Automatic Classification of Tumor Response From Radiology Reports With Rule-Based Natural Language Processing Integrated Into the Clinical Oncology Workflow.

Gery Laurent, F. Craynest, M. Thobois, Nawale Hajjaji

ELECTRONIC HEALTH RECORDS

original reports

Automatic Classification of Tumor Response From Radiology Reports With Rule-Based Natural Language Processing Integrated Into the Clinical Oncology Workflow

Gery Laurent, MS[1]; Franck Craynest, BCA[1]; Maxime Thobois, BCA[1]; and Nawale Hajjaji, MD, PhD[2,3]

abstract

PURPOSE Imaging reports in oncology provide critical information about the disease evolution that should be timely shared to tailor the clinical decision making and care coordination of patients with advanced cancer. However, tumor response stays unstructured in free-text and underexploited. Natural language processing (NLP) methods can help provide this critical information into the electronic health records (EHR) in real time to assist health care workers.

METHODS A rule-based algorithm was developed using SAS tools to automatically extract and categorize tumor response within progression or no progression categories. 2,970 magnetic resonance imaging, computed tomography scan, and positron emission tomography French reports were extracted from the EHR of a large comprehensive cancer center to build a 2,637-document training set and a 603-document validation set. The model was also tested on 189 imaging reports from 46 different radiology centers. A tumor dashboard was created in the EHR using the Timeline tool of the vis.js javascript library.

RESULTS An NLP methodology was applied to create an ontology of radiographic terms defining tumor response, mapping text to five main concepts, and application decision rules on the basis of clinical practice RECIST guidelines. The model achieved an overall accuracy of 0.88 (ranging from 0.87 to 0.94), with similar performance on both progression and no progression classification. The overall accuracy was 0.82 on reports from different radiology centers. Data were visualized and organized in a dynamic tumor response timeline. This tool was deployed successfully at our institution both retrospectively and prospectively as part of an automatic pipeline to screen reports and classify tumor response in real time for all metastatic patients.

CONCLUSION Our approach provides an NLP-based framework to structure and classify tumor response from the EHR and integrate tumor response classification into the clinical oncology workflow.

JCO Clin Cancer Inform 7:e2200139. © 2023 by American Society of Clinical Oncology

ASSOCIATED CONTENT

Appendix

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on December 21, 2022 and published at ascopubs.org/journal/cci on February 13, 2023: DOI https://doi.org/10.1200/CCI.22.00139

INTRODUCTION

The adoption of an electronic health records (EHR) system and the development of clinical applications for artificial intelligence have the potential to improve the delivery of medical care.[1-3] However, health care workers may experience difficulties with EHR, particularly with data overload and data fragmentation.[4,5] Moreover, some important data for clinical decision making stay in free text, thus limiting their sharing, such as data about the symptoms, the disease, drugs used, or surgical procedures.[6] The large amount of health data generated require user-friendly approaches[7] to organize and visualize the relevant information to assist the EHR navigation of health care workers[8] and reduce safety concerns linked to fragmented displays as raised by the Committee on Patient Safety and Health Information Technology[9] of the Institute of Medicine.

In oncology, tailored clinical decisions and coordination are paramount in providing appropriate care in due time, particularly in patients with advanced disease. On top of clinical symptoms and biological data, tumor response provides critical information about the cancer disease situation. This key oncology end point drives clinical care and guides the diagnostic reasoning used to interpret symptoms and biological signs in everyday practice. The diagnostic pattern and level of emergency in front of a clinical situation could significantly change depending on the disease status, especially when the disease is progressing. Timely communication and sharing of tumor response status among the health care team are crucial to avoid diagnostic errors and improve clinical decisions. However, in routine practice, this information is unstructured and stays confined to the imaging report, and is unexploited.[10]

CONTEXT

Key Objective

To develop a natural language processing–based framework to structure tumor response from free-text radiology reports of electronic health records and integrate tumor response into the clinical oncology workflow.

Knowledge Generated

Our contributions are the extraction annotation scheme and a robust algorithm to categorize tumor response automatically from free-text imaging records. The model was developed for non-English language texts, and is effective, fast, interpretable, and shareable.

Relevance

The use of natural language processing to create or enrich a clinical dashboard with high precision is quite novel and responds to a medical need of making sense of and organizing complex information to assist health care workers in monitoring and caring for patients with advanced cancer.

