

Original Paper

Natural Language Processing and Graph Theory: Making Sense of Imaging Records in a Novel Representation Frame

Laurent Binsfeld Gonçalves, MA; Ivan Nestic, MSc; Marko Obradovic, MSc; Bram Stieltjes, MD, PhD; Thomas Weikert, MD; Jens Bremerich, MD, PhD

Clinic of Radiology & Nuclear Medicine, University Hospital Basel, University of Basel, Basel, Switzerland

Corresponding Author:

Laurent Binsfeld Gonçalves, MA
Clinic of Radiology & Nuclear Medicine
University Hospital Basel
University of Basel
Petersgraben, 4
Basel, 4031
Switzerland
Phone: 352 621517916
Email: laurent.binsfeld@gmail.com

Abstract

Background: A concise visualization framework of related reports would increase readability and improve patient management. To this end, temporal referrals to prior comparative exams are an essential connection to previous exams in written reports. Due to unstructured narrative texts' variable structure and content, their extraction is hampered by poor computer readability. Natural language processing (NLP) permits the extraction of structured information from unstructured texts automatically and can serve as an essential input for such a novel visualization framework.

Objective: This study proposes and evaluates an NLP-based algorithm capable of extracting the temporal referrals in written radiology reports, applies it to all the radiology reports generated for 10 years, introduces a graphical representation of imaging reports, and investigates its benefits for clinical and research purposes.

Methods: In this single-center, university hospital, retrospective study, we developed a convolutional neural network capable of extracting the date of referrals from imaging reports. The model's performance was assessed by calculating precision, recall, and F1-score using an independent test set of 149 reports. Next, the algorithm was applied to our department's radiology reports generated from 2011 to 2021. Finally, the reports and their metadata were represented in a modifiable graph.

Results: For extracting the date of referrals, the named-entity recognition (NER) model had a high precision of 0.93, a recall of 0.95, and an F1-score of 0.94. A total of 1,684,635 reports were included in the analysis. Temporal reference was mentioned in 53.3% (656,852/1,684,635), explicitly stated as not available in 21.0% (258,386/1,684,635), and omitted in 25.7% (317,059/1,684,635) of the reports. Imaging records can be visualized in a directed and modifiable graph, in which the referring links represent the connecting arrows.

Conclusions: Automatically extracting the date of referrals from unstructured radiology reports using deep learning NLP algorithms is feasible. Graphs refined the selection of distinct pathology pathways, facilitated the revelation of missing comparisons, and enabled the query of specific referring exam sequences. Further work is needed to evaluate its benefits in clinics, research, and resource planning.

(*JMIR Med Inform* 2022;10(12):e40534) doi: [10.2196/40534](https://doi.org/10.2196/40534)

KEYWORDS

radiology; deep learning; NLP; radiology reports; imaging record; temporal referrals; date extraction; graph theory; health care information system; resource planning.

Introduction

Radiology departments generate tremendous amounts of reports every day. Narrative radiology reports are the primary communication medium between radiologists and referring physicians, thus playing a central role in patient care and containing a large variety of health care information [1,2]. From 1996 to 2010, image study volume for computed tomography (CT) and magnetic resonance imaging (MRI) increased by 280% to 380% [3]. Radiology embraced digital workflows and electronic information transfer to referring colleagues early on, which virtually eradicated analog data in this field [4]. This early commitment provides enormous quantities of digitalized reporting data containing interpretative image descriptions. However, the extraction of this information is hampered because unstructured reports are poorly computer-readable [5]. Semantic reports contain valuable information at a granular level (eg, multiple temporal referrals) that can be evoked for the overall report or specific findings in multiple document locations. This multilocal information cannot easily be determined on a whole document level [6].

Natural language processing (NLP) is one solution to the problem of extracting specific information from the plethora of free-text radiology reports. NLP is defined as the analysis of linguistic data, most commonly in the form of textual data, using computational methods [7-13]. NLP has evolved from rule-based to machine learning algorithms [14-20], deep learning being a subset of the latter that applies multilayer neural networks [21,22]. Its capability to automatically extract structured information has been described in many medical research settings [23-29]. Especially in radiology, there are numerous instances where it has demonstrated excellent text mining performances, including the detection of incidental findings and recommendations [30-32], actionable findings [33], specific findings [34-41], quality assessment of reports [42,43], and the generation of curated data sets [44-49].

The quantitative accumulation of radiology reports per patient over the years has led to a highly interconnected network of exams. Modern picture archiving and communication systems (PACS) represent the different exams as a list sorted by their acquisition date. Most systems can highlight the previous exams of roughly the same region in the study description to the user. This type of comparative visualization does not consider multiregional studies or often-encountered findings at the margins of the acquired field of view. It does not foreground the dates to which the radiologist compared his findings in the report. This last part especially is a significant shortcoming for clinicians reviewing patient history. They have to read every report carefully to see to which point in time the radiologist compared tumor progress, for example, or if the images from an external institute were available to the radiologist at the exact time when reading the follow-up exam.

One crucial connection in this context is dated referrals to prior exams. The good practice guidelines for radiological reporting from the European Society of Radiology [50] and the 2020 revised American College of Radiology practice parameter for communication of diagnostic imaging findings emphasize the

need for comparison with previous investigations, including the date of previous reports and mentioning the absence of previous imaging. By using comparison studies, radiologists make more observations, gain confidence in their interpretation, and provide more diagnoses [51-55]. One study found that the diagnostic accuracy, sensitivity, and specificity in mammography increased as the false-positive rate decreased [56]. Various recent studies relied on NLP techniques to extract the temporality of measurements in imaging reports (ie, attributing an observation on the current or prior exam) [39-62]. However, to the best of our knowledge, no methods that extract every referring date from semantic radiological texts have been researched. Moreover, no studies in the literature have focused on the overall temporal indexing of the report assessed, in most instances, by the radiologist at the beginning of the report.

One solution to displaying connections between a multitude of different reports is graph representation. Graph theory defines graphs as a set of properties stored in nodes connected by edges, which represent a relationship between the connected nodes [63,64]. A review paper from 2020 found that graphs, as defined by graph theory, are hardly used to represent patient data in a clinical context; in the literature review, only 11 papers matched the description [65].

This study aimed to develop a novel and concise visualization framework of related reports.

To this end, we applied a self-designed NLP algorithm capable of extracting the referencing dates from unstructured radiology reports on all the reports generated for 10 years at a university hospital. This information was an essential input for a relational graph in which the nodes represent the radiology reports with their associated metadata and the dated referrals are their connecting edges. Finally, we investigated the potential benefits of such a graph representation and storage for clinical and research purposes.

Methods

Ethics Approval

Institutional review board approval and the requirement for informed consent were waived (institutional review board: Ethikkommission Nordwest- und Zentralschweiz) since no patient identifiers were used. Collected data consisted of plain text from radiology reports and randomized metadata, neither of which could be tracked back to radiologists, individual patients, nor referring clinicians.

Data Set Acquisition and Description

We extracted all radiology reports from January 2011 to December 2021 as well as a selection of their associated Digital Imaging and Communications in Medicine (DICOM) metadata (ie, randomized patient ID, modality type, body region, study date) from the hospital database. All reports were written in German and derived from all the imaging modalities (ie, ultrasound, radiography, mammography, x-ray angiography, CT, MRI, nuclear medicine exams, and positron emission tomography [PET]-CT). The reports were a mix of unstructured free-text reports and standardized templates, either containing subheadings for distinct organs with prewritten normal findings

(eg, CT chest-abdomen) or checklists for standardized reporting features (eg, Liver Imaging Reporting and Data System for liver MRI). The broad structure of the reports was usually divided into 5 sections: medical history, medical question, examination protocol, radiological finding, and impression.

Every radiology exam had a predefined body region and modality type in its DICOM metadata. There were 14 body regions and 9 modalities (see [Multimedia Appendix 1](#)).

Construction of a Temporal Reference Extraction Algorithm

Data Selection for Training

We randomly selected 5187 reports from the previously extracted radiology reports.

Data Annotation

An internally developed data annotation tool, “xtag,” was used. A second-year medical resident (LBG) manually labeled 5187 reports with 5 classes indicating the temporal reference ([Table 1](#)). The annotation classes “date,” “today,” “yesterday,” and “no previous” were applied on the text sequence level (ie, annotating sequences of numbers or words). The annotation class “missing” was applied at the document level and was exclusive, meaning that no other annotation could be applied. On the other hand, “date,” “today,” “yesterday,” and “no previous” could be applied multiple times per report. To assess the necessity of a second reading, a fifth-year medical resident in radiology (TW) annotated 100 randomly selected reports. This process yielded 100% agreement among readers. Considering the simplicity of the task and based on this result, we refrained from a second reading of the whole data set.

Table 1. Annotated classes and their defined meaning.

Class	Meaning
Date	Precise numerical date referring to a comparative exam; any numerical or partially numerical format was accepted.
Today	Non-numerical temporal reference to a comparative study done on the same day as the actual report (ie, any literal expression meaning today)
Yesterday	Non-numerical temporal reference to a comparative study done on the day before the actual report (ie, any literal expression meaning yesterday)
No previous	Explicit statement that no comparable previous exams are available
Missing	No mention of a comparative study

Data Format

The training pipelines required the annotations to conform to the IOB2 format [66,67]. The predictions were also produced in the same format (further technical information can be found in [Multimedia Appendix 2](#) [5,68-72]).

Algorithm Training and Testing

We excluded 2392 reports from the annotated data set, as they did not contain temporal links. We split the data into a training/validation data set of 2646 reports (94.6%) and an independent test data set of 149 reports (5.4%). We estimated that 5% for an independent and second test data set is a valid representation, as we verified the algorithm's robustness using the 5-fold cross-validation [73]. We also considered the low output variability of the problem to be solved. We used the Spacy sentencizer to text into sentences before training. We then used the ktrain library to produce a bidirectional long short-term memory (LSTM) [74] model starting with pretrained fastText word embeddings [75] (for details, see [Multimedia Appendix 2](#)). We applied various rule-based date extraction algorithms on the predicted date sequences to extract as many dates as possible. The non-numerical classes today and yesterday were converted into a numerical format using the date of the referring report as a reference. The dates missing the year specification were assigned the same year as the referencing report. The prediction was ignored if the day or month was missing. A grid search algorithm tested different combinations of learning rates and batch sizes to find the near optimal parameters for our training algorithm. A 5-fold cross-validation

[76] of the training data set with 20% of the reports as validation in each round was performed to evaluate the model's performance on large independent data sets. The data set split into folds was done at the report level. The model was tested on the independent test data set in a final evaluation step. The following performance evaluation metrics were used to assess the trained model's quality: precision, recall, and F_1 -score [77].

