

Exploring the Integration of Ontologies and Medical Records through Named Entity Recognition and Relation Extraction

Olivia Bennett^{1*}, Harry Collins², Jack Foster¹, Ethan Wright³

¹ Department of Medical Informatics and Natural Language Processing, Faculty of Medicine, University of Nottingham, Nottingham, United Kingdom.

² Department of Ontology Integration and Clinical Data Science, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

³ Department of Named Entity Recognition and Relation Extraction in Healthcare, Faculty of Medicine, Newcastle University, Newcastle, United Kingdom.

*E-mail ✉ olivia.bennett@nottingham.ac.uk

Abstract

Unstructured natural language information contained in electronic health records holds considerable value for health research and innovation. However, manually reviewing these records is impractical and hinders access to the insights they contain. This study introduces a method to automatically extract information from unstructured clinical documents through Domain Entity Recognition and Relation Extraction, then organize the outputs using a domain-specific ontology. The work focuses on oncology, a field that demands significant attention. The primary contribution is a comprehensive methodology that combines multiple resources to achieve this goal. We created a novel annotated dataset of medical evolution notes in Brazilian Portuguese, featuring 1622 documents, 146,769 entities, and 111,716 relations. In the exams domain, the approach reached 78.24% accuracy for entity and relation extraction. Healthcare experts reviewed the method for entity recognition and relation extraction and viewed it favorably, deeming the overall methodology useful for medical practitioners.

Keywords: Named entity recognition, Deep learning, Electronic health record, Relation extraction, Natural language processing, Ontology

Introduction

Medical records generally document health events through natural-language statements. These records are increasingly moving to digital formats [1], creating extensive collections of important health data. Information inside Electronic Health Records (EHRs), including clinical reports and medical observations, supports various purposes such as disease registries, epidemiological research, drug safety monitoring, clinical trials, and health audits [2, 3]. Methods that rely

on clinical experts to manually extract medical details from patient files face clear constraints in scalability, time requirements, and cost [3-5]. Given this situation, the potential for reusing EHR data has created a need for automated processing of free-text narratives. Such automation opens new opportunities to assist healthcare providers [6, 7].

Named Entity Recognition (NER) and Relation Extraction (RE) methods have become effective tools for automatically retrieving meaningful details from unstructured health information. NER identifies key medical entities, for example, patient identifiers, disease names, and medication names. RE, in turn, detects and extracts connections between these entities, such as links between patients and diagnoses or between medications and dosages. Both approaches are increasingly used in healthcare to improve information access, clinical decision support, and patient management [8-12].

Access this article online

<https://smerpub.com/>

Received: 26 January 2026; Accepted: 02 May 2026

Copyright CC BY-NC-SA 4.0

How to cite this article: Bennett O, Collins H, Foster J, Wright E. Exploring the Integration of Ontologies and Medical Records through Named Entity Recognition and Relation Extraction. J Med Sci Interdiscip Res. 2026;6(1):172-89. <https://doi.org/10.51847/Hu5VoSiaNG>

Various techniques can aid NER and RE, including BERT or Bi-LSTM models. BERT architectures have delivered strong performance in NLP tasks, especially NER within specialized domains like healthcare [9, 13-15]. Bi-LSTM models use both forward and backward context to identify intricate patterns during relation extraction [16, 17]. While BERT and Bi-LSTM are commonly used, their application in Portuguese-language contexts still needs further investigation and faces challenges due to the scarcity of suitable datasets. Although established information extraction techniques exist, few methodologies effectively combine complementary methods. Therefore, we designed an approach to incorporate natural language data from oncology records into a domain ontology. The method employs Deep Learning, using Transformer models such as BERT along with a Bi-LSTM model, to perform named entity recognition and relation extraction. The extracted information is then mapped into an ontology. This methodology offers several advancements to the field. First, it includes the creation of previously unavailable datasets annotated with entities and relations from medical evolution notes, encompassing 1622 documents, 146,769 entities, and 111,716 relations. Second, it establishes an ontology framework tailored to the oncology domain for organizing related medical information. Third, it presents a strategy for combining NER and RE resources with the domain ontology. A notable aspect is the use of genuine oncology data in Portuguese for training. These outcomes create opportunities for future tools that can benefit medicine by supplying a structured, dependable knowledge foundation for clinicians and investigators. The rest of the paper is structured as follows. Section 2 reviews related work. Section 3 outlines the research setting. Section 4 details the development of Portuguese-language datasets derived from authentic sources. Section 5 covers the experiments. Section 6 presents and analyzes the experimental findings. Section 7 provides conclusions and directions for future research.

Related works

We performed a systematic literature review to gain a thorough perspective on existing studies. From an initial pool of 961 publications, we filtered and examined 23 relevant papers addressing Entity Recognition, Relation Extraction, and the integration of medical information. The review protocol merged guidelines from computing research [18] with those from the health domain [19].

This combined approach improved the reliability of the findings relative to standard literature searches [20].

The analysis revealed extensive use of BERT-based models for NER and RE in healthcare because of their strong performance [9, 14]. A major benefit of BERT is its ability to be fine-tuned on targeted datasets, which is especially helpful when annotated data is scarce [21, 22]. Transformer-based techniques have also shown utility in mining biomedical data [23]. While BERT is adaptable across many NLP applications, specialized variants such as BioBERT [21] achieve outstanding outcomes in biomedical settings. This underscores the importance of transformer models in NER and RE studies [24]. Model design can benefit from expanded contextual understanding, as demonstrated by MedBERT [25], a contextualized embedding model trained on the MedNLI medical natural language inference dataset. MedBERT improves NER and RE by better capturing medical context, resulting in notable accuracy gains. Moreover, domain-adapted contextual embeddings address the particular characteristics of medical terminology and serve as valuable assets for information extraction from EHRs [26].

Despite identifying valuable contributions involving BERT, the review uncovered several gaps. One concerns limited support for Portuguese, where resources remain scarce. Another concerns the nature of the data used, as many studies rely on non-health data or information not drawn from actual clinical environments. A third gap involves integrating extracted data using formal structures like ontologies, which could improve extraction quality by providing more organized and semantically rich representations of entities and their relationships.

Research context

This section outlines the core components of the proposed methodology for incorporating natural language data into a domain ontology focused on oncology. The study was conducted in collaboration with a software company operating the Gemed Oncology System, an EHR platform for oncology. The company serves clinics and hospitals by supplying tools for the electronic documentation of medical, nursing, and pharmaceutical care information. Consequently, the research constitutes a case study in oncology supported by real-world data and systems from this partner.

Figure 1 provides an overview of the research context, encompassing medical clinics or hospitals and the

requirements of healthcare staff. The methodology incorporates several essential aspects: medical notes produced by professionals within an EHR system, natural language processing of these notes to automatically perform named entity and relation extraction, and the

subsequent population of a domain ontology with the extracted information. The central aim of this strategy is to help healthcare professionals use their clinical documentation more effectively.

