He K, Yao L, Zhang J, Li Y, Li C. Construction of genealogical knowledge graphs from obituaries: multitask neural network extraction system. *J Med Internet Res.* 2021;23(8):e25670. [FREE Full text] [doi: [10.2196/25670](https://doi.org/10.2196/25670)] [Medline: [34346903](https://pubmed.ncbi.nlm.nih.gov/34346903/)]
50. Silva JF, Almeida JR, Matos S. Extraction of family history information from clinical notes: deep learning and heuristics approach. *JMIR Med Inform.* 2020;8(12):e22898. [FREE Full text] [doi: [10.2196/22898](https://doi.org/10.2196/22898)] [Medline: [33372893](https://pubmed.ncbi.nlm.nih.gov/33372893/)]
51. Dai H-J, Lee Y-Q, Nekkanti C, Jonnagaddala J. Family history information extraction with neural attention and an enhanced relation-side scheme: algorithm development and validation. *JMIR Med Inform.* 2020;8(12):e21750. [doi: [10.2196/21750](https://doi.org/10.2196/21750)] [Medline: [33258777](https://pubmed.ncbi.nlm.nih.gov/33258777/)]
52. Zhan K, Peng W, Xiong Y, Fu H, Chen Q, Wang X, et al. Novel graph-based model with biaffine attention for family history extraction from clinical text: modeling study. *JMIR Med Inform.* 2021;9(4):e23587. [FREE Full text] [doi: [10.2196/23587](https://doi.org/10.2196/23587)] [Medline: [33881405](https://pubmed.ncbi.nlm.nih.gov/33881405/)]
53. Yang Z, Pou-Prom C, Jones A, Banning M, Dai D, Mamdani M, et al. Assessment of natural language processing methods for ascertaining the expanded disability status scale score from the electronic health records of patients with multiple sclerosis: algorithm development and validation study. *JMIR Med Inform.* 2022;10(1):e25157. [FREE Full text] [doi: [10.2196/25157](https://doi.org/10.2196/25157)] [Medline: [35019849](https://pubmed.ncbi.nlm.nih.gov/35019849/)]
54. Santus E, Li C, Yala A, Peck D, Soomro R, Faridj N, et al. Do neural information extraction algorithms generalize across institutions? *JCO Clin Cancer Inform.* 2019;3:1-8. [FREE Full text] [doi: [10.1200/CCI.18.00160](https://doi.org/10.1200/CCI.18.00160)] [Medline: [31310566](https://pubmed.ncbi.nlm.nih.gov/31310566/)]
55. Mahajan P, Rana D. Feature optimization in CNN using MROA for disease classification. *IDT.* 2023;17(2):301-315. [doi: [10.3233/idt-220097](https://doi.org/10.3233/idt-220097)]
56. Landolsi MY, Hlaoua L, Romdhane LB. Hybrid method to automatically extract medical document tree structure. *Eng Appl Artif Intell.* 2023;120:105922. [doi: [10.1016/j.engappai.2023.105922](https://doi.org/10.1016/j.engappai.2023.105922)]
57. Wunnava S, Qin X, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. *Drug Saf.* 2019;42(1):113-122. [doi: [10.1007/s40264-018-0765-9](https://doi.org/10.1007/s40264-018-0765-9)] [Medline: [30649736](https://pubmed.ncbi.nlm.nih.gov/30649736/)]
58. Viani N, Miller TA, Napolitano C, Priori SG, Savova GK, Bellazzi R, et al. Supervised methods to extract clinical events from cardiology reports in Italian. *J Biomed Inform.* 2019;95:103219. [FREE Full text] [doi: [10.1016/j.jbi.2019.103219](https://doi.org/10.1016/j.jbi.2019.103219)] [Medline: [31150777](https://pubmed.ncbi.nlm.nih.gov/31150777/)]
59. Wang S, Pang M, Pan C, Yuan J, Xu B, Du M, et al. Information extraction for intestinal cancer electronic medical records. *IEEE Access.* 2020;8:125923-125934. [doi: [10.1109/access.2020.3005684](https://doi.org/10.1109/access.2020.3005684)]
60. Ju M, Short AD, Thompson P, Bakerly ND, Gkoutos GV, Tsaprouni L, et al. Annotating and detecting phenotypic information for chronic obstructive pulmonary disease. *JAMIA Open.* 2019;2(2):261-271. [FREE Full text] [doi: [10.1093/jamiaopen/ooz009](https://doi.org/10.1093/jamiaopen/ooz009)] [Medline: [31984360](https://pubmed.ncbi.nlm.nih.gov/31984360/)]