Extraction of the Referenced Modality and Body Region

The referenced modality was extracted using a simple rule-based approach. After extracting the temporal references from the report, the algorithm searched for a mention of the modality in the sentence with the date reference. The previous report's body part was derived from its metadata and was assumed to be the same as the referencing report's body region.

Graduation of the Predicted Link's Confidence

We graduated the prediction's confidence as follows: (1) date, modality, and body part; (2) date and modality; (3) date and body part; (4) date. This confidence graduation was established as a link property, in which 1 was the most confident and 4 was the least confident. The link was discarded if it was impossible to generate it based on these 4 principles. This approach permitted narrowing and increasing the accuracy of the referenced reports if more than one exam was acquired on the referenced date.

Algorithm Application and Data Extraction on the Complete Data Set

The preparatory steps for extraction of the temporal information by the trained model were the same as for the training part. The result of the model's application to all the reports from 2011 to 2021 was a per-token table with labels for each token in the IOB2 format. Predictions that did not comply with the IOB2 format were removed.

Populating the Graph Database

The graph database system used was Neo4j (Version 4.4.). All the reports and a selection of their associated metadata from 2011 to 2021 were imported via the py2neo library. The metadata consisted of the exam's acquisition date, name, modality, and body region, as well as its randomized patient ID. The reports and their metadata were assigned to the vertices, and unidirectional edges from referencing reports to referenced reports were created. We assigned 3 properties to the edges: The first consisted of the inferred class called "reference class," the second showed the extracted string, and the third displayed the prediction's confidence.

Interactive Exploration of the Graph

The assessment of the potential benefits of patient data visualization in a graph was explored interactively. The aim

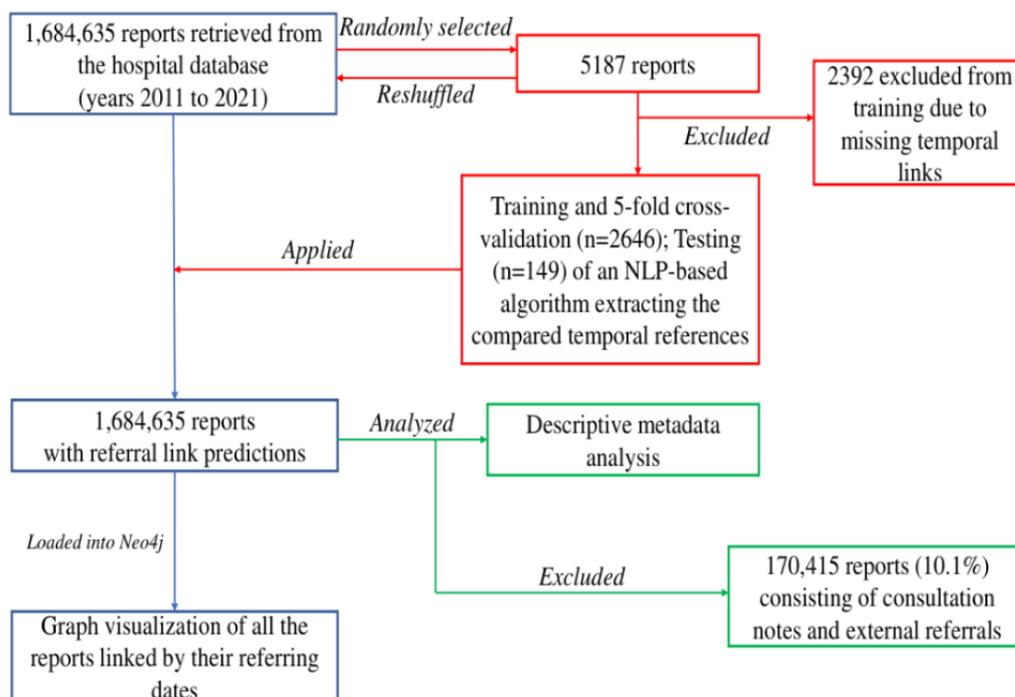
was to offer, at one glance, a well-ordered overview of the patient's imaging history with the related reports; enable comparison to previous exams; and represent the desired pathology pathway in a concise way (eg, oncological or postoperative follow-up imaging). In addition, it reveals to the clinician and the radiologist at what point in time the radiologist made his or her comparison. The user should be able to restrict his or her search to individually adaptable filters in the report's metadata (eg, body region, modality type, report date, or keywords in the reports' text). Another important feature would be to provide precisely filtered examinations in a concise order, in which every exam has its precisely defined position in a sequence. A final goal was to assess missed comparisons to previous exams, which was hoped to be achieved visually by spotting the missing link in the graph and by self-designable search algorithms.

Results

Data Set

In total, 1,684,635 reports from 264,655 distinct patients were extracted. We excluded 170,415 (10.1%) reports from the metadata analysis because they consisted of consultation notes and external referrals (detailed count in [Multimedia Appendix 3](#)). [Figure 1](#) shows the detailed methodical flowchart.

Figure 1. Study flowchart of 1,684,635 patient reports retrieved from the hospital database (2011 to 2021). NLP: natural language processing.



Annotation Distribution

A total of 7860 annotations were applied to 5187 reports from 2011 to 2019. Class distribution of the training data set was as follows: 44% date reference, 27% no previous comparative exam, 23% missing temporal link and 6% referral class "today." We removed the semantic referencing class "yesterday" from our data set as there were not enough training samples (34/5187, 0.7%).

Temporal Information Extraction Algorithm

Hyperparameter Optimization

The algorithm's output yielded an optimal learning rate of 1e-2 and a batch size of 1024. The random state was fixed for reproducibility. The maximal number of training epochs was limited to a never reached limit of 30.

Training and Testing

The stagnation in the validation performance of 3 epochs was targeted for the early stopping. During training, the model is stored after each epoch. After the completion of the training

process, the best-performing epoch weights were used for the final model. The same procedure was used for all the steps in which training was involved. After 5-fold cross-validation (results in [Multimedia Appendix 4](#)), the algorithm's performance was tested on the previously unused test data set ([Table 2](#)).

Table 2. Test results on 149 previously unused reports.

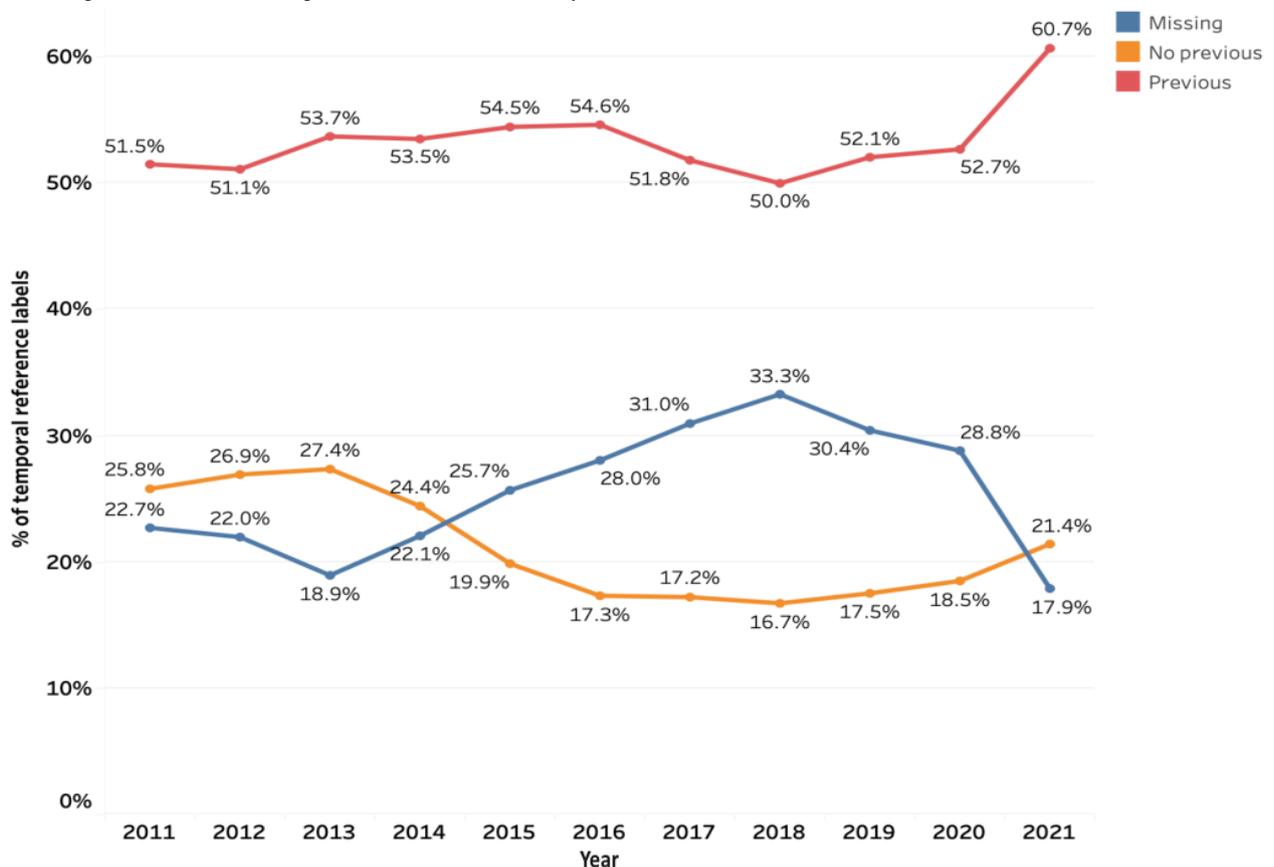
Variable	Precision (95% CI)	Recall (95% CI)	F ₁ -score (95% CI)
Date	0.93 (0.89-0.93)	0.9 (0.86-0.93)	0.93 (0.91-0.94)
No previous	0.94 (0.95-0.97)	0.98 (0.96-0.98)	0.96 (0.93-0.98)
Today	0.76 (0.73-0.88)	0.85 (0.79-0.90)	0.83 (0.79-0.93)
Micro average	0.93 (0.91-0.94)	0.92 (0.90-0.95)	0.94 (0.89-0.93)
Macro average	0.86 (0.84-0.95)	0.91 (0.87-0.95)	0.91 (0.80-0.94)
Weighted average	0.93 (0.91-0.94)	0.93 (0.90-0.94)	0.94 (0.91-0.95)