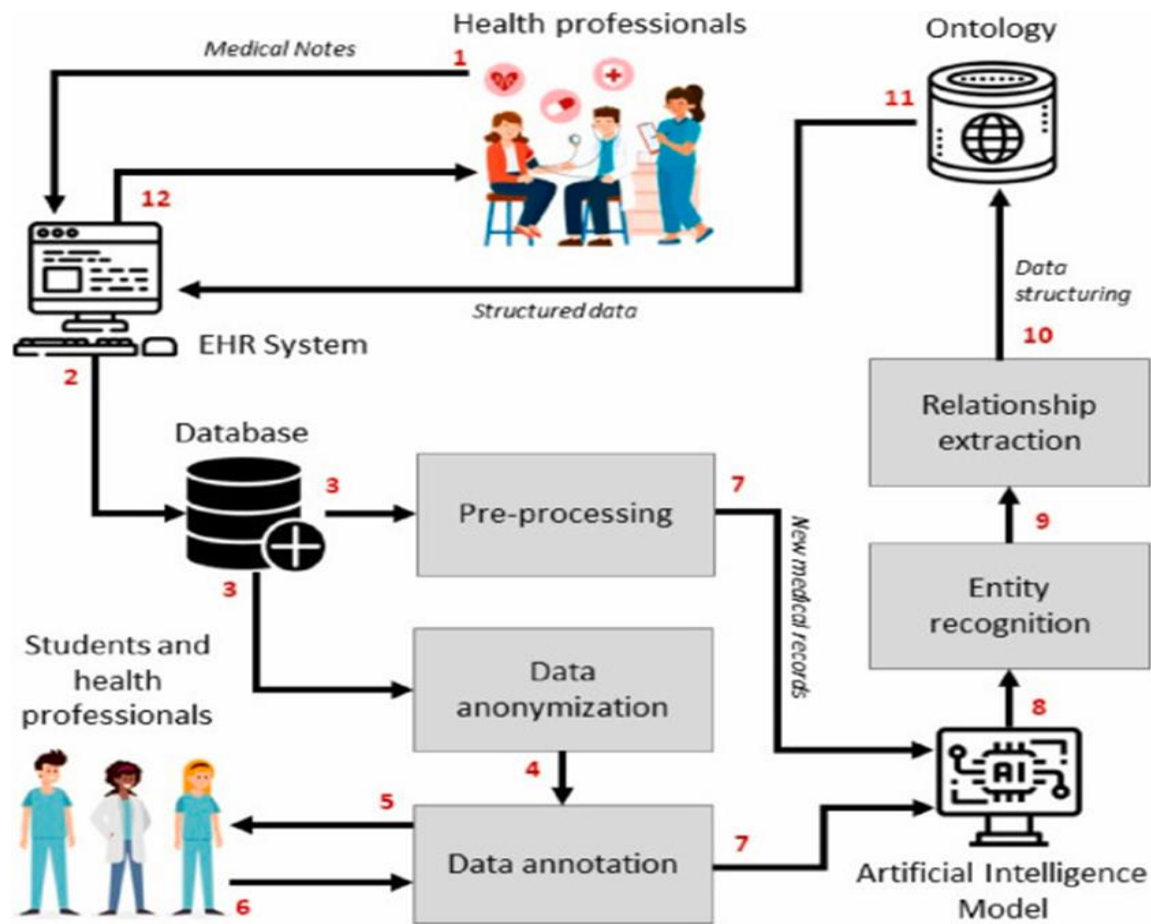


Figure 1. Overview of the research context.

The initial phase [1] of the methodology focuses on healthcare professionals' activities as they document patient progress notes and the outcomes of those activities. Essential patient-care details are entered into the GEMED EHR system as clinical notes that combine free-text entries with structured information. These notes capture clinical events recorded by medical staff. Professionals utilize a flexible form within the system to log each clinical event. The system then saves the information in a database during the next phase [2]. The system retains comprehensive patient health details, including medical notes, electronic prescriptions, and consultation histories.

The following phase of the methodology [3] consists of automated procedures to prepare the data for named

entity recognition and relation extraction. Key activities include data anonymization, which ensures that identifying information is detached from the individuals involved. Additional general preprocessing steps are also applied. This study implements anonymization to prevent the exposure or unique identification of patients and healthcare providers. After these steps, the data becomes suitable for annotation and dataset creation. During the annotation phase [4], the project received assistance from experts at Interprocess GEMED and Universidade do Vale do Rio dos Sinos [5]. Annotated data were pulled from the GEMED database to label clinical information according to predefined categories [6]. The resulting annotated collection was named Dataset Gemed Onco (DGO).

In the next stage, automated information extraction takes place [7]. The proposed solution integrates two models: BERT and Bi-LSTM. BERT was trained for entity recognition [8] and feature extraction [9] to identify relevant elements in the text. Meanwhile, Bi-LSTM classifies relationships between the identified entities.

After extraction, the methodology moves to data representation by organizing the extracted details within a domain ontology [10]. An oncology-specific ontology was selected and developed to model the entities, concepts, and particular relationships uncovered during extraction. Once structured, the EHR system can query the ontology [11] and leverage its contents for various purposes, primarily data retrieval, but also to assist in diagnosis or event forecasting [12].

We established an evaluation strategy to assess the model from both healthcare and computational perspectives. In the healthcare domain, oncology specialists reviewed the model and its outcomes in practical case studies. Data were gathered through questionnaires and ongoing interactions with professionals to gauge their views on the model's applicability. On the computational side, standard performance metrics were applied, including Accuracy, Recall, F1-Score, macro-average, and weighted-average.

Developed datasets

Two datasets were created. The first addresses relations in the exams domain. Professionals from Interprocess and final-semester medical students from Universidade do Vale do Rio dos Sinos annotated it. The second dataset targets Diagnostic Features of breast cancer and was annotated by healthcare specialists. Approval from the research ethics committee was not necessary for this study, as the original data underwent full anonymization throughout the process, eliminating any links to specific patients and preventing individual identification.

For the first dataset, referred to as Gemed Onco - Exams or DGO-E, we examined 1022 documents detailing patient evolutions from an oncology clinic. These documents reflect authentic information recorded by healthcare professionals. Before use, all documents were anonymized by removing names and identifiers linked to patients or staff. Discussions with oncology experts helped determine the dataset's focus. The decision was

made to emphasize exam-related data, given its identifiability within the text and its importance for a company prototype.

The entities of interest defined for annotation in the DGO-E dataset include:

- **Exame (Exam):** Refers to the name of the examination. Example: PSA;
- **Resultado (Result):** Denotes the numerical or descriptive outcomes of an exam. Examples: 193mil (193 thousand), 1547;
- **Data (Date):** Specifies the date when an exam occurred. Examples: 07/18/2022 (2022/18/07), 07/2022 (2022/07);
- **Membro (Member):** Identifies a part of the human body where the exam was conducted. Examples: Crânio (Skull), lombar (lumbar);
- **Tempo (Time):** Indicates the timeframe of an action, such as how long ago an exam took place. Examples: 1 mês atrás (1 month ago), no último ano (last year).

The relations of interest established for annotation were:

- **Resultado (Result):** Connects the outcome value to the specific type of exam;
- **Quando (When):** Links the date and time information to the corresponding exam;
- **Localização (Location):** Associates a body part with the exam.

Once the parameters were established, the annotation process for the dataset commenced. Three scholarship students nearing completion of their medical studies at Unisinos, along with a nurse and a biomedical doctor from Empresa Interprocess, carried out the annotations. UBIAI Tools [1] software was employed for this purpose. The tool supports collaborative annotation and offers a user-friendly interface. It allows for manual labeling of entities, relations, and categories, and includes capabilities such as Entity Dictionary, Rule-Based Matching, and Relation Dictionary for assisted annotations. Manual annotation was selected for this research to ensure the highest possible data quality.

Figure 2 shows an example annotation and highlights the software interface.

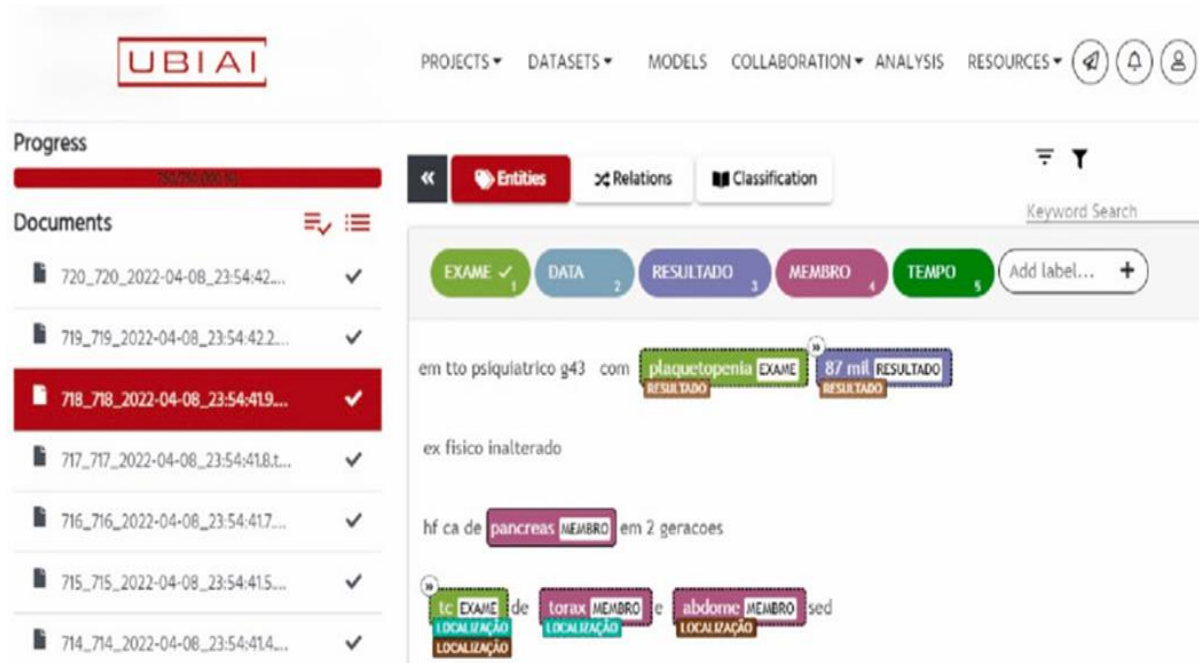


Figure 2. UBAI software with an annotated document.