61. Lopes F, Teixeira C, Gonçalo Oliveira H. Comparing different methods for named entity recognition in portuguese neurology text. *J Med Syst.* 2020;44(4):77. [doi: [10.1007/s10916-020-1542-8](https://doi.org/10.1007/s10916-020-1542-8)] [Medline: [32112285](https://pubmed.ncbi.nlm.nih.gov/32112285/)]
62. Alfattni G, Belousov M, Peek N, Nenadic G. Extracting drug names and associated attributes from discharge summaries: text mining study. *JMIR Med Inform.* 2021;9(5):e24678. [FREE Full text] [doi: [10.2196/24678](https://doi.org/10.2196/24678)] [Medline: [33949962](https://pubmed.ncbi.nlm.nih.gov/33949962/)]
63. Chen T, Dredze M, Weiner JP, Hernandez L, Kimura J, Kharrazi H. Extraction of geriatric syndromes from electronic health record clinical notes: assessment of statistical natural language processing methods. *JMIR Med Inform.* 2019;7(1):e13039. [FREE Full text] [doi: [10.2196/13039](https://doi.org/10.2196/13039)] [Medline: [30862607](https://pubmed.ncbi.nlm.nih.gov/30862607/)]
64. Li Z, Ren J. Fine-tuning ERNIE for chest abnormal imaging signs extraction. *J Biomed Inform.* 2020;108:103492. [FREE Full text] [doi: [10.1016/j.jbi.2020.103492](https://doi.org/10.1016/j.jbi.2020.103492)] [Medline: [32645382](https://pubmed.ncbi.nlm.nih.gov/32645382/)]
65. Jouffroy J, Feldman SF, Lerner I, Rance B, Burgun A, Neuraz A. Hybrid deep learning for medication-related information extraction from clinical texts in French: MedExt algorithm development study. *JMIR Med Inform.* 2021;9(3):e17934. [doi: [10.2196/17934](https://doi.org/10.2196/17934)]
66. Li C, Ma K. Entity recognition of Chinese medical text based on multi-head self-attention combined with BILSTM-CRF. *Math Biosci Eng.* 2022;19(3):2206-2218. [FREE Full text] [doi: [10.3934/mbe.2022103](https://doi.org/10.3934/mbe.2022103)] [Medline: [35240782](https://pubmed.ncbi.nlm.nih.gov/35240782/)]
67. Zannotto BS, Beck da Silva Etges AP, Dal Bosco A, Cortes EG, Ruschel R, De Souza AC, et al. Stroke outcome measurements from electronic medical records: cross-sectional study on the effectiveness of neural and nonneural classifiers. *JMIR Med Inform.* 2021;9(11):e29120. [FREE Full text] [doi: [10.2196/29120](https://doi.org/10.2196/29120)] [Medline: [34723829](https://pubmed.ncbi.nlm.nih.gov/34723829/)]
68. Steinkamp JM, Chambers C, Lalevic D, Zafar HM, Cook TS. Toward complete structured information extraction from radiology reports using machine learning. *J Digit Imaging.* 2019;32(4):554-564. [FREE Full text] [doi: [10.1007/s10278-019-00234-y](https://doi.org/10.1007/s10278-019-00234-y)] [Medline: [31218554](https://pubmed.ncbi.nlm.nih.gov/31218554/)]
69. Peterson KS, Lewis J, Patterson OV, Chapman AB, Denhalter DW, Lye PA, et al. Automated travel history extraction from clinical notes for informing the detection of emergent infectious disease events: algorithm development and validation. *JMIR Public Health Surveill.* 2021;7(3):e26719. [FREE Full text] [doi: [10.2196/26719](https://doi.org/10.2196/26719)] [Medline: [33759790](https://pubmed.ncbi.nlm.nih.gov/33759790/)]
70. Malmasi S, Ge W, Hosomura N, Turchin A. Comparing information extraction techniques for low-prevalence concepts: The case of insulin rejection by patients. *J Biomed Inform.* 2019;99:103306. [FREE Full text] [doi: [10.1016/j.jbi.2019.103306](https://doi.org/10.1016/j.jbi.2019.103306)] [Medline: [31618679](https://pubmed.ncbi.nlm.nih.gov/31618679/)]
71. Chen Y, Hao L, Zou VZ, Hollander Z, Ng RT, Isaac KV. Automated medical chart review for breast cancer outcomes research: a novel natural language processing extraction system. *BMC Med Res Methodol.* 2022;22(1):136. [FREE Full text] [doi: [10.1186/s12874-022-01583-z](https://doi.org/10.1186/s12874-022-01583-z)] [Medline: [35549854](https://pubmed.ncbi.nlm.nih.gov/35549854/)]
72. Stercx L, Vandewiele G, Dehaene I, Janssens O, Ongenaes F, De Backere F, et al. Clinical information extraction for preterm birth risk prediction. *J Biomed Inform.* 2020;110:103544. [FREE Full text] [doi: [10.1016/j.jbi.2020.103544](https://doi.org/10.1016/j.jbi.2020.103544)] [Medline: [32858168](https://pubmed.ncbi.nlm.nih.gov/32858168/)]
73. Cen X, Yuan J, Pan C, Tang Q, Ma Q. Contextual embedding bootstrapped neural network for medical information extraction of coronary artery disease records. *Med Biol Eng Comput.* 2021;59(5):1111-1121. [doi: [10.1007/s11517-021-02359-1](https://doi.org/10.1007/s11517-021-02359-1)] [Medline: [33893606](https://pubmed.ncbi.nlm.nih.gov/33893606/)]
74. Dandala B, Joopudi V, Tsou CH, Liang JJ, Suryanarayanan P. Extraction of information related to drug safety surveillance from electronic health record notes: joint modeling of entities and relations using knowledge-aware neural attentive models. *JMIR Med Inform.* 2020;8(7):e18417. [FREE Full text] [doi: [10.2196/18417](https://doi.org/10.2196/18417)] [Medline: [32459650](https://pubmed.ncbi.nlm.nih.gov/32459650/)]
75. Zhou J, Guo X, Duan L, Yao Y, Shang Y, Wang Y, et al. Moving toward a standardized diagnostic statement of pituitary adenoma using an information extraction model: a real-world study based on electronic medical records. *BMC Med Inform Decis Mak.* 2022;22(1):319. [FREE Full text] [doi: [10.1186/s12911-022-02031-0](https://doi.org/10.1186/s12911-022-02031-0)] [Medline: [36476365](https://pubmed.ncbi.nlm.nih.gov/36476365/)]
76. Kraljevic Z, Searle T, Shek A, Roguski L, Noor K, Bean D, et al. Multi-domain clinical natural language processing with MedCAT: The medical concept annotation toolkit. *Artif Intell Med.* 2021;117:102083. [doi: [10.1016/j.artmed.2021.102083](https://doi.org/10.1016/j.artmed.2021.102083)] [Medline: [34127232](https://pubmed.ncbi.nlm.nih.gov/34127232/)]
77. Dewaswala N, Chen D, Bhopalwala H, Kaggal VC, Murphy SP, Bos JM, et al. Natural language processing for identification of hypertrophic cardiomyopathy patients from cardiac magnetic resonance reports. *BMC Med Inform Decis Mak.* 2022;22(1):272. [FREE Full text] [doi: [10.1186/s12911-022-02017-y](https://doi.org/10.1186/s12911-022-02017-y)] [Medline: [36258218](https://pubmed.ncbi.nlm.nih.gov/36258218/)]
78. Chen L, Song L, Shao Y, Li D, Ding K. Using natural language processing to extract clinically useful information from chinese electronic medical records. *Int J Med Inform.* 2019;124:6-12. [doi: [10.1016/j.ijmedinf.2019.01.004](https://doi.org/10.1016/j.ijmedinf.2019.01.004)] [Medline: [30784428](https://pubmed.ncbi.nlm.nih.gov/30784428/)]
79. Zhang Q, Wu M, Lv P, Zhang M, Yang H. Research on named entity recognition of chinese electronic medical records based on multi-head attention mechanism and character-word information fusion. *IFS.* 2022;42(4):4105-4116. [doi: [10.3233/jifs-212495](https://doi.org/10.3233/jifs-212495)]
80. spaCy. Language processing pipelines. URL: <https://spacy.io/usage/processing-pipelines/> [accessed 2023-11-01]
81. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. In: Walker M, Ji H, Stent A, editors. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana. Association for Computational Linguistics; 2018:2227-2237.