Temporal Referencing Analysis

A temporal reference to comparable exams was mentioned in 53.3% (656,852/1,232,297), explicitly stated as not available in 21.0% (258,386/1,232,297), and omitted in 25.7% (317,059/1,232,297) of the reports. Variability over the years was asserted ([Figure 2](#)). The modalities with the least amount of missing references were mammography (41,197/545,636, 7.6%), PET/CT (1850/18,500, 10.3%), and CT (278,286/2,399,017, 11.6%). On the other hand, angiography (33,924/40,872, 83.2%) and ultrasound (94,080/254,270, 37.2%) had the most missing references ([Table S4 in Multimedia Appendix 5](#)). The body regions with the lowest amount of

missing references were trunk (3072/39,639, 7.8%), breast (5727/70,617, 8.1%), and thorax (25,646/276,060, 9.3%). On the other hand, the heart (19,030/26,090, 72.9%) and neck (14,716/23,230, 63.4%) regions had the most missing links ([Table S5 in Multimedia Appendix 5](#)). Modalities primarily referred to the same modality except for angiography referring to plain radiographs in 39.8% (1790/4503), PET/CT referring to MRI in 45.1% (456/1013), and nuclear medicine exams referring to CT in 33.9% (3500/10294; [Table S6 in Multimedia Appendix 5](#)). Every body region predominantly referred to the same body region. The most extreme example was “breast,” which was referenced in 99.0% (59,619/60,221) of the cases by the other breast studies.

Figure 2. Temporal reference of the reports (n=1,514,220) over the years.



Analysis of the Median Time Period of the Referencing Reports

The median period between referencing reports from 2011 to 2021 was determined in days, per modality (Table 3) and body

region (Table 4). The most extended periods were found in mammography (372 days) and the corresponding body region breast (370 days). The shortest periods were observed in plain radiograph reports (19 days) and the thorax region (10 days).

Table 3. Median time period between referencing reports (n=757,249) per modality.

Modality	Time period (days), median (Q1-Q3)	IQR	P value
Computed radiography	19 (2-118)	116	<.001
X-ray angiography	35 (7-137)	130	.048
Computed tomography (CT)	42 (3-231)	228	<.001
Magnetic resonance	65 (3-344)	341	.002
Nuclear medicine	114 (8-440)	432	.06
PET ^a /CT	129.5 (30-366)	336	<.001
Ultrasound	344 (24-386)	362	.002
Mammography	372 (352-722)	368	.03

^aPET: positron emission tomography.

Table 4. Median time period between referencing reports (n=757,249) per body region.

Body region	Time period (days), median (Q1-Q3)	IQR	P value
Thorax	10 (2-156)	154	.01
Upper extremity	11 (1-48)	47	.01
Abdomen	35 (4-237)	233	.006
Spine	35 (3-207)	204	<.001
Pelvis	39 (3-146)	143	.001
Lower extremity	42 (6-136)	130	.01
Head	65 (2-364)	362	<.001
Trunk	89 (34-196)	162	.03
Heart	125 (8-378)	370	=.40
Whole body	128 (8-427.3)	419.3	<.001
Neck	182 (29-395)	366	.045
Breast	370 (348-550)	202	.009

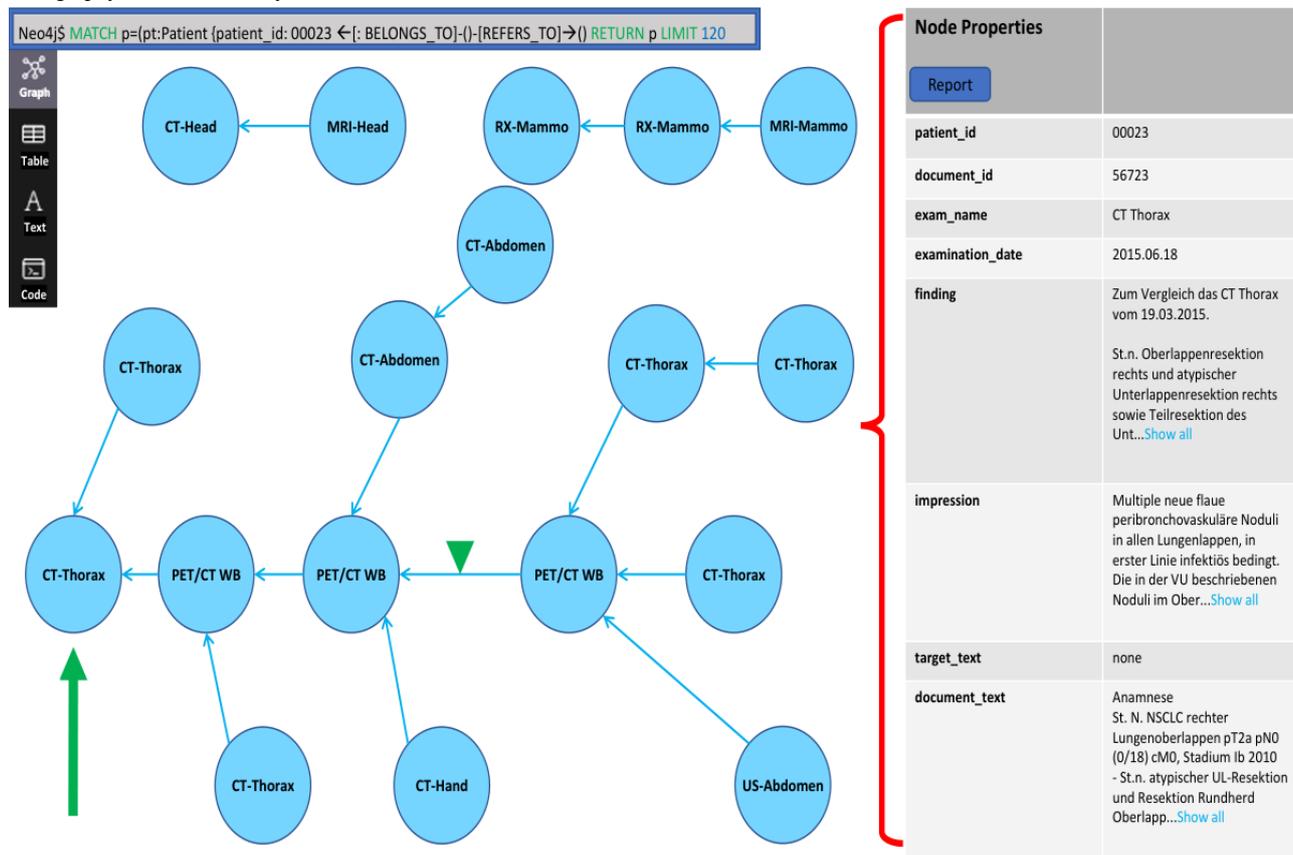
Exploration of Imaging Records in a Graph

General Overview

All the imaging reports and metadata from 2011 to 2021 were successfully loaded into a directed graph. The blue nodes represented the different patient reports labeled with their examination name (eg, CT-chest or MRI-head), and the

connecting links were their automatically extracted referral dates. The interface was individually adaptable (eg, the user could freely position the nodes as desired, and the colors of the individual components and displayed metadata were customizable). The total number of distinct patient reports could be selected at the beginning of any query. This view permitted a rapid visual assessment of the earliest comparison exam (Figure 3).

Figure 3. Single patient case example (lung cancer with cancer-related studies) including the full user interface in which the blue box in the top left is the query interface; the blue nodes contain the examination name, represent all the imaging studies stored in the picture archiving and communication system (PACS), and are ordered from oldest to newest on the right; the connecting blue arrows represent their referral links; and a node's metadata (examination name, acquisition date, text of the selected and referenced reports, and finding and impression sections) appears on the right side when the node is clicked on. CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; RX: x-ray; US: ultrasonography; WB: whole body.

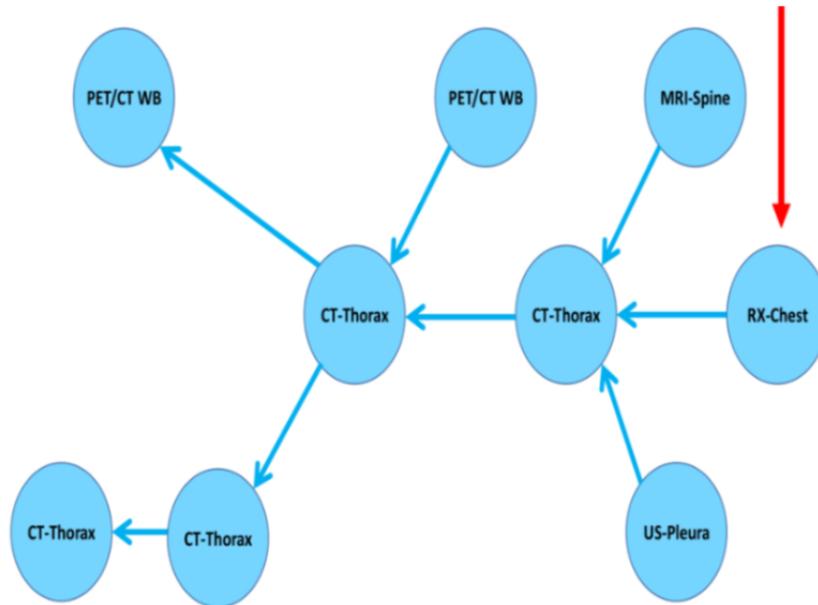


Multiparametric Filtered Representations

Narrowing the reports down to the most relevant and thus facilitating visualization are of utmost relevance with the high number of exams per patient. By clicking on the node of interest, the user could opt to display solely the linked reports (visualized in Figure 4). Another possible method of restricting the view

and looking for specific findings was a search filter related to the associated metadata and specific words in the report's text. One possible concept would be to look for specific exams with no previous reference and a defined pathologic condition as a keyword in the report's text, which would speed up the selection of the first exam associated with this condition.

Figure 4. Filtered representation of related studies as a partial screenshot within the user interface, in which all the related prior exams appear at one glance by clicking on the last node referring to a lung cancer (red arrow). Although the user selected the most recent study, clicking on every other node in this network would have resulted in the same view. CT: computed tomography; MRI: magnetic resonance imaging; PET: positron emission tomography; US: ultrasonography; WB: whole body.