In developing the second collection, known as Dataset Gemed Onco - Diagnostic Features (DGO-DF), we partnered with specialists in health data labeling and medical annotation. After an initial review, the focus shifted to compiling information specifically for Breast Cancer diagnostics to contribute meaningful resources to this vital field of medicine. The annotated elements in this resource consist of:

- **Característica do diagnóstico (Diagnosis Features):** Highlights important details and characteristics tied to an individual cancer or oncology situation.

Example: Ca de cólon Ell alto risco (Ell colon cancer high risk).

- **Órgão (Organ):** Specifies the body part impacted by the malignancy.

Example: Paciente notou nódulo em mama Esquerda (Patient noticed a lump in the left breast).

- **Local do Órgão (Organ Location):** Details the exact position within the affected body part.

Example: Identificou um nódulo em mama Direita (Identified a breast lump on the right).

- **Estadiamento (Staging):** Captures the stage or classification level of the cancer.

Example: Ca de ovário EIV resistente a platina (Platinum resistant ovarian cancer EIV).

- **Tempo (Time):** Records how much time has passed since symptoms first appeared or were identified.

Example: Paciente notou há 6 meses nódulo na mama E com crescimento moderado (Patient noted 6 months ago a moderately growing breast E lump).

Given the background of the participating medical experts, the effort here was limited to entity labeling alone. A total of 600 clinical evolution notes were obtained from an oncology center, all pre-selected for Breast Cancer diagnoses. Every document underwent anonymization procedures to safeguard patient identities and fully meet ethical requirements.

The annotation work was carried out with the MAE Annotation Tool [2]. This Multi-document Annotation Environment (MAE) provides a sophisticated interface for tagging textual entities and excels at processing multiple files simultaneously.

Such expert-created Portuguese language datasets are crucial for building deep learning systems adapted to Portuguese, supporting advanced natural language processing needs. Involving domain professionals in manual labeling delivers superior accuracy, which strengthens the resulting models, improves outcome reliability, and accelerates advances in medical applications.

To further ensure quality, all 1022 annotated files in DGO-E underwent thorough review and confirmation by experienced oncology professionals. These reviewers achieved full agreement regarding the precision and uniformity of the marked entities and connections

throughout the materials. This verification is essential to confirm the dataset's reliability for future studies. The graph in **Figure 3** provides a visual summary of the distribution, using colored bars to indicate entity counts across different categories. Resultado (result) and Exame (Exams) stand out as the most populated groups, whereas Membro (member), Data (date), and especially Tempo (time) show markedly lower numbers. This breakdown reflects the common patterns seen in everyday clinical documentation from the project's source materials.

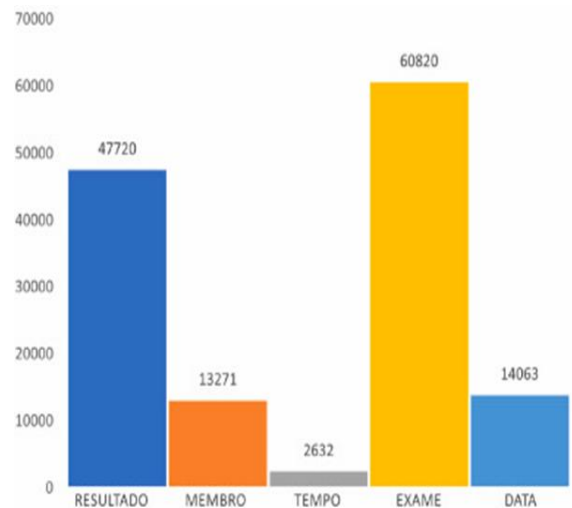


Figure 3. DGO-E examination entities comparison.

This collection contains a large volume of labeled entities, with the categories Exame and Resultado standing out due to their high annotation counts (60,820 and 47,720, respectively). Such figures demonstrate that the resource effectively captures a wide range of associated details commonly found in electronic health records (EHRs). By contrast, the Membro and Data categories received comparatively modest annotation totals (13,271 and 14,063). Despite their lower frequency, these labels remain clinically meaningful, offering valuable insights into involved body regions and timelines connected to medical tests. The notably limited count of Tempo annotations (2,632) presents certain difficulties for training models, yet it accurately mirrors the natural distribution present in the source data.

Regarding relation extraction (RE) within oncological progress notes, the aim was to detect meaningful connections among the labeled entities. Typical examples include linking an Exame to its corresponding Resultado, or associating a Data with a particular Exame.

The graph in **Figure 4** provides a comparative overview of these relations.

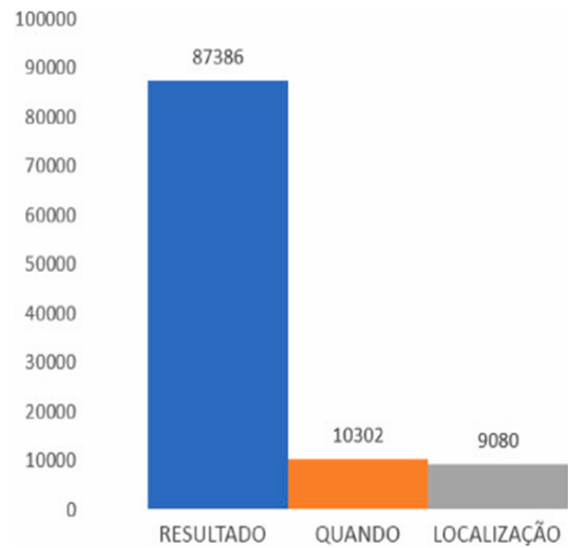


Figure 4. Comparison of DGO-E exam lists.

Regarding the relations labeled in the dataset, the Resultado category stood out with a very high number of annotations (87,386). In comparison, the Quando and Localização relations appeared far less frequently (10,302 and 9,080, respectively). These details proved particularly useful for better understanding and evaluating examination outcomes.

For Named Entity Recognition (NER) tasks, the labeling process relies on the widely adopted IOB format. This approach provides a uniform, well-defined method for indicating the start and end points of entities within text. It helps ensure precise parsing, avoids overlaps or omissions of entity components, and guarantees that every pertinent segment is properly captured.

For DGO-DF, once annotation was complete, oncology specialists carefully examined and confirmed the full set of 600 labeled documents. This verification process was critical to upholding the overall quality and uniformity of the annotations, confirming that key details were accurately recognized and tagged. Finally, the complete dataset was exported in IOB format. **Figure 5** shows a side-by-side comparison of the entities.