Tumor response in advanced cancer disease is mostly monitored using computed tomography (CT) scans, magnetic resonance imaging (MRI), and positron emission tomography (PET) imaging techniques. Although radiology reports in oncology comply with the international RECIST classification system in four categories: complete response, partial disease, stable disease, and progression,[11] the radiologist's interpretation of tumor response is unstructured in the imaging report. Moreover, the reports offer a nuanced analysis of the disease status with a rich clinical narrative containing some variability. Tumor response is not routinely encoded into the electronic medical record because it would require manual curation and additional human effort from the health care team. Providing tumor response status in real time to the multiple actors involved in the trajectory of patients with advanced cancer may improve the quality of care. In a report by Henkel et al,[12] providing organized information, including imaging, to urologists caring for patients with prostate cancer improved physicians' effectiveness.

Natural language processing (NLP), a branch of artificial intelligence, gathers a variety of approaches ranging from statistical analyses to deep learning methods and is devoted to processing human languages by computers.[13,14] Advancements in NLP methods can help supplement the tumor response information gap. Previous reports have shown the performance of NLP to extract tumor information from imaging records in the context of cancer, on the basis of English language text.[15,16]

Our work showed that it is possible to extract, categorize, and visualize tumor response from the EHR to provide real-time information to the health care team with high accuracy. In this study, we describe a rule-based NLP method to annotate, extract, and classify tumor response from free-text radiology reports of the EHR platform of a French regional comprehensive cancer center. Our contributions are our extraction annotation scheme and a robust algorithm to categorize tumor response for further integration and use in the patient EHR.

METHODS

Study Population

This project was reviewed and approved by the institutional review board of Oscar Lambret Center. Patients participating provided a written nonopposition to the reuse of their EHR data for research purposes. The study complied with European General Data Protection Regulation. The reports were deidentified before processing (names, surnames, permanent IDs, and dates of birth were removed). Deidentification was based on selective extraction of data from the radiology report, specifically the ID of the report (different from the patients' ID) and the conclusion section, which does not contain identification data.

Patients at our institution are included in trajectories of care according to the tumor type and the stage of the disease. All patients recorded in a trajectory metastatic in the EHR of the hospital, regardless of the primary tumor type, during the year 2020 were included without additional selection criteria. As described in Figure 1, 4,414 reports were found (corresponding to 1,088 patients), but only 2,970 reports (corresponding to 925 patients) were CT scans, MRI, or PET, and had a conclusion section. Among these patients with metastatic solid tumors, the majority (39%) had breast cancer, 15% had sarcoma, 13% had head and neck cancer, 11% had lung cancer, 10% had urologic cancer, 6% had digestive cancer, and 6% brain tumors. Chemotherapy was the most frequent treatment received before the imaging report (65%), followed by targeted therapy (20%). Immunotherapy represented only 9% and radiotherapy 6% of the treatments administered.

MRI reports, CT scan reports, and PET reports were extracted from EHRs because they are conventional techniques for the monitoring of metastatic cancers. Monitoring imaging is either performed internally in our center or externally in radiology centers close to patients' places of residence. A total of 2,970 internal reports were extracted and an 80:20 split ratio strategy was applied using a simple random stratified sampling on the basis of