82. Lee EK, Uppal K. CERC: an interactive content extraction, recognition, and construction tool for clinical and biomedical text. *BMC Med Inform Decis Mak.* 2020;20(Suppl 14):306. [FREE Full text] [doi: [10.1186/s12911-020-01330-8](https://doi.org/10.1186/s12911-020-01330-8)] [Medline: [33323109](https://pubmed.ncbi.nlm.nih.gov/33323109/)]
83. Chen Y-P, Lo Y-H, Lai F, Huang C-H. Disease concept-embedding based on the self-supervised method for medical information extraction from electronic health records and disease retrieval: algorithm development and validation study. *J Med Internet Res.* 2021;23(1):e25113. [FREE Full text] [doi: [10.2196/25113](https://doi.org/10.2196/25113)] [Medline: [33502324](https://pubmed.ncbi.nlm.nih.gov/33502324/)]
84. Sykes D, Grivas A, Grover C, Tobin R, Sudlow C, Whiteley W, et al. Comparison of rule-based and neural network models for negation detection in radiology reports. *Nat Lang Eng.* 2021;27(2):203-224. [doi: [10.1017/s1351324920000509](https://doi.org/10.1017/s1351324920000509)]
85. Chaturvedi J, Chance N, Mirza L, Vernugopan V, Velupillai S, Stewart R, et al. Development of a corpus annotated with mentions of pain in mental health records: natural language processing approach. *JMIR Form Res.* 2023;7:e45849. [FREE Full text] [doi: [10.2196/45849](https://doi.org/10.2196/45849)] [Medline: [37358897](https://pubmed.ncbi.nlm.nih.gov/37358897/)]
86. Newman-Griffis D, Divita G, Desmet B, Zirikly A, Rosé CP, Fosler-Lussier E. Ambiguity in medical concept normalization: an analysis of types and coverage in electronic health record datasets. *J Am Med Inform Assoc.* 2021;28(3):516-532. [FREE Full text] [doi: [10.1093/jamia/ocaa269](https://doi.org/10.1093/jamia/ocaa269)] [Medline: [33319905](https://pubmed.ncbi.nlm.nih.gov/33319905/)]
87. Jaber A, Martínez P. Disambiguating clinical abbreviations using a one-fits-all classifier based on deep learning techniques. *Methods Inf Med.* 2022;61(S 01):e28-e34. [FREE Full text] [doi: [10.1055/s-0042-1742388](https://doi.org/10.1055/s-0042-1742388)] [Medline: [35104909](https://pubmed.ncbi.nlm.nih.gov/35104909/)]
88. 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. [FREE Full text] [doi: [10.1186/s12859-022-05035-9](https://doi.org/10.1186/s12859-022-05035-9)] [Medline: [36384464](https://pubmed.ncbi.nlm.nih.gov/36384464/)]