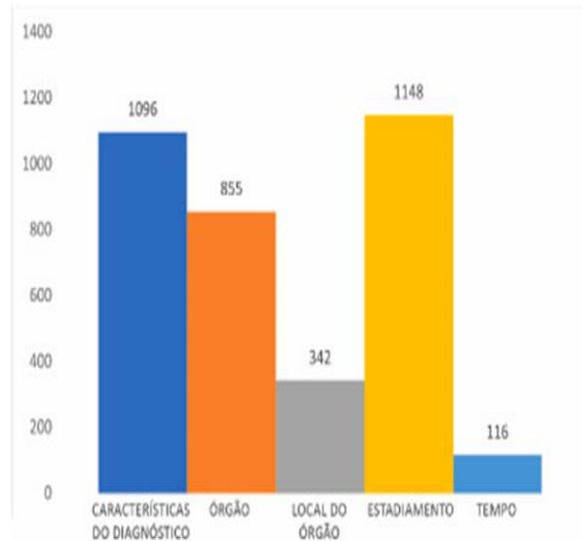


Figure 5. Comparison of DGO-DF entities.

The dataset features a substantial number of labeled entities, with *Estadiamento* being particularly prominent at 1,148 instances. This category is especially important for Breast Cancer assessment, providing critical details on disease stage and progression. Other notable entities include *Característica do diagnóstico*, appearing 1,096 times, and *Órgão*, with 855 occurrences. Together, these provide insight into the condition's specific traits and the involved body part.

In contrast, the *Local do órgão* entity appeared less frequently, recorded in 342 cases. This may indicate that precise positional details regarding the cancer-affected organ receive less emphasis in the reviewed medical documentation. Still, its relevance can vary depending on the clinical scenario.

The *Tempo* entity appeared 116 times, conveying details about the timeframe linked to Breast Cancer identification. Although its count is lower than others, temporal aspects remain a key consideration in Breast Cancer studies, providing essential data that can influence patient treatment strategies and outcome predictions.

Experiments

This section outlines the experiments conducted and their value across two primary dimensions: developing NER and RE models, and creating and applying an ontology to support data integration.

Named entity recognition

For the initial experiment, we trained a BERT model to handle NER tasks and generate text feature representations. The NER setup relied on the BertForTokenClassification model [3], as described in Ref. [27]. This architecture benefits from broad library compatibility, with a Python tokenizer and support for Jax, PyTorch, and TensorFlow.

Multiple BERT variants were tested during this phase. The top-performing configuration from the entity recognition trials was later selected to support subsequent relation classification using BiLSTM. The specific models evaluated included BERT-base-multilingual-cased [28], BioBERT-PT [28, 29], and BERTimbau [30, 31].

Model construction for NER began with the spaCy [4] library in Python. spaCy is a comprehensive NLP toolkit that supports sophisticated text parsing and information extraction. It was used for data preparation and preprocessing throughout the training cycles. Two separate training runs were executed, each spanning 50 epochs, enabling the model to recognize entities within the corpus progressively. The core architecture chosen was BERT, a Transformer-based design. Training occurred in a supervised setting using texts paired with their corresponding entity labels.

The first training round utilized 716 labeled samples, representing 70% of the available NER dataset. Data was loaded from JSON files containing the raw texts and class labels, then structured into Pandas dataframes—a tabular format with rows and columns—for easier handling and manipulation.

Before training, a data-cleaning phase identified and addressed outliers or corrupted entries, though none were detected in this case. Tokenization followed, breaking the texts into smaller units such as words or subwords. The BERT tokenizer was then applied to convert inputs into token sequences, incorporating special markers for sequence start and end. All token sequences were padded or truncated to a uniform maximum length.

Because of BERT's input constraints, the processing logic was adapted to handle text in overlapping windows of 128 tokens, assigning each token's classification based on the window where it was closest to the center. In practice, the maximum sequence length was configured to 512 tokens with a stride of 128. For instance, a token at position 400 would be evaluated across multiple windows, and its final output would be taken from the iteration in which it sat closest to the center.

Training ran for 50 epochs using a batch size of 16. Optimization used the AdamW algorithm [32] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, a weight decay of 0.1, and an initial learning rate of $1e-5$, with linear scheduling and a 100-step warmup phase.

Fine-tuning adapted the pre-trained BERT model to the specific NER task using the DGO-E dataset and its labeled examples. The original classification head was substituted with a new layer tailored for entity tagging. This process adjusted the model weights through optimization to reduce the loss. Data shuffling was implemented to mitigate overfitting and improve generalization to unseen examples. Entities were handled jointly so the model could learn comprehensive recognition patterns. Supporting libraries such as re, random, numpy, pandas, matplotlib, sklearn, TensorFlow, and Keras were integrated to streamline development and training activities.

Model performance was assessed using the validation set outlined in Section 6.1. The resulting model supports inference on new documents, generating predictions alongside contextual embeddings derived from the [CLS] token representation.

An additional training run was performed on the unmodified DGO-DF dataset. Since DGO-DF includes only entity annotations, the assessment concentrated solely on the NER component.

Relation extraction

In the relation extraction experiment, a BiLSTM-based method was used to detect entity pairs and classify the connections between them. The classification process considers the entity types present in the text along with the specific entities chosen for relation analysis. The BiLSTM model processes each pair, trio, or larger group of entities in separate runs. Special tokens were inserted to mark the beginning and end of entities, enabling the BERT model to identify them clearly.

This strategy was realized using the PyTorch library through the Hugging Face API, with a token-level classifier layered on top of the BERT models. Training data came primarily from the DGO-E dataset, building on the prior NER work. Data preparation involved converting the inputs into a suitable structure, such as data tables, to encode relationships among entities in the text properly. Words were then transformed into numerical vectors via Word2Vec and GloVe embedding techniques. These vector representations encode

semantic meaning, helping the model capture associations between terms.

The BiLSTM architecture was constructed using layers of LSTM units. Its bidirectional design includes one layer that scans the sequence forward and another that processes it backward. This setup allows the model to gather contextual cues from both directions for each word, leading to a richer understanding of the surrounding text. A loss function was specified to quantify the difference between the relations predicted by the BiLSTM and the ground-truth relations in the dataset. Weights for the BiLSTM were randomly initialized near zero following Xavier initialization guidelines [33] to avoid issues like vanishing or exploding gradients. Weight updates occurred through the backpropagation algorithm [34], which distributes loss gradients across the network. Training consisted of multiple epochs, with each epoch iterating over the training data and adjusting weights according to computed gradients, progressively refining the model's ability to recognize entity relationships.

The Adam optimizer [32] guided weight adjustments, incorporating hyperparameters such as learning rate, learning rate decay, momentum (β_1), and momentum decay (β_2). The learning rate determines the step size for weight changes, while decay gradually lowers it over time. Momentum influences the impact of past gradients, and its decay progressively reduces that influence. The chosen parameters were $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e-7$, with an initial learning rate of $1e-3$ and a decay of $1e-5$. Experiments identified optimal values of $S = 128$ and $P = 64$, which delivered the strongest results. The BiLSTM was trained for 75 epochs.

To regularize the BiLSTM and curb overfitting, L1 and L2 regularization methods were added to penalize excessive weight magnitudes in the loss function. Additionally, Dropout was implemented by randomly deactivating units during training to limit unit co-adaptation and enhance generalization. Data was processed in batches (mini-batches) to improve computational efficiency, enabling parallel operations and better utilization of hardware resources.

Information integration with ontology

Ontology development followed the methodology outlined in Ref. [35]. The process began with an analysis of 3,309 documents comprising medical consultations and notes from healthcare professionals. These materials contained 212,829 words across 28,178 lines. This

examination helped pinpoint domain-specific terms and information sets that could be automatically extracted. Clinicians often need these supplementary details during routine practice—for instance, descriptions of symptom perception or medication use.

Consultations with healthcare and technology experts yielded a collection of important phrases and concepts that formed the foundation for the work. Interviews were conducted to determine key questions and establish a solid basis for domain modeling. Manual text analysis was then used to map this information to core concepts and relationships in the knowledge base. This careful approach ensured the resulting knowledge base was precise and dependable, providing a strong platform for future automated information extraction efforts.

The ontology was built using Protégé [5] software and includes 181 classes, 14 data properties, and 12 object properties. It covers oncology-related classes such as

diagnosis, course, test, drug, patient, procedure, protocol, symptom, and organ. The structure also defines relationships between entities—for example, links between patients and symptoms, patients and drugs, drugs and measurement units, exams or procedures and body locations, and patients and diagnoses or exams. Data properties capture specific attributes like exam identifiers, exam characteristics, treatment cycles, event dates, cancer staging, exam types, symptom intensity, body locations, patient identifiers, periods, medication dosage, and test results.