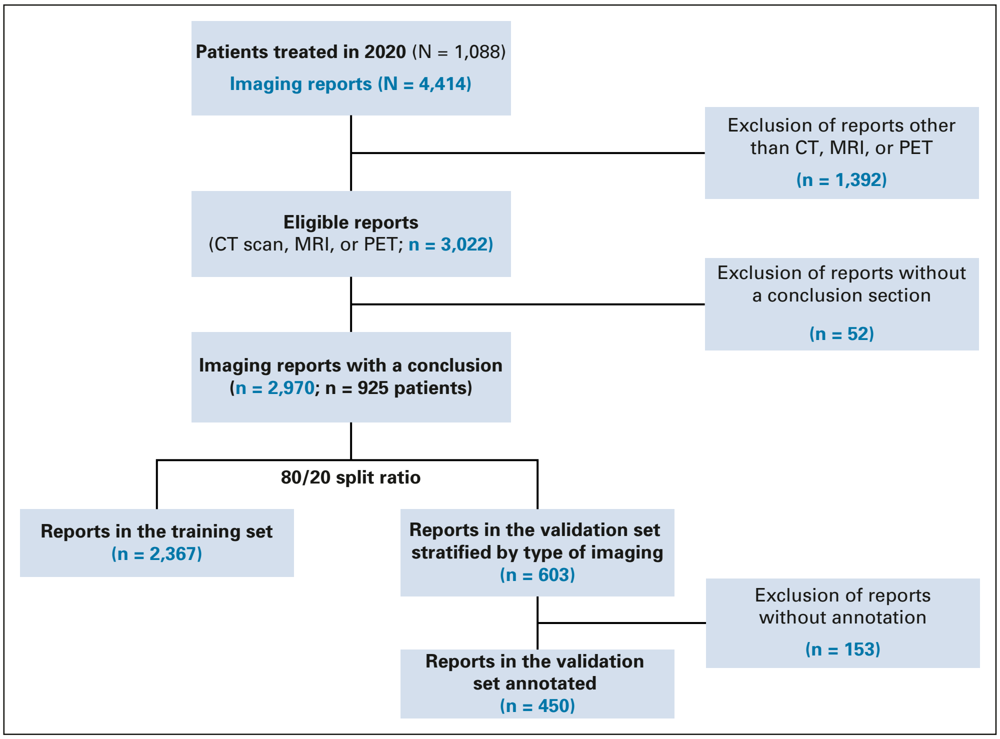

FIG 1. Flowchart depicting patients and internal imaging reports. CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography.

the document imaging type. Two thousand three hundred sixty-seven reports were used as a training set for the model, while the 603 remaining documents were kept as a validation set. In addition, 189 external imaging records were extracted to test the performance of the algorithm on reports from other French radiology centers using different reporting templates and linguistic variability.

Preprocessing

EHR reports were stored in PDF format within the hospital database. An optimal character recognition (OCR) step via WINDEV software program was performed to convert the reports from the PDF format into text format.[17] The OCR technology ABBYY FlexiCapture 12 Release 3 was used for data capture. The quality of extraction in French reported by the manufacturer ranges between 75% and 90%.[18] The radiologist template included a conclusion section indicating his interpretation of the tumor response analysis on the basis of the RECIST system.[11] Because of the difficulty to differentiate atypical tumor response to immunotherapy,[19,20] we decided to classify pseudoprogression as a progression for further interpretation by oncologists.

A conclusion section summarizing the findings is mandatory in French imaging reports.[21] For this reason and to limit the number of rules, we extracted only the conclusion section, which was found in 98.3% of the reports extracted. This also allowed for minimizing the annotation workload. Internal records follow a general template format that facilitates the extraction of the conclusion with keywords within the reports: only the part of the text after the mention of the conclusion section was kept.

Text preprocessing steps were performed for standardization and to minimize spelling mistakes: lowercasing, accent removal, and special characters removal.

SAS Visual Text Analytics Rule-Based Algorithm

The rule-based algorithm was developed using the SAS Visual Text Analytics environment used in our institution. It offers the capability to use a SAS proprietary rule-writing language called LITI (language interpretation for textual information) to define keywords and parameters for concept extraction.[22] Custom concepts using medical-specific dictionaries can be created and matches within the report can be extracted alongside rule-based relations between concept nodes.

Evaluation Metrics

To assess the model performance, accuracy, precision recall, and F1 score were calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{N};$$

$$\text{Precision} = \frac{TP}{TP + FP};$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1} - \text{Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})}$$

where TP—number of true positives; TN—number of true negatives; FP—number of false positives; FN—number of false negatives; N—total number of documents.

Dynamic Visualization Module

A dynamic browser-based visualization tumor dashboard was created in the EHR using the vis.js javascript library.[23] The timeline component of the library is easy to use to handle large amounts of dynamic data and customize their representation. The type of imaging (CT scan, MRI, and PET), tumor response for each imaging report classified by the algorithm with a color code, and the name of the therapy received were chronologically organized with this tool to build a synthetic view of the monitoring of the disease over time.

RESULTS

Concept Dictionary Building

The first iteration of a concept dictionary was built using medically approved vocabulary after consultation with a practicing medical oncologist. Conclusions from 2,367 imaging reports (MRI, CT scan, and PET) were selected as being eligible to be manually analyzed as part of the training set. This helped enrich the vocabulary list with terms and expressions more specific to the hospital EHR structure and format. More than 235 words and terms were used as a backbone for the extraction of three concepts: the presence of a tumor lesion, the location of the lesion, and tumor response (Fig 2A). Tumor response was categorized into a progression response and a no progression response, which gathered stable disease, partial response, and complete response. To refine the classification, the tumor response concept was further subdivided into a direct response concept when the dictionary allowed a direct categorization, or an indirect response concept in case the combination with the lesion or the location concepts were needed to ensure the link with the metastatic disease. Two different sets of rules were developed: whether the response was mentioned in an independent manner (direct) or linked with a lesion/location (indirect; Fig 2A).