Specific Exam Sequence Selection

Selecting highly customizable sequences of referring exams with specific metadata attributes (eg, chest x-ray followed by

chest CT) was possible. This can be refined, for example, with a period restriction or restricted time interval between the related exams (Figures 5 and 6).

Figure 5. Specific exam sequence selection as a partial screenshot within the user interface, in which we used the query field to randomly select 300 reports (blue nodes) of head computed tomography (CT) referenced by a head magnetic resonance image (MRI) that was acquired no longer than 3 days later and contained the keyword “infarct” in the impression field.

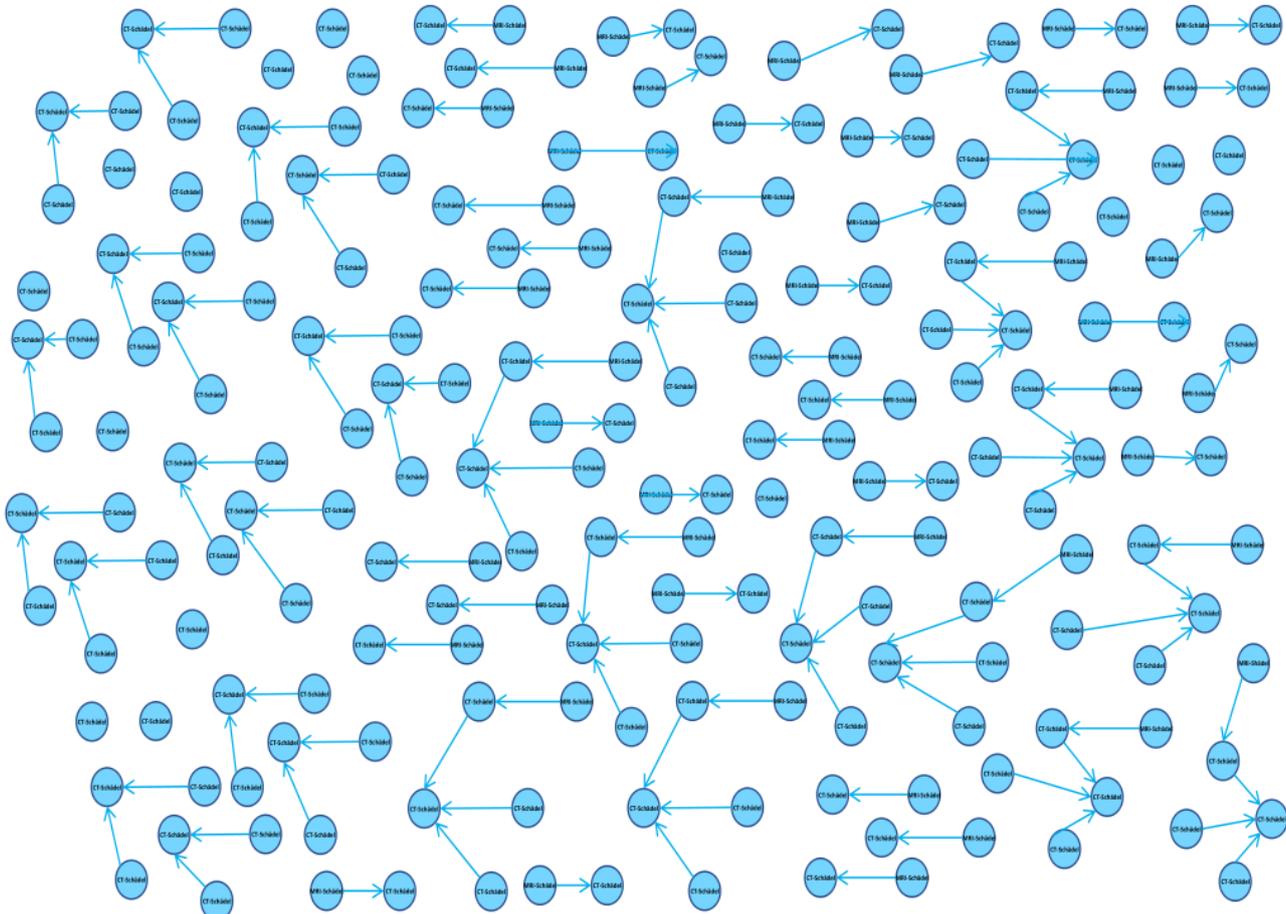
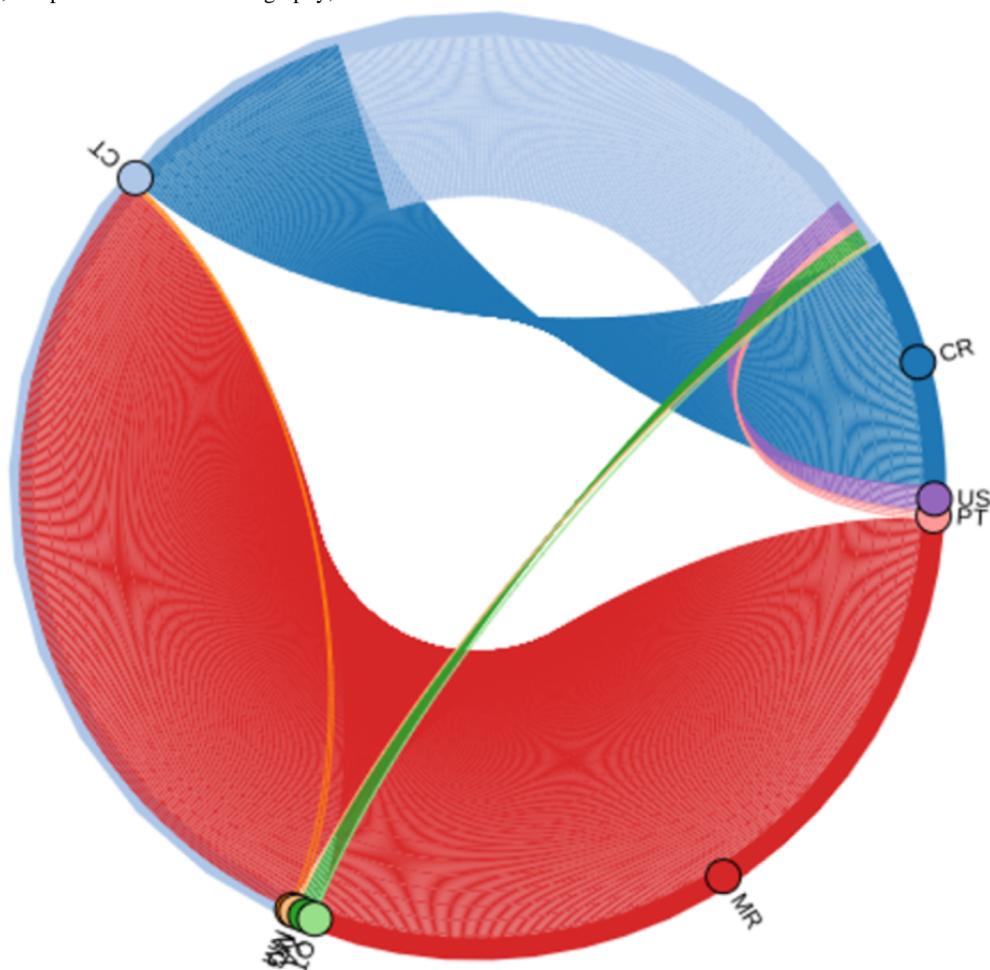


Figure 6. Chord diagram representing the connections between modalities in the head region referencing a head computed tomography (CT) image (light blue rim) during the 7 days after its acquisition. The size of the arc is proportional to the number of referenced reports. Most referring reports are head magnetic resonance images (MRIs), followed by other head CT images. CR: computed radiography; MR: magnetic resonance; NM: nuclear medicine; OT: other; PT: positron emission tomography; US: ultrasound.



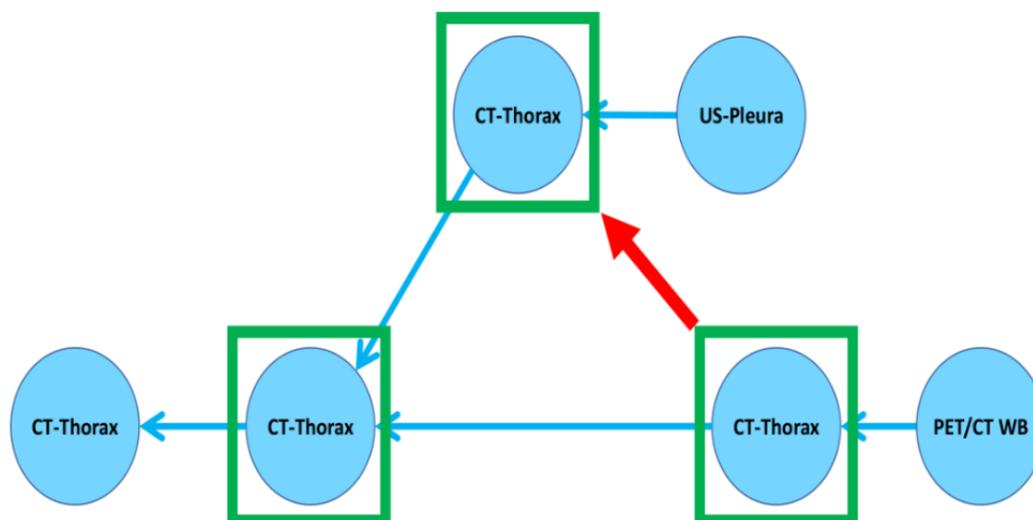
Visual and Filter-Aided Detection of Missing Comparative Connections

Selective queries with sequential filters and graph visualization permitted a rapid assessment of situations in which referral links were missing (Figure 7). This feature was helpful when

preceding comparative exams had been overlooked due to the poor list-like appearance of exam history in PACS or radiology information systems as well as when previous external images were imported into the PACS after the acquisition and reading of the following exam.