Overall, the ontology supplies a structured framework for representing and connecting oncology-related data, facilitating improved interpretation and analysis of clinical information. As an illustration, **Figure 6** highlights central concepts around Edema along with their primary relationships and selected instances from the ontology.

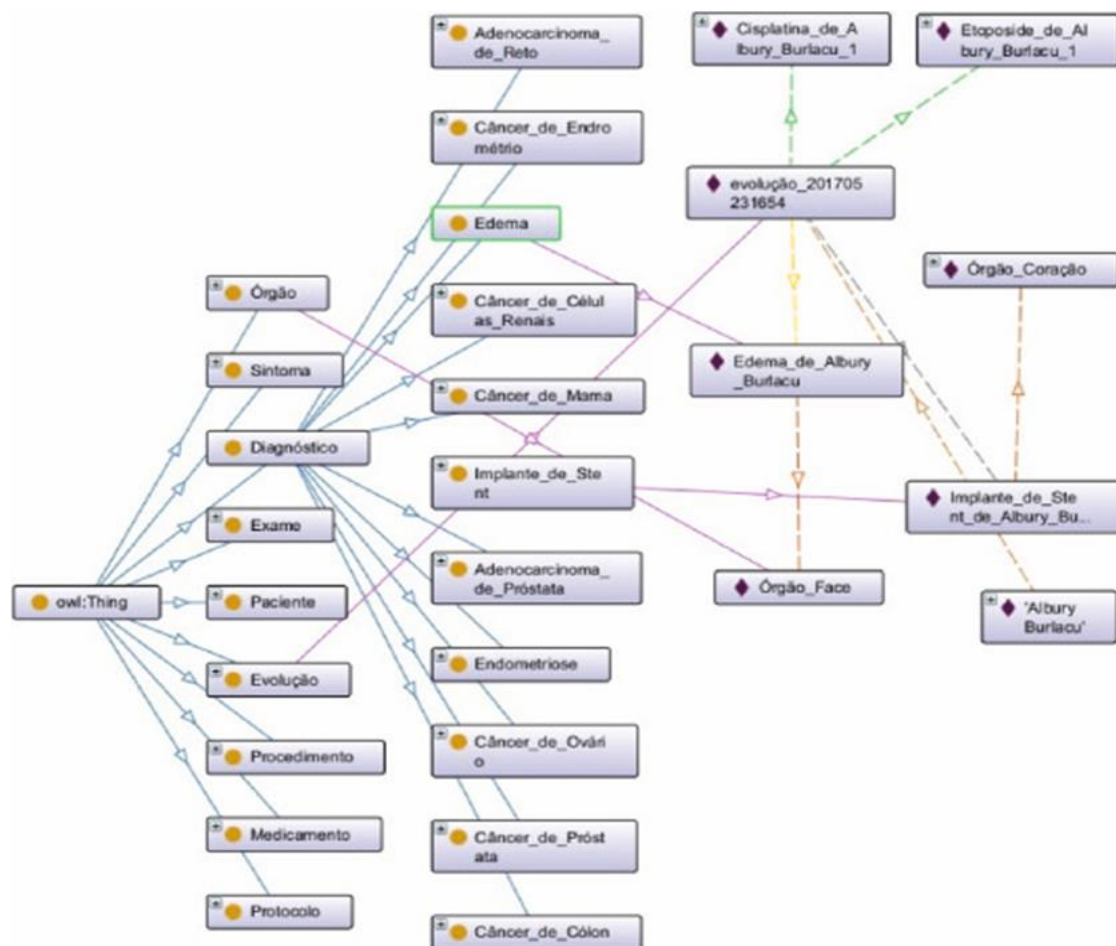


Figure 6. Core concepts of edema.

The edema concept is classified as a type of Diagnóstico and is associated with the Órgão Face. The ontology captures details about the progression of this concept through specific instances.

The ontology’s formally organized structure supports integrating unstructured clinical information, helping healthcare providers interpret and examine the circumstances of cancer patients. Thanks to its clearly defined relationships, the ontology can underpin the creation of medical decision support tools, ultimately helping enhance treatment and follow-up care for individuals with cancer.

Integration with the ontology relied on custom libraries developed as part of this project. These tools enable recording and managing data within the ontology, ensuring that extracted entities and relations are stored correctly. Tests for automatic insertion were conducted successfully.

The model leverages results from the DGO-E experiment to identify entities and relations linked to examinations. Once extracted, this information becomes available for entry into the ontology. By default, a new ontology instance is generated containing evolution details such as the patient’s GEMED ID (omitted here due to anonymization), the evolution date, and exams connected via the “Present” relation. The identified exams and their associated links are then placed into this new instance. Instance names are typically formed by combining the exam or evolution name, date, time, and patient GEMED ID (currently disregarded because of anonymization). Integration testing with DGO-DF was postponed for future efforts, as that dataset lacks annotated relations.

The automated ontology integration module was implemented in Python 3 with the Owlready2 library. Owlready2 facilitates ontology-oriented programming in Python by offering direct access to OWL ontologies. It features an efficient quad store based on SQLite3—an embedded relational database—optimized for RDF data storage and querying, which delivers strong performance and low memory usage even with sizable ontologies. The library also connects with UMLS and other medical terminologies via its PyMedTermino2 submodule.

Additionally, the algorithm incorporates the HermiT reasoner [36], a dedicated logical inference engine for OWL ontologies. HermiT applies description logic to conclude concepts and relations, supporting tasks such as classifying instances, inferring properties, detecting

inconsistencies, and checking equivalences while adhering to OWL standards.

By linking the model to the ontology, ontological knowledge deepens the model’s analyses and inferences. The combination of a well-structured ontology and HermiT’s automated reasoning capabilities promotes semantic understanding of the represented knowledge. It allows coherent, organized handling of the data produced by the model.

Once model outputs are generated, the OWL ontology is loaded. The outputs are then mapped to the corresponding concepts and properties, involving the creation of instances and the establishment of relationships between them. Owlready2 enables navigation and querying of the ontology’s elements, making it straightforward to insert the model-generated data into the predefined structured format.

Results and Discussion

We evaluated outcomes from two primary perspectives: computational and healthcare-related. Computational assessment covered the NER and RE experiments as well as ontology construction and integration. Meanwhile, usability reviews and result comparisons involved collaboration with medical professionals.

Evaluation of NER and RE experiments

This section examines the effectiveness of the NER and RE models when applied to medical texts. Performance was measured on the DGO-E and DGO-DF datasets using standard metrics: Accuracy, Recall, F1 Score, Support, Macro Average, and Weighted Average. To assess DGO-E NER outcomes, 305 annotated evolution notes (30% of the dataset) served as the test set. Results for Accuracy, Recall, F1 Score, Support, Macro Mean, and Weighted Mean are presented in **Table 1**. The entities Exame, Órgão, Resultado, Data, and Tempo achieved accuracy scores between 70.01% and 78.92%. Their recall values ranged from 70.38% to 77.27%, and F1-scores ranged from 70.59% to 78.77%. Support numbers, indicating the count of instances per entity, varied from 3,311 to 14,648. Overall model accuracy stood at 78.24%. These figures show that the model delivered solid results in identifying and categorizing the targeted entities.

Table 1. NER performance of the DGO-E.

Empty Cell	Accuracy	Recall	F1-score	Support
Exame	78.92 %	77.27 %	78.77 %	14648

Membro	78.26 %	77.74 %	78.25 %	3311
Resultado	70.01 %	70.38 %	70.59 %	11504
Data	77.76 %	75.14 %	76.33 %	3501
Tempo	73.33 %	72.17 %	73.56 %	592
Accuracy			78.24 %	35994
Macro avg	72.78 %	76.01 %	75.10 %	35994
Weighted avg	78.32 %	78.76 %	78.66 %	35994

During training, the BiLSTM model was evaluated on a held-out validation set comprising 30% of the DGO-E data not used in training. The model's relation predictions were compared against the ground-truth

labels. Cross-entropy loss was applied to quantify the difference between the predicted outputs and the actual relation labels. **Table 2** shows the corresponding outcomes.