Negation detection further refined the previously mentioned rules and contributed to the removal of false positives. The detection of a necrotic lesion and vocabulary negating metastasis (such as lack of metastasis) were specific rules added to complement this approach.

Keyword addition and rule tweaking occurred throughout the project, notably after comparison with a manually annotated test set of 670 randomly selected documents from the training set. Besides, preliminary algorithm results were analyzed with an emphasis on documents without any mention of tumor response to detect potential new missing keywords (false negative) and records with a positive prediction for both progression and no progression response to detect potential rule conflicts.

Information Extraction and Data Aggregation

Randomly selected conclusions were reviewed by two domain experts and one oncologist to analyze the wording structure and develop rule-based linked relations between concepts. Although the direct response concept led to the creation of a final independent response concept, the combined response concept is composed of a response concept tied to a lesion or a location concept (Fig 2B). A relation between a response concept and a lesion/location concept is based on concept order, the distance between words, and being part of the same sentence. A sentence example can be found in Figure 2C. The final model is composed of 239 keywords and five main sets of rules with two to five subrules each: a lesion rule, a localization rule (all localizations and brain localizations), a response rule (complete, partial, stable, progression, and lack of evolution), a negation detection rule, and a response-localization rule.

For each imaging report, a detailed list of detected responses was generated and stored. A data aggregation step on a per document basis was then developed to remove duplicate matching and evaluate the EHR global response.

The reports were classified into three categories in the final aggregated table following a priority rule on the basis of response severity: progression, no progression, or no response detection (Fig 2D). In the case of a dissociated response with both progression and no progression responses being detected within the same report, the overall response was systematically considered as being a progression response. This priority rule enabled the automatic classification of imaging records reporting multiple tumor responses in patients with two or more metastatic locations. The keywords list also includes some words or expressions mitigating response such as "non significatif" or "discret" (both classified as no progression) or a hypothetical word such as "suspect" (classified as progression).

Evaluation of the Algorithm Performance

The evaluation of the algorithm was performed using a validation set composed of 603 internal documents generated by 49 different radiologists working at our center with various levels of experience (from residents to senior radiologists) and different reporting styles and structures within the institution. One hundred fifty-three imaging records were excluded because of a lack of annotation of the tumor response such as in the case of baseline imaging or of not interpretable lesions. The 450 remaining imaging records were annotated by a medical oncologist. The performance metrics performed on this final validation set were summarized in Table 1. The model was able to classify 76.7% (n = 345) of the reports and achieved an overall accuracy of 0.88 out of the