Abbreviations

AI: artificial intelligence

ANN: artificial neural network

BERT: Bidirectional Encoder Representations from Transformers

BiLSTM-CRF: Bidirectional Long Short-Term Memory - Conditional Random Field

CNN: convolutional neural network

DL: deep learning

EHR: electronic health record

EMR: electronic medical record

ELMo: embeddings from language models

ICD-10: International Statistical Classification of Diseases, Tenth Revision

IE: information extraction

LLM: large language model

LM: language model

ML: machine learning

NER: named entity recognition

NLP: natural language processing

PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

PT-BR: Brazilian Portuguese

PT-PT: European Portuguese

RNN: recurrent neural network

SLR: systematic literature review

WOSCC: Web of Science Core Collection

Edited by A Castonguay; submitted 03.05.24; peer-reviewed by A Kocian, DR Carvalho; comments to author 13.06.24; revised version received 14.07.24; accepted 06.08.24; published 21.10.24

Please cite as:

Nunes M, Bone J, Ferreira JC, Elvas LB

Health Care Language Models and Their Fine-Tuning for Information Extraction: Scoping Review

JMIR Med Inform 2024;12:e60164

URL: <https://medinform.jmir.org/2024/1/e60164>

doi: [10.2196/60164](https://doi.org/10.2196/60164)

PMID: [39432345](https://pubmed.ncbi.nlm.nih.gov/39432345/)

©Miguel Nunes, Joao Bone, Joao C Ferreira, Luis B Elvas. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 21.10.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.