Table 2. DGO-E relation recognition performance.

Empty Cell	Accuracy	Recall	F1-score	Support
Resultado	76.61 %	75.87 %	75.24 %	20983
Quando	73.45 %	73.55 %	74.12 %	2100
Localização	76.75 %	75.01 %	75.28 %	2501
Accuracy			76.17 %	25584
Macro avg	76.78 %	76.00 %	75.10 %	25584
Weighted avg	76.33 %	78.17 %	78.22 %	25584

Table 2 shows the outcomes for relation recognition on the DGO-E dataset. The Resultado, Quando, and Localização relations achieved accuracy levels ranging from 73.45% to 76.75%, recall values between 73.55% and 75.87%, and F1-scores from 74.12% to 75.28%. Support counts for these relations varied from 2,100 to 20,983. Overall model accuracy reached 76.17%. These

findings demonstrate that the model attained performance levels consistent with existing literature for identifying and categorizing relations within this collection.

For further assessment, we trained the model on the unmodified DGO-DF dataset. **Table 3** displays the NER performance results obtained for DGO-DF.

Table 3. DGO-DF NER performance.

Empty Cell	Accuracy	Recall	F1-score	Support
Característica do Diagnóstico	70.22 %	70.37 %	71.24 %	230
Órgão	73.71 %	72.74 %	72.98 %	189
Local do Órgão	70.12 %	70.17 %	70.99 %	71
Estadiamento	71.73 %	70.13 %	71.00 %	246
Tempo	71.70 %	71.19 %	71.46 %	24
Accuracy			72.87 %	760
Macro avg	72.78 %	72.00 %	72.10 %	760
Weighted avg	72.33 %	72.17 %	72.33 %	760

Table 3 shows the NER outcomes for the DGO-DF dataset. The entities Característica do Diagnóstico,

Órgão, Local do órgão, Estadiamento, and Tempo obtained accuracy rates ranging from 70.12% to 73.71%,

recall values between 70.17% and 72.74%, and F1-scores from 70.99% to 72.98%. Support counts for these entities varied from 24 to 230. The model's overall accuracy stood at 72.87%. These outcomes suggest that the model achieved performance levels aligned with those reported in the relevant literature for recognizing and classifying entities in the DGO-DF collection.

Taken together, the results are considered positive. An examination of the metrics reveals comparable performance across categories, with Accuracy, Recall, and F1-score values clustered closely together. Most categories surpassed the 70% threshold across these measures, demonstrating the model's ability to detect both entities and relations reliably.

Broadly speaking, the BERT model exhibited strong capabilities in NER and RE tasks across the two domains under study. These findings should be interpreted in light of the unique traits of the datasets involved. Within the medical field, NER and RE tasks are inherently difficult given the intricate and varied nature of clinical documentation. In this light, the achieved results can be viewed as successful relative to the complexity of the challenge.

These outcomes are tied to the datasets and tasks at hand. They cannot be directly compared to other studies because the DGO-E and DGO-DF collections have distinctive features and focus on proprietary healthcare and patient information. Nevertheless, even without straightforward comparability, the results align well with the state of the art, as evidenced by Ref. [37], which reported 64% accuracy on Portuguese psychiatry records for NER, and Ref. [38], which attained 70% accuracy in relation extraction from Portuguese clinical progress notes containing pharmacogenomic information.

In contrast, many models in the literature rely on large general-purpose pre-training corpora such as Wikipedia or synthetic data. In contrast, specialized datasets

manually annotated by healthcare experts provide rich domain-specific context. This constitutes a meaningful contribution to the field. While direct comparisons remain limited for the reasons noted, the findings help advance research in the area and can serve as a benchmark for upcoming studies. Additionally, model performance may be affected by the volume of available training data; more annotated examples could further improve results by providing more opportunities for learning and generalization.

Ontology integration

This section details the evaluation process carried out for the DGO-E experiment in conjunction with the ontology. The assessment took place within a practical scenario involving healthcare professionals, specifically biomedical experts, specialized nurses, and medical students. Following appropriate briefing, two distinct evaluations were conducted. The first measured the model's accuracy relative to professionals' judgments. The second explored specialist acceptance of integrating unstructured data into an ontology; further details are provided in section 6.2.1.

For the integration assessment, 22 evolution notes were randomly selected and exported from the SGO. A healthcare professional reviewed the texts and manually annotated the entities in each document. These same notes were subsequently processed by the model, enabling a direct comparison between the expert annotations and the model's performance.

Figure 7 presents a side-by-side comparison for evolution 12, displaying the annotations made by the health professional on the left and the entities detected by the model on the right. Colored highlighting distinguishes entity types: red for exames, yellow for resultados, blue for membros, pink for datas, and green for tempos.

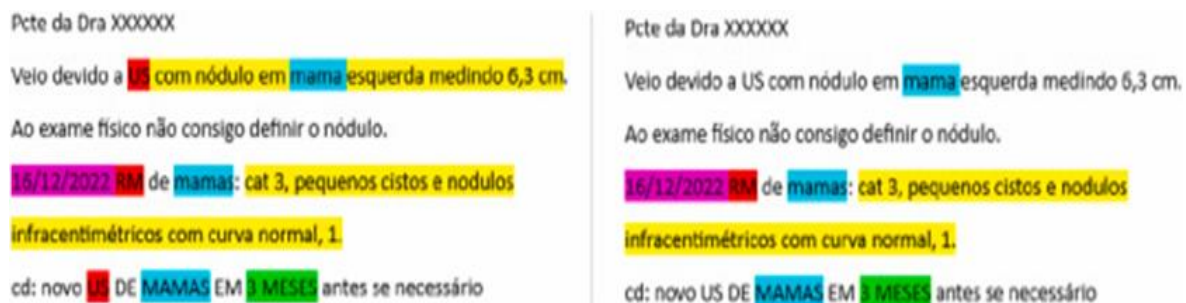


Figure 7. Comparison of evolution 12.

Figure 7 highlights that the model effectively captured most entities flagged by the healthcare professional. It performed particularly well when recognizing items in the Membro, Tempo, and Data groups. On the other hand, it overlooked the US (Ultrassom - Ultrasound)

exams and did not detect the resultado describing “nodule in the left breast measuring 6.3 cm”. **Table 4** provides a complete side-by-side overview of the expert annotations versus the model outputs.

Table 4. Model accuracy comparison. The opening column displays the document identifier. In the Exame, Resultado, Tempo, Membro, and Data columns, the left-side number reflects the count of annotations provided by the health professional (HP). In contrast, the right-side number corresponds to those produced by the model (Mod). Each row further reports the average accuracy achieved for the markings in that particular document.

Exame		Resultado		Tempo		Membro		Data		Accuracy
HP	Mod	HP	Mod	HP	Mod	HP	Mod	HP	Mod	
5	3	15	4	0	0	0	0	18	18	65.79 %
30	21	29	9	0	0	12	12	14	14	65.88 %
15	9	13	4	0	0	5	3	8	8	58.54 %
2	2	2	0	0	0	1	1	1	1	66.67 %
6	3	6	4	0	0	2	2	6	5	70.00 %
5	2	5	2	0	0	0	0	8	7	61.11 %
2	0	2	1	0	0	3	1	2	2	44.44 %
2	0	6	3	0	0	2	2	14	14	79.17 %
28	17	28	8	0	0	12	11	16	16	61.90 %
2	1	2	0	0	0	1	0	7	6	58.33 %
4	2	3	1	0	0	1	1	5	2	46.15 %
3	1	2	1	1	1	3	3	1	1	70.00 %
2	1	6	3	0	0	2	1	3	2	53.85 %
3	0	3	1	0	0	1	1	10	9	64.71 %
21	9	20	6	0	0	7	7	15	15	58.73 %
2	2	3	1	0	0	1	0	8	8	78.57 %
5	2	5	3	0	0	3	3	3	3	68.75 %
5	3	5	1	0	0	5	4	4	4	63.16 %
2	1	2	0	0	0	1	1	1	1	50.00 %
5	2	5	1	0	0	0	0	7	6	52.94 %
16	10	16	5	0	0	7	5	6	6	57.78 %
2	2	2	2	0	0	1	1	2	2	100.00 %
55.68 %		33.33 %		100 %		84.28 %		94.33 %		73.52 %