Figure 7. Single patient case example illustrating a missing temporal reference (red arrow) between subsequent reports (blue nodes, ordered by the earliest acquisition on the left to the latest on the right) of computed tomography (CT) studies of the thorax (green boxes). It is easy to detect a suspected missed link between the earlier CT-Thorax report at the top was not referenced by the later CT-Thorax report at the bottom right as well as via the search queries exposing temporal inconsistencies in the referrals (blue box at the top). PET: positron emission tomography; US: ultrasonography; WB: whole body.

```
Neo4j$ MATCH (:Organ {name: 'TH'}) <-(r2:Report)->(r1:Report)->(r3:Report)->(:Organ {name: 'TH'}) WHERE r1.examination_date < r2.examination_date AND r2.examination_date > r3.examination_date AND NOT EXISTS ((r3) -- (r2)) AND ((r1)->(:Modality {name: 'CT'}) -- (r2)) AND ((r3)->(:Modality {name: 'CT'})) AND r1.patient_id=1003211 RETURN r1, r2, r3 LIMIT 30.
```



Discussion

Principal Findings

As shown in this paper, representing imaging records in a directed graph is feasible. Connecting them via their referring dates improved visualization of related imaging pathways and detected missed exam comparisons. We also showed that automated extraction of referring dates from written radiology reports using a deep learning-based NLP algorithm, the groundwork needed to create the representation, is feasible and achievable with high significance (F_1 -score of 0.94).

Considering the extraction of concepts of temporality using NLP, our method can be compared with a publication from 2019 by Bozkurt et al [60]. Their main focus was extracting measurements and their core descriptors, among other things, their temporal context, for which they used rule-based NLP with predefined regular expressions. They solely focused on 2 temporal aspects (ie, current or prior), and their pipeline had a high F_1 -score of 0.85. Our approach uses a date-extracting LSTM. It focuses on all the referring dates in a written report, including the ones without a precise measurement, for example, the lesions that cannot be measured due to an amorphous configuration or the overall comparison date of the report. In addition, our algorithm has the crucial and unique advantage of detecting the explicit absence and missingness of a comparative exam from written text. Furthermore, we extracted every date of comparison from the report, thus permitting a comminuted and precise linkage for constructing a general graph.

Our approach, however, has the main disadvantage of not attributing the comparison date to specific findings or

measurements, which will slow down the focused review of specific entities in complex patient histories. Another disadvantage of our more granular extraction method is the high complexity of the task, which consecutively increases its dependency on correctly spelled referencing dates. Following this logic, omitted or wrongly chosen dates would have a greater impact on the integrity of the machine learning model and the graph in addition to the effect of varying writing habits or report templates between different institutions or radiologists. Although the reporting guidelines favor precisely dated comparisons, the radiologist does not always explicitly write the exact date of the compared finding in the text. As this omission mostly happens in comparison to the most recent report, which would be mentioned at the beginning of the report as the last referenced report, our method covers the majority of these cases. These aspects may render the overall applicability of our model more complex and susceptible to smaller errors than the temporality extraction algorithms developed so far.

In 2006, Lakhani et al [78] explored, in their large-scale database analysis of 1.8 million reports, how often radiologists compared with prior studies using a SQL approach. They found that 42.5% of reports completely omitted any reference to previous studies, 38.7% mentioned a comparison, and no relevant comparison was explicitly pointed out in only 18.8%. Although not entirely comparable, as they focused on a purely semantic approach of referring information extraction, it provides a good approximation, because if the reports contained phrases hinting toward a comparison, the date of the compared exam was most probably mentioned. In our study, reports referenced the date of a comparable exam (53%), explicitly stated that there was no previous exam (21%) more often, and were less prone to miss the referring link (26%). The best year for indicating

temporal references was 2021, with only 17.9% of reports missing a reference, down from 30.4% in 2019 and 28.8% in 2020. This tendency toward more temporal referrals could result from the increased emphasis on comparison exam consultation and report structuring in current reporting guidelines and digitalization, with many previous studies easily accessible. However, these percentages of prior exam consultation based on the written references in radiology reports are most probably underestimates. Haygood et al [79] concluded in their study from 2018 that the assumption that an older radiologic image or medical document was not consulted during radiologic interpretation merely because it is not cited in the report was not valid. This causes medicolegal issues for the reader. Radiologists were found negligent by juries for failure to compare a new chest radiograph with all previous chest radiographs [80]. Without written proof, this gets more difficult to defend. Another relevant aspect is the different extraction approaches. Our sentence-based named-entity recognition system analyzed data on a granular level, thus not missing single dates meant for comparison in parenthesis or other dates without a clear semantic indication of referral like the SQL approaches required.

Studies analyzing errors in the radiology reporting process emphasized the importance of comparing findings [52-56]. The good practice guidelines from the European Society of Radiology [50] and the 2020 revised American College of Radiology practice parameter for communication of diagnostic imaging findings support this affirmation. Kim and Mansfield [55] found that 5% of all errors in radiology resulted from failure to consult prior radiographic studies that could have led to the correct diagnosis. However, a critical review of the previous radiologists' findings or impressions should prevail when comparing previous exams. One must be careful not to follow an incorrect path; this error, called "satisfaction of report," accounted for 6% of all errors reported in radiology in the study by Kim and Mansfield [55]. The widespread availability of previous exams in modern PACS renders an excuse for failing to compare findings with prior exams obsolete. The automatic selection of comparative exams offered by modern PACS is inherently biased because it primarily considers the locoregional aspect, thus losing focus on multiregionality. For example, a CT of the cervical spine or shoulder may be overlooked as a potential comparison source when evaluating apical lung masses, or abdominal radiographs when interpreting hips. The same logic applies to clinicians or radiologists reviewing the imaging history of a given finding, especially in oncology, which has many multiregional studies and findings.

These complex considerations call for a well-arranged and organized visualization system. Poor usability and hampered visualization of patient data reduce the motivation of thoroughly reviewing them, which remains a challenge in health care and is associated with increased error rates due to missing pertinent details, user fatigue, and frustration [78,79]. A study from 2022 analyzing the impact of intensive care unit clinical information systems showed that poor interface design and visual representations are major sources of dissatisfaction among users [80]. Our explorations indicate that grouping related exams

together in a graph could help improve this fundamental and increasingly pressing user-friendliness issue.

We hope that, by reinforcing the radiologist's organizing role and improving the case overview by replacing the list appearance of imaging history, he or she will tend to omit the referring links less often, thus minimizing comparison error. Another critical aid is the improved detection of omitted connections in situations in which, for example, previously acquired external scans were loaded into the PACS after reading the following exam. This would be of great value for the subsequent physicians reviewing the imaging history. Temporal referrals in a report prove to the reader that the radiologist has not forgotten to compare a specific finding. This is a valuable asset, considering that a finding's relevance is often determined by its temporal course. For example, lung nodules, brain atrophy changes, or vascular aneurysms showing no dynamic changes over a long period are less alarming, especially in infants and older adults, for whom noninvasive imaging follow-ups are favored over invasive medical investigations. Optimizing visualization with a graph representation could save time as well as decrease unnecessary exams and radiation exposure for patients.

In specialized medicine, clinicians are more focused on specific regions or findings. Manually filtering out the irrelevant exams adds work and a source of potential error (eg, an orthopedic surgeon is more inclined to investigate images implying the healing process of a fracture or a neurologist the exams related to cerebral or spinal findings). Our graph enables the user with an exam of interest to select all the related studies and to omit, if desired, all the unrelated reports, thus substantially and instantly reducing the number of studies to be reviewed.

Our system can assist quality control and review of guideline adherence by rapidly filtering out selectable sequences of exams (eg, CT performed after an x-ray) refined by the possibility of restricting the search for an interstudy period. This highly customizable review based on the reports' metadata could also help research projects. For example, when evaluating the features of a brain lesion over time, one could filter out all the reports in the database in which the finding is described in the report text; these reports will then be shown, if desired independently of the patient, with their respective related reports. This approach rapidly and intuitively speeds up an otherwise fastidious query, offering the researcher a follow-up and quick method for taking the measurement steps on the associated images. The quantitative and qualitative predictions as well as the period of the related following radiology exams could be of great value for clinical management purposes, permitting optimal prediction of the necessary human and material resources.

Limitations

Our study has several limitations. The main limitation was that the analysis was based on a single tertiary care university hospital and depended strongly on our reporting customs. Second, reports were labeled by only 1 reader (a second-year resident). Given the low grade of complexity in labeling the referencing dates and the 100% agreement in a subset of 100 reports, we refrained from a second reading of the whole data

set. The more challenging task of determining the comparative study was done during the reporting process by at least one board-certified radiologist. Third, there were insufficient samples to train the non-numerical referencing dates expressing “yesterday.” This should be addressed in future work. One solution could be to use active learning algorithms prioritizing the model’s most uncertain predictions. Fourth, there was a lack of external validation. Also, to our knowledge, there is no comparable study in the literature. Nevertheless, the methodology should be reproducible in other radiology department setups to allow for future comparison. To this end, we have also made the codebase that allows for internal testing available ([Multimedia Appendix 2](#)). Fifth, the focus here was on the feasibility of an entire pipeline, including extraction and representation. Thus, we did not thoroughly evaluate its clinical usefulness but, instead, illustrated the potential usefulness in several use cases.

Future Prospects

The high performance of our NLP-based model at processing immense amounts of free-text data underlines its potential for future research projects. The process of filtering out comparative studies could be accelerated substantially, which could greatly benefit the development of image detection-based and NLP-based algorithms. The concept of a related graph database could optimize the engineering and designing of other medical software tools in radiology by improving visualization and user-friendliness, accelerating data selection in research projects,

and enhancing quality control and clinical review processes. An important amelioration could be the connection of the dates to the specific findings or measurements to which they are referring. Furthermore, it could enable resource planners to separately predict the necessary human and material resources. A significant asset of these databases is the easy-to-implement expansions (eg, integration of pathology reports or associated images). By giving users the possibility of correcting and adding links, it would be conceivable to create a continuously self-improving algorithm.

Conclusion

We established a proof of concept of an NLP-based algorithm capable of accurately extracting the dates of referrals on a granular level from unstructured radiology reports. We successfully generated customizable graphs of referring radiology reports, in which multiple filters may freely be applied, providing a well-arranged visual overview. This type of visualization permitted new possibilities for querying specific exam sequences, facilitated the detection of missed comparisons by the radiologist, and offers health care professionals a wide range of review opportunities. The radiologist’s awareness and motivation for the comparative aspect of his or her findings could be increased, and his or her worth for clinicians could be augmented by not solely providing information but also actively helping to organize it. Further work is needed to expand its features and evaluate its definite benefits in day-to-day clinical practice.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Defining body regions and modality types.