The model achieved an overall average accuracy of 73.52% in its responses, reflecting a solid level of agreement with the healthcare expert’s annotations. An in-depth review of individual evolutions also uncovered instances where the model produced notably divergent outcomes relative to the expert’s markings. When performance was broken down by category, the model performed weakest in identifying Resultado tags (33.33%), highlighting challenges in detecting and extracting intricate details from medical examinations. These exams typically involve intricate language,

specialized terminology, and numerous abbreviations. For instance, in evolution 9, the model matched only 8 out of the 28 Resultado annotations made by the health professional, underscoring a pronounced struggle in this area. Clinicians’ identifications in this category frequently arise from extended textual passages that lack predictable structures—a challenge worth investigating more thoroughly in subsequent studies. Conversely, the model reached perfect accuracy (100%) for Tempo tags, though this was based on just a single instance in the dataset. The Membro and Data categories

performed nearer to the overall average, with accuracies of 84.28% and 94.33%, respectively, demonstrating more reliable model behavior in those areas. For Exame, despite the frequent use of acronyms, the system delivered middling results at 55.68% accuracy. Evolution 22 stood out as a case where the model attained 100% correctness across every category; however, this success likely stems from the limited scope of the sample, which contained only 2 annotations per category and thus represented an easier scenario for the model. Lastly, **Figure 8** provides a visual breakdown of the performance across categories.

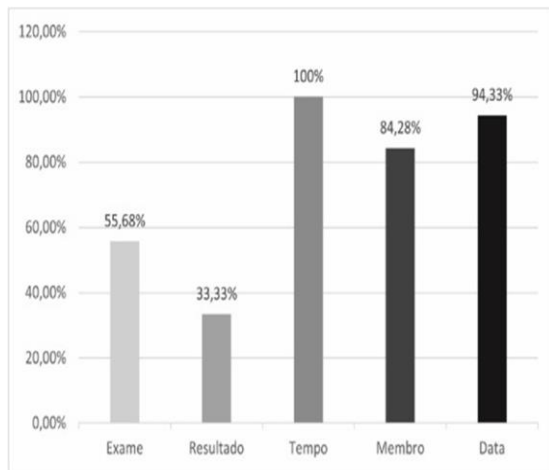


Figure 8. Comparison of accuracy by category.

A further key observation from the evaluation concerns the link between the intricacy of the content within the clinical evolutions and the model's effectiveness. Evolutions composed in straightforward plain-text style yielded stronger outcomes, whereas those pulled from tabular structures or featuring fragmented, non-continuous data exhibited reduced performance. This clearly illustrates the issue and points to potential limitations of the model when dealing with varying data organization and flow.

After assessing the model's precision, the evolutions were incorporated into the ontology. The following section demonstrates this procedure through a concrete case.

Once the model processed the input, it identified and pulled out the relevant entities. In this particular example, the system recognized and extracted these category entities and corresponding values: data(26/12/2018), data(01/07/2019), exame(cea), resultado(252), exame(vid d), resultado(249). Subsequently, the

identified relations were established and added to the ontology.

In the experiment, an individual representing the evolution was generated within the ontology and classified under *Evolução*. This individual encompassed the CEA exam (recognized as Antígeno Carcinoembrionário - Carcinoembryonic Antigen) and the VIT D exam (Vitamina D - Vitamin D), together with the date of the evolution. Additionally, distinct individuals were instantiated for each extracted exam from the text and for their associated results. **Figure 9** displays the extracted evolution relations in a graphical representation.

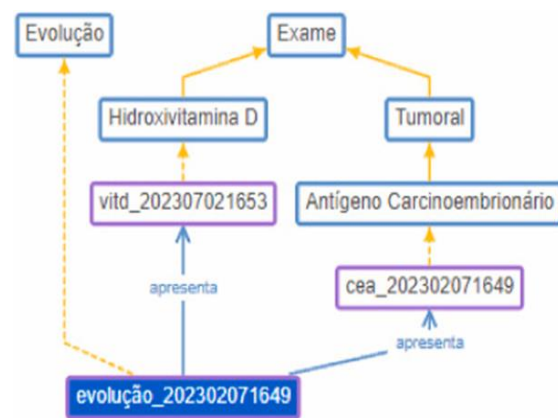


Figure 9. Flow of extracted evolution relations.

The ontology facilitates a more accurate and semantically rich organization of the extracted information. It also supports establishing connections between different entities, enhancing analytical capabilities and enabling sophisticated queries. By incorporating entities and relations derived from medical texts into the ontology, we achieved a more comprehensive and organized perspective on previously unstructured data, streamlining information retrieval and comprehension.

A preliminary prototype implementation plan, developed through alignment meetings with Interprocess, focuses on automating the structuring of data from physician-written exam notes during patient evolutions to automatically populate the SGO. Currently, healthcare staff complete this task manually. This module forms part of the patient's electronic health record (EHR), which includes core details, evolutions, prescriptions, schedules, referrals, documents, activities, and personal information. Within the exam section of the SGO, users can view a list of performed examinations, their outcomes, the responsible laboratory, and each test date.

The proposed solution uses an information extraction model to analyze recorded medical exams and automatically organize key details. As a result, the exams appear in a well-structured list. This eliminates manual data entry and gives healthcare professionals a clearer, more accessible overview. Overall, the approach aims to streamline workflows, reduce staff time demands, and minimize potential errors during data input.

The prototype integrated with the SGO is an initial version and is open to modifications and improvements based on user feedback and observed requirements. Collaboration with Interprocess will support joint development efforts. Upon full adoption, the SGO is anticipated to become more robust and effective, ultimately contributing to enhanced patient care and more efficient clinical information management in a secure, interconnected environment.

Assessment of acceptance by professionals

The ontology underwent evaluation using a scenario-based model. In this approach, an assessor systematically outlines user interaction scenarios with the system and appraises the actions needed to complete each one through a cognitive framework [39]. Given the scenario-defined objectives and the collected responses, the evaluator can pinpoint critical obstacles that might impede goal achievement and assess the acceptability of the system's features and resources.

The five health domain specialists previously mentioned completed the questionnaire. Before the assessment, these experts received a comprehensive overview of the developed system, an introduction to ontologies, and demonstrations of the model's operation, along with sample outputs. The acceptance evaluation aimed to gauge the specialists' receptiveness toward incorporating unstructured data into an ontology. Participants responded to a tailored questionnaire to collect relevant evidence. The questions included: I know of some software that performs this type of task; The use of the model can facilitate the formalization of knowledge in the health area; The use of the model can enable the development of new applications in the health area; The proposed model is helpful for the health area; The use of ontology as a basis for structuring data can simplify the understanding and visualization of health information; If you need structured information, you will use the template; It is easy to understand the structuring of the data with the model.

The findings indicate that 60% of respondents strongly disagreed when asked about their familiarity with any software capable of handling the proposed task, suggesting that the majority were unaware of existing tools for this purpose. This highlights a gap in awareness and adoption within the field, presenting an opportunity for further research. The remaining 40% agreed, indicating that a minority had some previous exposure to similar technologies.

Concerning the model's potential to support the formalization of knowledge in healthcare, 100% of participants fully agreed. This strong consensus reflects broad recognition of the model's value in advancing knowledge formalization in this domain and signals notable interest in its application. Similarly, regarding the model's capacity to foster the creation of new applications in healthcare, 100% of participants fully agreed. This unanimous view underscores the model's perceived viability as a foundation for innovative health-related tools.

In summary, the outcomes reveal a favorable view of the proposed model's strengths, particularly its ease of comprehension, practical utility in healthcare, potential for new application development, and improved data structuring. These findings offer valuable guidance for integrating the model into the SGO and underscore important areas for discussion.

Conclusion

This paper introduced a novel approach for incorporating natural language data into a domain-specific ontology. The case study used deep learning methods, including Transformer-based models such as BERT and the Bi-LSTM architecture, to perform named entity recognition (NER) and relation extraction (RE) tasks.