[\[DOCX File , 14 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Data format and machine learning.

[\[DOCX File , 14 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Distinct count in the whole data set.

[\[DOCX File , 35 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Five-fold cross-validation results.

[\[DOCX File , 13 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Metadata analysis of the temporal references.

[\[DOCX File , 19 KB-Multimedia Appendix 5\]](#)

References

1. Grieve FM, Plumb AA, Khan SH. Radiology reporting: a general practitioner's perspective. *Br J Radiol.* Jan 2010;83(985):17-22. [[FREE Full text](#)] [doi: [10.1259/bjr/16360063](https://doi.org/10.1259/bjr/16360063)] [Medline: [19470574](https://pubmed.ncbi.nlm.nih.gov/19470574/)]

2. Lee B, Whitehead MT. Radiology reports: What YOU think you're saying and what THEY think you're saying. *Curr Probl Diagn Radiol*. May 2017;46(3):186-195. [doi: [10.1067/j.cpradiol.2016.11.005](https://doi.org/10.1067/j.cpradiol.2016.11.005)] [Medline: [28069356](https://pubmed.ncbi.nlm.nih.gov/28069356/)]
3. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. *JAMA*. Jun 13, 2012;307(22):2400-2409. [FREE Full text] [doi: [10.1001/jama.2012.5960](https://doi.org/10.1001/jama.2012.5960)] [Medline: [22692172](https://pubmed.ncbi.nlm.nih.gov/22692172/)]
4. Avrin DE, Urbania TH. Demise of film. *Acad Radiol*. Mar 2014;21(3):303-304. [doi: [10.1016/j.acra.2013.12.008](https://doi.org/10.1016/j.acra.2013.12.008)] [Medline: [24507419](https://pubmed.ncbi.nlm.nih.gov/24507419/)]
5. Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. *JMIR Med Inform*. Mar 31, 2020;8(3):e17984. [FREE Full text] [doi: [10.2196/17984](https://doi.org/10.2196/17984)] [Medline: [32229465](https://pubmed.ncbi.nlm.nih.gov/32229465/)]
6. Short RG, Bralich J, Bogaty D, Befera NT. Comprehensive word-Level classification of screening mammography reports using a neural network sequence labeling approach. *J Digit Imaging*. Oct 18, 2019;32(5):685-692. [FREE Full text] [doi: [10.1007/s10278-018-0141-4](https://doi.org/10.1007/s10278-018-0141-4)] [Medline: [30338478](https://pubmed.ncbi.nlm.nih.gov/30338478/)]
7. Verspoor K, Cohen KB. Natural Language Processing. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, editors. *Encyclopedia of Systems Biology*. New York, NY: Springer; 2013;1495-1498.
8. Goldberg Y. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*. Nov 20, 2016;57:345-420. [FREE Full text] [doi: [10.1613/jair.4992](https://doi.org/10.1613/jair.4992)]
9. Hao T, Huang Z, Liang L, Weng H, Tang B. Health natural language processing: methodology development and applications. *JMIR Med Inform*. Oct 21, 2021;9(10):e23898. [FREE Full text] [doi: [10.2196/23898](https://doi.org/10.2196/23898)] [Medline: [34673533](https://pubmed.ncbi.nlm.nih.gov/34673533/)]
10. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *Int J Med Inform*. May 2019;125:37-46. [FREE Full text] [doi: [10.1016/j.ijmedinf.2019.02.008](https://doi.org/10.1016/j.ijmedinf.2019.02.008)] [Medline: [30914179](https://pubmed.ncbi.nlm.nih.gov/30914179/)]
11. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform*. Apr 27, 2019;7(2):e12239. [FREE Full text] [doi: [10.2196/12239](https://doi.org/10.2196/12239)] [Medline: [31066697](https://pubmed.ncbi.nlm.nih.gov/31066697/)]
12. Bae JH, Han HW, Yang SY, Song G, Sa S, Chung GE, et al. Natural language processing for assessing quality indicators in free-text colonoscopy and pathology reports: development and usability study. *JMIR Med Inform*. Apr 15, 2022;10(4):e35257. [FREE Full text] [doi: [10.2196/35257](https://doi.org/10.2196/35257)] [Medline: [35436226](https://pubmed.ncbi.nlm.nih.gov/35436226/)]
13. Chen J, Gong Z, Liu W. A nonparametric model for online topic discovery with word embeddings. *Information Sciences*. Dec 2019;504:32-47. [doi: [10.1016/j.ins.2019.07.048](https://doi.org/10.1016/j.ins.2019.07.048)]
14. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. *Insights Imaging*. Apr 04, 2019;10(1):44. [FREE Full text] [doi: [10.1186/s13244-019-0738-2](https://doi.org/10.1186/s13244-019-0738-2)] [Medline: [30949865](https://pubmed.ncbi.nlm.nih.gov/30949865/)]
15. Chen P, Zafar H, Galperin-Aizenberg M, Cook T. Integrating natural language processing and machine learning algorithms to categorize oncologic response in radiology reports. *J Digit Imaging*. Apr 27, 2018;31(2):178-184. [FREE Full text] [doi: [10.1007/s10278-017-0027-x](https://doi.org/10.1007/s10278-017-0027-x)] [Medline: [29079959](https://pubmed.ncbi.nlm.nih.gov/29079959/)]
16. Luo JW, Chong JJ. Review of natural language processing in radiology. *Neuroimaging Clin N Am*. Nov 2020;30(4):447-458. [doi: [10.1016/j.nic.2020.08.001](https://doi.org/10.1016/j.nic.2020.08.001)] [Medline: [33038995](https://pubmed.ncbi.nlm.nih.gov/33038995/)]
17. Dahl FA, Rama T, Hurlen P, Brekke PH, Husby H, Gundersen T, et al. Neural classification of Norwegian radiology reports: using NLP to detect findings in CT-scans of children. *BMC Med Inform Decis Mak*. Mar 04, 2021;21(1):84. [FREE Full text] [doi: [10.1186/s12911-021-01451-8](https://doi.org/10.1186/s12911-021-01451-8)] [Medline: [33663479](https://pubmed.ncbi.nlm.nih.gov/33663479/)]
18. Liu H, Zhang Z, Xu Y, Wang N, Huang Y, Yang Z, et al. Use of BERT (Bidirectional Encoder Representations from Transformers)-based deep learning method for extracting evidences in Chinese radiology reports: development of a computer-aided liver cancer diagnosis framework. *J Med Internet Res*. Jan 12, 2021;23(1):e19689. [FREE Full text] [doi: [10.2196/19689](https://doi.org/10.2196/19689)] [Medline: [33433395](https://pubmed.ncbi.nlm.nih.gov/33433395/)]
19. Bressemer K, Adams L, Gaudin R, Tröltzsch D, Hamm B, Makowski MR, et al. Highly accurate classification of chest radiographic reports using a deep learning natural language model pre-trained on 3.8 million text reports. *Bioinformatics*. Jan 29, 2021;36(21):5255-5261. [doi: [10.1093/bioinformatics/btaa668](https://doi.org/10.1093/bioinformatics/btaa668)] [Medline: [32702106](https://pubmed.ncbi.nlm.nih.gov/32702106/)]
20. Barash Y, Guralnik G, Tau N, Soffer S, Levy T, Shimon O, et al. Comparison of deep learning models for natural language processing-based classification of non-English head CT reports. *Neuroradiology*. Oct 25, 2020;62(10):1247-1256. [doi: [10.1007/s00234-020-02420-0](https://doi.org/10.1007/s00234-020-02420-0)] [Medline: [32335686](https://pubmed.ncbi.nlm.nih.gov/32335686/)]
21. Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. *Radiographics*. 2017;37(7):2113-2131. [doi: [10.1148/rg.2017170077](https://doi.org/10.1148/rg.2017170077)] [Medline: [29131760](https://pubmed.ncbi.nlm.nih.gov/29131760/)]
22. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. Jan 7, 2019;25(1):24-29. [doi: [10.1038/s41591-018-0316-z](https://doi.org/10.1038/s41591-018-0316-z)] [Medline: [30617335](https://pubmed.ncbi.nlm.nih.gov/30617335/)]
23. Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak*. Jul 26, 2006;6(1):30. [FREE Full text] [doi: [10.1186/1472-6947-6-30](https://doi.org/10.1186/1472-6947-6-30)] [Medline: [16872495](https://pubmed.ncbi.nlm.nih.gov/16872495/)]
24. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ*. Apr 24, 2015;350(apr24 11):h1885. [FREE Full text] [doi: [10.1136/bmj.h1885](https://doi.org/10.1136/bmj.h1885)] [Medline: [25911572](https://pubmed.ncbi.nlm.nih.gov/25911572/)]

25. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA*. Aug 24, 2011;306(8):848-855. [doi: [10.1001/jama.2011.1204](https://doi.org/10.1001/jama.2011.1204)] [Medline: [21862746](https://pubmed.ncbi.nlm.nih.gov/21862746/)]
26. Jung K, LePendu P, Iyer S, Bauer-Mehren A, Percha B, Shah N. Functional evaluation of out-of-the-box text-mining tools for data-mining tasks. *J Am Med Inform Assoc*. Jan 2015;22(1):121-131. [FREE Full text] [doi: [10.1136/amiajnl-2014-002902](https://doi.org/10.1136/amiajnl-2014-002902)] [Medline: [25336595](https://pubmed.ncbi.nlm.nih.gov/25336595/)]
27. Carrell D, Halgrim S, Tran D, Buist DSM, Chubak J, Chapman WW, et al. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. *Am J Epidemiol*. Mar 15, 2014;179(6):749-758. [FREE Full text] [doi: [10.1093/aje/kwt441](https://doi.org/10.1093/aje/kwt441)] [Medline: [24488511](https://pubmed.ncbi.nlm.nih.gov/24488511/)]
28. Huang SH, LePendu P, Iyer SV, Tai-Seale M, Carrell D, Shah NH. Toward personalizing treatment for depression: predicting diagnosis and severity. *J Am Med Inform Assoc*. Nov 01, 2014;21(6):1069-1075. [FREE Full text] [doi: [10.1136/amiajnl-2014-002733](https://doi.org/10.1136/amiajnl-2014-002733)] [Medline: [24988898](https://pubmed.ncbi.nlm.nih.gov/24988898/)]
29. Agarwal V, Podchyska T, Banda JM, Goel V, Leung TI, Minty EP, et al. Learning statistical models of phenotypes using noisy labeled training data. *J Am Med Inform Assoc*. Nov 12, 2016;23(6):1166-1173. [FREE Full text] [doi: [10.1093/jamia/ocw028](https://doi.org/10.1093/jamia/ocw028)] [Medline: [27174893](https://pubmed.ncbi.nlm.nih.gov/27174893/)]
30. Trivedi G, Dadashzadeh ER, Handzel RM, Chapman WW, Visweswaran S, Hochheiser H. Interactive NLP in clinical care: identifying incidental findings in radiology reports. *Appl Clin Inform*. Aug 04, 2019;10(4):655-669. [FREE Full text] [doi: [10.1055/s-0039-1695791](https://doi.org/10.1055/s-0039-1695791)] [Medline: [31486057](https://pubmed.ncbi.nlm.nih.gov/31486057/)]
31. Dutta S, Long WJ, Brown DF, Reisner AT. Automated detection using natural language processing of radiologists recommendations for additional imaging of incidental findings. *Ann Emerg Med*. Aug 2013;62(2):162-169. [doi: [10.1016/j.annemergmed.2013.02.001](https://doi.org/10.1016/j.annemergmed.2013.02.001)] [Medline: [23548405](https://pubmed.ncbi.nlm.nih.gov/23548405/)]
32. Trivedi G, Hong C, Dadashzadeh ER, Handzel RM, Hochheiser H, Visweswaran S. Identifying incidental findings from radiology reports of trauma patients: An evaluation of automated feature representation methods. *Int J Med Inform*. Sep 2019;129:81-87. [FREE Full text] [doi: [10.1016/j.ijmedinf.2019.05.021](https://doi.org/10.1016/j.ijmedinf.2019.05.021)] [Medline: [31445293](https://pubmed.ncbi.nlm.nih.gov/31445293/)]
33. Visser JJ, de Vries M, Kors JA. Automatic detection of actionable findings and communication mentions in radiology reports using natural language processing. *Eur Radiol*. Jun 06, 2022;32(6):3996-4002. [doi: [10.1007/s00330-021-08467-8](https://doi.org/10.1007/s00330-021-08467-8)] [Medline: [34989840](https://pubmed.ncbi.nlm.nih.gov/34989840/)]
34. Li AY, Elliot N. Natural language processing to identify ureteric stones in radiology reports. *J Med Imaging Radiat Oncol*. Jun 05, 2019;63(3):307-310. [doi: [10.1111/1754-9485.12861](https://doi.org/10.1111/1754-9485.12861)] [Medline: [30720244](https://pubmed.ncbi.nlm.nih.gov/30720244/)]
35. Dublin S, Baldwin E, Walker RL, Christensen LM, Haug PJ, Jackson ML, et al. Natural language processing to identify pneumonia from radiology reports. *Pharmacoepidemiol Drug Saf*. Aug 01, 2013;22(8):834-841. [FREE Full text] [doi: [10.1002/pds.3418](https://doi.org/10.1002/pds.3418)] [Medline: [23554109](https://pubmed.ncbi.nlm.nih.gov/23554109/)]
36. Grundmeier R, Masino A, Casper T, Dean J, Bell J, Enriquez R, et al. Identification of long bone fractures in radiology reports using natural language processing to support healthcare quality improvement. *Appl Clin Inform*. Dec 18, 2017;07(04):1051-1068. [doi: [10.4338/aci-2016-08-ra-0129](https://doi.org/10.4338/aci-2016-08-ra-0129)]
37. Kolanu N, Brown AS, Beech A, Center JR, White CP. Natural language processing of radiology reports for the identification of patients with fracture. *Arch Osteoporos*. Jan 06, 2021;16(1):6. [doi: [10.1007/s11657-020-00859-5](https://doi.org/10.1007/s11657-020-00859-5)] [Medline: [33403479](https://pubmed.ncbi.nlm.nih.gov/33403479/)]
38. Senders JT, Cho LD, Calvachi P, McNulty JJ, Ashby JL, Schulte IS, et al. Automating clinical chart review: an open-source natural language processing pipeline developed on free-text radiology reports from patients with glioblastoma. *JCO Clin Cancer Inform*. Jan 2020;4:25-34. [FREE Full text] [doi: [10.1200/CCI.19.00060](https://doi.org/10.1200/CCI.19.00060)] [Medline: [31977252](https://pubmed.ncbi.nlm.nih.gov/31977252/)]
39. Sevenster M, Buurman J, Liu P, Peters J, Chang P. Natural language processing techniques for extracting and categorizing finding measurements in narrative radiology reports. *Appl Clin Inform*. Dec 19, 2017;06(03):600-610. [doi: [10.4338/aci-2014-11-ra-0110](https://doi.org/10.4338/aci-2014-11-ra-0110)]
40. Cheng LTE, Zheng J, Savova GK, Erickson BJ. Discerning tumor status from unstructured MRI reports--completeness of information in existing reports and utility of automated natural language processing. *J Digit Imaging*. Apr 30, 2010;23(2):119-132. [FREE Full text] [doi: [10.1007/s10278-009-9215-7](https://doi.org/10.1007/s10278-009-9215-7)] [Medline: [19484309](https://pubmed.ncbi.nlm.nih.gov/19484309/)]
41. Senders JT, Karhade AV, Cote DJ, Mehrtash A, Lamba N, DiRisio A, et al. Natural language processing for automated quantification of brain metastases reported in free-text radiology reports. *JCO Clinical Cancer Informatics*. Dec 2019;(3):1-9. [doi: [10.1200/cci.18.00138](https://doi.org/10.1200/cci.18.00138)]
42. Lacson R, Prevedello LM, Andriole KP, Gill R, Lenoci-Edwards J, Roy C, et al. Fleischner Society. Factors associated with radiologists' adherence to Fleischner Society guidelines for management of pulmonary nodules. *J Am Coll Radiol*. Jul 2012;9(7):468-473. [doi: [10.1016/j.jacr.2012.03.009](https://doi.org/10.1016/j.jacr.2012.03.009)] [Medline: [22748786](https://pubmed.ncbi.nlm.nih.gov/22748786/)]
43. Duszak R, Nossal M, Schofield L, Picus D. Physician documentation deficiencies in abdominal ultrasound reports: frequency, characteristics, and financial impact. *J Am Coll Radiol*. Jun 2012;9(6):403-408. [doi: [10.1016/j.jacr.2012.01.006](https://doi.org/10.1016/j.jacr.2012.01.006)] [Medline: [22632666](https://pubmed.ncbi.nlm.nih.gov/22632666/)]
44. Petkov VI, Penberthy LT, Dahman BA, Poklepovic A, Gillam CW, McDermott JH. Automated determination of metastases in unstructured radiology reports for eligibility screening in oncology clinical trials. *Exp Biol Med (Maywood)*. Dec 09, 2013;238(12):1370-1378. [FREE Full text] [doi: [10.1177/1535370213508172](https://doi.org/10.1177/1535370213508172)] [Medline: [24108448](https://pubmed.ncbi.nlm.nih.gov/24108448/)]