The proposed methodology was assessed by both health domain experts and computational metrics. Tests involved NER and RE techniques applied to authentic Portuguese-language medical records. The research contributed a new annotated dataset in Brazilian Portuguese focused on the medical field, which supported the experimental phase. The resulting model attained 78.24% accuracy in the exam domain and 72.87% in the diagnosis domain for NER and RE tasks. Furthermore, we constructed an oncology-oriented ontology and linked it to the model, comprising around 181 classes, 14 data properties, 12 object properties, and over 200 individuals. Specialist evaluation in the health

domain yielded a 73.52% success rate aligned with their annotations, while the usability study indicated strong acceptance.

The methodology developed here offers an effective solution for handling unstructured NER and RE processes in healthcare. The encouraging outcomes open avenues for future tools that can support the medical community by delivering a well-organized and dependable knowledge repository for professionals and researchers alike. The results demonstrate the value and relevance of this work to the health sector, while also revealing multiple directions for advancing relation extraction and refining strategies for structuring natural language data.

This research provides a foundation for building applications that support diverse analytical and operational activities. It leverages natural language processing to pull varied details from unstructured text and convert them into structured formats enriched with relational connections. Although centered on Portuguese texts, the study yields observations that may help manage medical documents in languages beyond English.

Future work

Future efforts include broadening the DGO-E through additional annotations of entities and relations within oncology. Greater data availability would support enhancements to the model's effectiveness. Another direction involves enriching the DGO-DF with annotations for entity relations. Once the datasets are expanded, the BiLSTM model could be further trained using the new annotated material. Such steps are expected to boost information extraction capabilities and deepen the understanding of relations within EHR evolutions.

A more extensive exploration of the ontology created in this study is also recommended. Beyond targeting particular entities and relations, data from the ontology could be utilized to produce oncology-focused questions. This could lead to an interactive system that engages users and supplies pertinent responses and insights drawn from the ontology's conceptual framework.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Ngiam KY, Khor W. Big data and machine learning algorithms for healthcare delivery. *Lancet Oncol.* 2019;20(5):e262-73.
2. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc.* 2016;23(5):1007-15.
3. Solares JRA, Raimondi FED, Zhu Y, Rahimian F, Canoy D, Tran J, et al. Deep learning for electronic health records: a comparative review of multiple deep neural architectures. *J Biomed Inform.* 2020;101:103337.
4. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc.* 2019;26(4):364-79.
5. Dhole G, Uke N. NLP based retrieval of medical information for diagnosis of human diseases. *Int J Renew Energy Technol.* 2014;3(10):243e8.
6. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform.* 2017;73:14-29.
7. Liu S, Wang X, Hou Y, Li G, Wang H, Xu H, et al. Multimodal data matters: language model pre-training over structured and unstructured electronic health records. *IEEE J Biomed Health Inform.* 2022;27(1):504-14.
8. Bitterman DS, Miller TA, Mak RH, Savova GK. Clinical natural language processing for radiation oncology: a review and practical primer. *Int J Radiat Oncol Biol Phys.* 2021;110(3):641-55.
9. 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):1208-16.
10. Qiao B, Zou Z, Huang Y, Fang K, Zhu X, Chen Y. A joint model for entity and relation extraction based on BERT. *Neural Comput Appl.* 2022;1-11.

11. Kim YM, Lee TH, Na SO. Constructing novel datasets for intent detection and NER in a Korean healthcare advice system: guidelines and empirical results. *Appl Intell.* 2023;53(1):941-61.
12. Leng J, Wang D, Ma X, Yu P, Wei L, Chen W. Bi-level artificial intelligence model for risk classification of acute respiratory diseases based on Chinese clinical data. *Appl Intell.* 2022;52(11):13114-31.
13. Li Z, Chen H, Qi R, Lin H, Chen H. Docr-bert: document-level r-bert for chemical-induced disease relation extraction via Gaussian probability distribution. *IEEE J Biomed Health Inform.* 2021;26(3):1341-52.
14. Qasim R, Bangyal WH, Alqarni MA, Almazroi AA. A fine-tuned BERT-based transfer learning approach for text classification. *J Healthc Eng.* 2022.
15. Gao S, Alawad M, Young MT, Gounley J, Schaefferkoetter N, Yoon HJ, et al. Limitations of transformers on clinical text classification. *IEEE J Biomed Health Inform.* 2021;25(9):3596-607.
16. Sun T, Wang D. Research on relation extraction method based on multi-channel convolution and BiLSTM model. In: 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). IEEE; 2020. p. 1082-90.
17. Chen L, Zhang W, Ye H. Accurate workload prediction for edge data centers: Savitzky-Golay filter, cnn and BiLSTM with attention mechanism. *Appl Intell.* 2022;52(11):13027-42.
18. Kitchenham BA, Charters S. Guidelines for performing systematic literature reviews in software engineering. Tech. Rep. EBSE 2007-001, Keele University and Durham University Joint Report; 2007.
19. Prisma. Preferred reporting items for systematic reviews and meta-analyses [Internet]. 2021 [cited 2021 Apr 8]. Available from: <http://prisma-statement.org/PRISMAStatement/Checklist.aspx>
20. Cooper ID. What is a “mapping study?”. *J Med Libr Assoc.* 2016;104(1):76.
21. Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D, et al. Behrt: transformer for electronic health records. *Sci Rep.* 2020;10(1):1-12.
22. Morgan S, Ranasinghe T, Zampieri M. Wlv-rit at germeval 2021: multitask learning with transformers to detect toxic, engaging, and fact-claiming comments. *arXiv:2108.00057.* 2021.
23. Xue K, Zhou Y, Ma Z, Ruan T, Zhang H, He P. Fine-tuning BERT for joint entity and relation extraction in Chinese medical text. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2019. p. 892-7.
24. 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.
25. Herlihy C, Rudinger R. Mednli is not immune: natural language inference artifacts in the clinical domain. *arXiv:2106.01491.* 2021.
26. Huang K, Altosaar J, Ranganath R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. *arXiv:1904.05342.* 2019.
27. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805.* 2018.
28. de Souza JVA, Schneider ETR, Cezar JO, Silva LE, Gumiel YB, Paraiso EC, et al. A multilabel approach to Portuguese clinical named entity recognition. *J Health Inform.* 2020;12.
29. Ji Z, Wei Q, Xu H. BERT-based ranking for biomedical entity normalization. *AMIA Summit Transl Sci Proc.* 2020;2020:269.
30. Lopes É, Correa U, de Freitas LA. Exploring BERT for aspect extraction in Portuguese language. In: The International FLAIRS Conference Proceedings. 2021;34.
31. Souza F, Nogueira R, Lotufo R. Bertimbau: pretrained BERT models for Brazilian Portuguese. In: Brazilian Conference on Intelligent Systems. Springer; 2020. p. 403-17.
32. Loshchilov I, Hutter F. Decoupled weight decay regularization. *arXiv:1711.05101.* 2017.
33. Xu G, Wang C, He X. Improving clinical named entity recognition with global neural attention. In: Web and Big Data: Second International Joint Conference, APWeb-WAIM 2018. Springer; 2018. p. 264-79.
34. Li S, Wang W, Lu B, Du X, Dong M, Zhang T, et al. Long-term structural health monitoring for bridge based on back propagation neural network and long and short-term memory. *Struct Health Monit.* 2023;22(4):2325-45.

35. Man D. Ontologies in computer science. *Didactica Mathematica*. 2015;31(1).
36. He Y, Chen J, Dong H, Horrocks I, Allocca C, Kim T, et al. Deeponto: a Python package for ontology engineering with deep learning. *arXiv:2307.03067*. 2023.
37. Niero LHP, Guilherme IR, Oliveira LES, de Araújo Filho GM. Psybertpt: a clinical entity recognition model for psychiatric narratives. In: 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2023. p. 672-7.
38. Bettoni GN. Extração de informação em evoluções clínicas e integração com dados farmacogenômicos [Master's thesis]. Pontifícia Universidade Católica do Rio Grande do Sul; 2022.
39. Sugimura V, Ishigaki VK. New web-usability evaluation method: scenario-based walkthrough. *Fujitsu Sci Tech J*. 2005;41(1):105-14.