45. Zhou Y, Amundson PK, Yu F, Kessler MM, Benzinger TLS, Wippold FJ. Automated classification of radiology reports to facilitate retrospective study in radiology. *J Digit Imaging*. Dec 30, 2014;27(6):730-736. [FREE Full text] [doi: [10.1007/s10278-014-9708-x](https://doi.org/10.1007/s10278-014-9708-x)] [Medline: [24874407](https://pubmed.ncbi.nlm.nih.gov/24874407/)]
46. Hripcsak G, Austin JHM, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology*. Jul 2002;224(1):157-163. [doi: [10.1148/radiol.224101118](https://doi.org/10.1148/radiol.224101118)] [Medline: [12091676](https://pubmed.ncbi.nlm.nih.gov/12091676/)]
47. Chapman BE, Lee S, Kang HP, Chapman WW. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. *J Biomed Inform*. Oct 2011;44(5):728-737. [FREE Full text] [doi: [10.1016/j.jbi.2011.03.011](https://doi.org/10.1016/j.jbi.2011.03.011)] [Medline: [21459155](https://pubmed.ncbi.nlm.nih.gov/21459155/)]
48. Chen MC, Ball RL, Yang L, Moradzadeh N, Chapman BE, Larson DB, et al. Deep learning to classify radiology free-text reports. *Radiology*. Mar 2018;286(3):845-852. [doi: [10.1148/radiol.2017171115](https://doi.org/10.1148/radiol.2017171115)] [Medline: [29135365](https://pubmed.ncbi.nlm.nih.gov/29135365/)]
49. Yu S, Kumamaru KK, George E, Dunne RM, Bedayat A, Neykov M, et al. Classification of CT pulmonary angiography reports by presence, chronicity, and location of pulmonary embolism with natural language processing. *J Biomed Inform*. Dec 2014;52:386-393. [FREE Full text] [doi: [10.1016/j.jbi.2014.08.001](https://doi.org/10.1016/j.jbi.2014.08.001)] [Medline: [25117751](https://pubmed.ncbi.nlm.nih.gov/25117751/)]
50. European Society of Radiology (ESR). Good practice for radiological reporting. Guidelines from the European Society of Radiology (ESR). *Insights Imaging*. Apr 6, 2011;2(2):93-96. [FREE Full text] [doi: [10.1007/s13244-011-0066-7](https://doi.org/10.1007/s13244-011-0066-7)] [Medline: [22347937](https://pubmed.ncbi.nlm.nih.gov/22347937/)]
51. Aideyan UO, Berbaum K, Smith WL. Influence of prior radiologic information on the interpretation of radiographic examinations. *Academic Radiology*. Mar 1995;2(3):205-208. [doi: [10.1016/s1076-6332\(05\)80165-5](https://doi.org/10.1016/s1076-6332(05)80165-5)]
52. White K, Berbaum K, Smith WL. The role of previous radiographs and reports in the interpretation of current radiographs. *Invest Radiol*. Mar 1994;29(3):263-265. [doi: [10.1097/00004424-199403000-00002](https://doi.org/10.1097/00004424-199403000-00002)] [Medline: [8175298](https://pubmed.ncbi.nlm.nih.gov/8175298/)]
53. Hunter T, Boyle R. The value of reading the previous radiology report. *AJR Am J Roentgenol*. Mar 01, 1988;150(3):697-698. [doi: [10.2214/ajr.150.3.697](https://doi.org/10.2214/ajr.150.3.697)] [Medline: [3257629](https://pubmed.ncbi.nlm.nih.gov/3257629/)]
54. Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *Radiographics*. Oct 2015;35(6):1668-1676. [doi: [10.1148/rg.2015150023](https://doi.org/10.1148/rg.2015150023)] [Medline: [26466178](https://pubmed.ncbi.nlm.nih.gov/26466178/)]
55. Kim YW, Mansfield LT. Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. *AJR Am J Roentgenol*. Mar 2014;202(3):465-470. [doi: [10.2214/AJR.13.11493](https://doi.org/10.2214/AJR.13.11493)] [Medline: [24555582](https://pubmed.ncbi.nlm.nih.gov/24555582/)]
56. Yankaskas BC, May RC, Matuszewski J, Bowling JM, Jarman MP, Schroeder BF. Effect of observing change from comparison mammograms on performance of screening mammography in a large community-based population. *Radiology*. Dec 2011;261(3):762-770. [FREE Full text] [doi: [10.1148/radiol.11110653](https://doi.org/10.1148/radiol.11110653)] [Medline: [22031709](https://pubmed.ncbi.nlm.nih.gov/22031709/)]
57. Sevenster M, Buurman J, Liu P, Peters J, Chang P. Natural language processing techniques for extracting and categorizing finding measurements in narrative radiology reports. *Appl Clin Inform*. Dec 19, 2017;06(03):600-610. [doi: [10.4338/aci-2014-11-ra-0110](https://doi.org/10.4338/aci-2014-11-ra-0110)]
58. Sevenster M, Bozeman J, Cowhy A, Trost W. A natural language processing pipeline for pairing measurements uniquely across free-text CT reports. *J Biomed Inform*. Feb 2015;53:36-48. [FREE Full text] [doi: [10.1016/j.jbi.2014.08.015](https://doi.org/10.1016/j.jbi.2014.08.015)] [Medline: [25200472](https://pubmed.ncbi.nlm.nih.gov/25200472/)]
59. Yim W, Kwan SW, Yetisgen M. Classifying tumor event attributes in radiology reports. *Journal of the Association for Information Science and Technology*. Sep 14, 2017;68(11):2662-2674. [doi: [10.1002/asi.23937](https://doi.org/10.1002/asi.23937)]
60. Bozkurt S, Alkim E, Banerjee I, Rubin DL. Automated detection of measurements and their descriptors in radiology reports using a hybrid natural language processing algorithm. *J Digit Imaging*. Aug 2019;32(4):544-553. [FREE Full text] [doi: [10.1007/s10278-019-00237-9](https://doi.org/10.1007/s10278-019-00237-9)] [Medline: [31222557](https://pubmed.ncbi.nlm.nih.gov/31222557/)]
61. Banerjee I, Bozkurt S, Caswell-Jin JL, Kurian AW, Rubin DL. Natural language processing approaches to detect the timeline of metastatic recurrence of breast cancer. *JCO Clinical Cancer Informatics*. Dec 2019;(3):1-12. [doi: [10.1200/cci.19.00034](https://doi.org/10.1200/cci.19.00034)]
62. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc*. 2004;11(5):392-402. [FREE Full text] [doi: [10.1197/jamia.M1552](https://doi.org/10.1197/jamia.M1552)] [Medline: [15187068](https://pubmed.ncbi.nlm.nih.gov/15187068/)]
63. Cox S, Ahalt SC, Balhoff J, Bizon C, Fecho K, Kebede Y, et al. Visualization environment for federated knowledge graphs: development of an interactive biomedical query language and web application interface. *JMIR Med Inform*. Nov 23, 2020;8(11):e17964. [FREE Full text] [doi: [10.2196/17964](https://doi.org/10.2196/17964)] [Medline: [33226347](https://pubmed.ncbi.nlm.nih.gov/33226347/)]
64. Sun H, Xiao J, Zhu W, He Y, Zhang S, Xu X, et al. Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: model development and performance evaluation. *JMIR Med Inform*. Jul 23, 2020;8(7):e17653. [FREE Full text] [doi: [10.2196/17653](https://doi.org/10.2196/17653)] [Medline: [32706714](https://pubmed.ncbi.nlm.nih.gov/32706714/)]
65. Schrodt J, Dudchenko A, Knaup-Gregori P, Ganzinger M. Graph-representation of patient data: a systematic literature review. *J Med Syst*. Mar 12, 2020;44(4):86. [FREE Full text] [doi: [10.1007/s10916-020-1538-4](https://doi.org/10.1007/s10916-020-1538-4)] [Medline: [32166501](https://pubmed.ncbi.nlm.nih.gov/32166501/)]
66. Ujiie S, Yada S, Wakamiya S, Aramaki E. Identification of adverse drug event-related Japanese articles: natural language processing analysis. *JMIR Med Inform*. Nov 27, 2020;8(11):e22661. [FREE Full text] [doi: [10.2196/22661](https://doi.org/10.2196/22661)] [Medline: [33245290](https://pubmed.ncbi.nlm.nih.gov/33245290/)]

67. Dai H, Lee Y, Nekkanti C, Jonnagaddala J. Family history information extraction with neural attention and an enhanced relation-side scheme: algorithm development and validation. *JMIR Med Inform.* Dec 01, 2020;8(12):e21750. [FREE Full text] [doi: [10.2196/21750](https://doi.org/10.2196/21750)] [Medline: [33258777](https://pubmed.ncbi.nlm.nih.gov/33258777/)]
68. Charniak E. Statistical Techniques for Natural Language Parsing. *AI Magazine.* 1997;18(4):33. [FREE Full text] [doi: [10.1609/aimag.v18i4.1320](https://doi.org/10.1609/aimag.v18i4.1320)]
69. Martinez AR. Part-of-speech tagging. *WIREs Comp Stat.* Sep 30, 2011;4(1):107-113. [doi: [10.1002/wics.195](https://doi.org/10.1002/wics.195)]
70. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. *J Biomed Inform.* Dec 2015;58 Suppl(Suppl):S11-S19. [FREE Full text] [doi: [10.1016/j.jbi.2015.06.007](https://doi.org/10.1016/j.jbi.2015.06.007)] [Medline: [26225918](https://pubmed.ncbi.nlm.nih.gov/26225918/)]
71. Maiya AS. ktrain: a low-code library for augmented machine learning. *Journal of Machine Learning Research.* 2022;23:1-6. [FREE Full text]
72. usb-ai / apophenator. GitHub. URL: <https://github.com/usb-ai/apophenator> [accessed 2022-12-05]
73. Lin C, Hsu C, Lou Y, Yeh S, Lee C, Su S, et al. Artificial intelligence learning semantics via external resources for classifying diagnosis codes in discharge notes. *J Med Internet Res.* Nov 06, 2017;19(11):e380. [FREE Full text] [doi: [10.2196/jmir.8344](https://doi.org/10.2196/jmir.8344)] [Medline: [29109070](https://pubmed.ncbi.nlm.nih.gov/29109070/)]
74. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* 2005;18(5-6):602-610. [doi: [10.1016/j.neunet.2005.06.042](https://doi.org/10.1016/j.neunet.2005.06.042)] [Medline: [16112549](https://pubmed.ncbi.nlm.nih.gov/16112549/)]
75. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics.* Dec 2017;5:135-146. [FREE Full text] [doi: [10.1162/tacl_a_00051](https://doi.org/10.1162/tacl_a_00051)]
76. Hastie T, Tibshirani R, Friedman J. Model Assessment and Selection. In: *The Elements of Statistical Learning.* New York, NY: Springer; 2009;219-259.
77. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging.* Aug 12, 2015;15(1):29. [FREE Full text] [doi: [10.1186/s12880-015-0068-x](https://doi.org/10.1186/s12880-015-0068-x)] [Medline: [26263899](https://pubmed.ncbi.nlm.nih.gov/26263899/)]
78. Lakhani P, Menschik ED, Goldszal AF, Murray JP, Weiner MG, Langlotz CP. Development and validation of queries using structured query language (SQL) to determine the utilization of comparison imaging in radiology reports stored on PACS. *J Digit Imaging.* Mar 1, 2006;19(1):52-68. [FREE Full text] [doi: [10.1007/s10278-005-7667-y](https://doi.org/10.1007/s10278-005-7667-y)] [Medline: [16132483](https://pubmed.ncbi.nlm.nih.gov/16132483/)]
79. Haygood TM, Mullins B, Sun J, Amini B, Bhosale P, Kang HC, et al. Consultation and citation rates for prior imaging studies and documents in radiology. *J. Med. Imag.* Jul 1, 2018;5(03):1. [doi: [10.1117/1.jmi.5.3.031409](https://doi.org/10.1117/1.jmi.5.3.031409)]
80. Berlin L. Must new radiographs be compared with all previous radiographs, or only with the most recently obtained radiographs? *AJR Am J Roentgenol.* Mar 2000;174(3):611-615. [doi: [10.2214/ajr.174.3.1740611](https://doi.org/10.2214/ajr.174.3.1740611)] [Medline: [10701597](https://pubmed.ncbi.nlm.nih.gov/10701597/)]

Abbreviations

- CT:** computed tomography
- DICOM:** Digital Imaging and Communications in Medicine
- LSTM:** long short-term memory
- MRI:** magnetic resonance imaging
- NLP:** natural language processing
- PACS:** picture archiving and communication system
- PET:** positron emission tomography

Edited by A Benis; submitted 29.06.22; peer-reviewed by I Banerjee, R Sauvayre; comments to author 21.07.22; revised version received 13.09.22; accepted 30.11.22; published 21.12.22

Please cite as:

Binsfeld Gonçalves L, Nestic I, Obradovic M, Stieltjes B, Weikert T, Bremerich J
Natural Language Processing and Graph Theory: Making Sense of Imaging Records in a Novel Representation Frame
JMIR Med Inform 2022;10(12):e40534
URL: <https://medinform.jmir.org/2022/12/e40534>
doi: [10.2196/40534](https://doi.org/10.2196/40534)
PMID: [36542426](https://pubmed.ncbi.nlm.nih.gov/36542426/)

©Laurent Binsfeld Gonçalves, Ivan Nestic, Marko Obradovic, Bram Stieltjes, Thomas Weikert, Jens Bremerich. Originally published in *JMIR Medical Informatics* (<https://medinform.jmir.org>), 21.12.2022. 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.