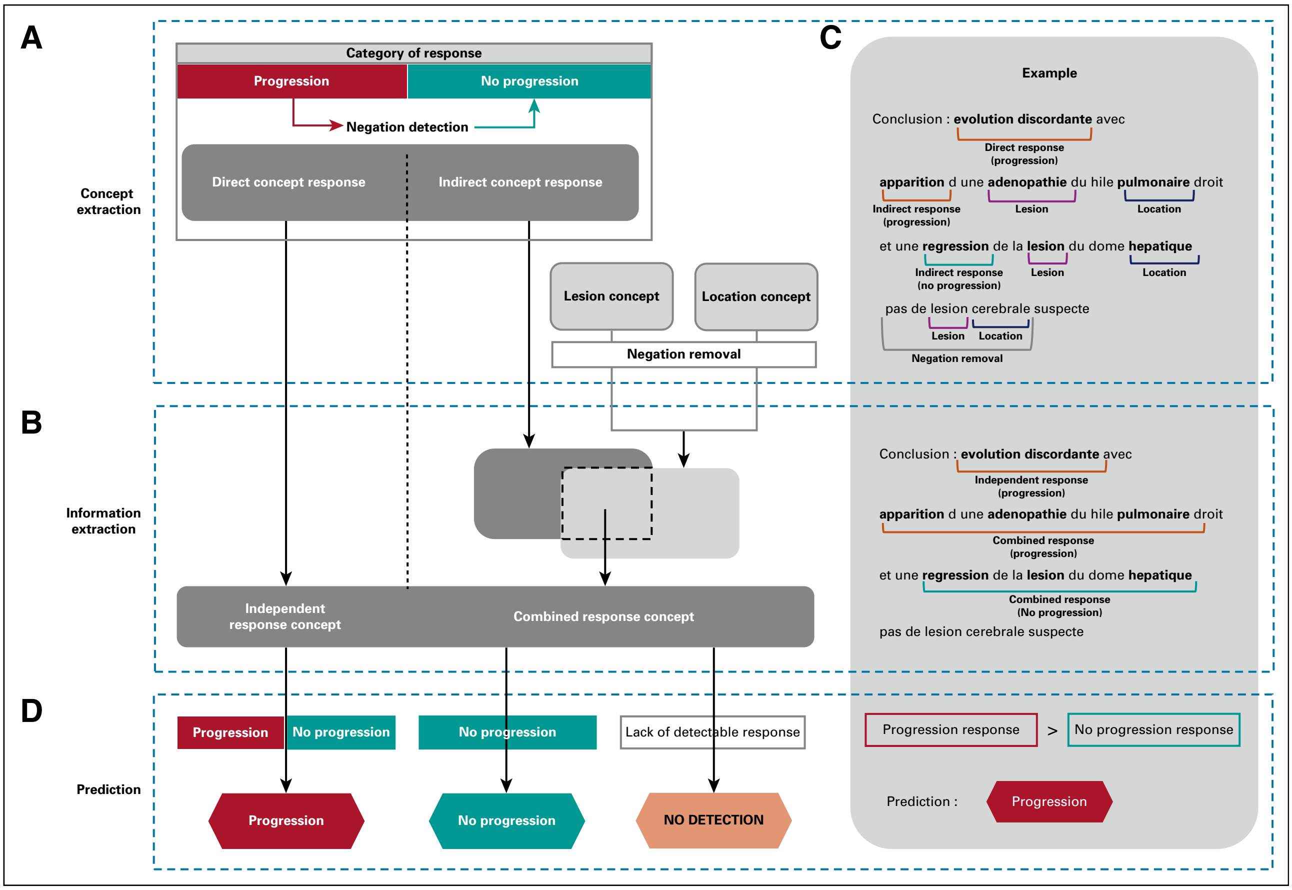

FIG 2. Algorithm process diagram with (A) the concept extraction, (B) the information extraction, (C) a sentence example, and (D) the classification prediction. Translation of the example: discordant evolution ("évolution discordante") with appearance of adenomegaly of the right pulmonary hilum ("apparition d'une adénomégalie du hile pulmonaire droit") and a regression of the liver dome lesion ("régression de la lésion du hile hépatique"), and no suspicious brain lesion ("pas de lésion cérébrale suspecte").

predicted documents. The model performed slightly better on TEP reports. Precision was similar on both progression and no progression classification.

Performance assessment on 189 external documents from 46 other radiology centers showed an overall accuracy of 0.82 (Table 1). The confusion matrix in Figures 3A and 3B details the results for internal and external reports, respectively.

When the algorithm was applied to the whole imaging report, that is, the results section and the conclusion section, the ability of the model to classify the reports raised to 91%. However, the accuracy dropped to 61.6% (Appendix Table A1).

Deployment and Data Visualization

To visualize and organize the data, a dynamic tumor response timeline was developed and integrated into the internal clinical dashboard used by oncologists and nurses at our institution. The aim was to improve the previous EHR imaging documents organized as a simple list of reports (Fig 4A).

The screenshot from the EHR of a patient treated since 2021 with chemotherapy for metastatic triple-negative programmed death ligand-1–negative and BRCA-negative breast cancer is shown as an example in Figure 4B. The tumor response categorized by the algorithm is visualized as three groups: progressive disease, lack of information, and no evolution of the disease (no progression). In this example, MRI and PET scans were used to monitor the metastatic disease over time. The classification by the algorithm showed that this patient had a metastatic disease rapidly progressing since 2021 both on PET and brain MRI, which is highly relevant to the health care team.

To make this visualization tool more friendly, the imaging report can be accessed directly by clicking on the sticks (Fig 4C), which includes reports not classified by the algorithm.

A metric regarding anticancer treatments has also been added to provide the previous and current therapeutic contexts. In this example, each progression triggered a chemotherapy change.

TABLE 1. Model Performance

Tumor Response by Type of Imaging Report	Total (No.)	True Positive (No.)	False Positive (No.)	Recall	Precision	F1 Score
Internal imaging reports						
Progression	127	102	25	0.80	0.91	0.85
MRI	23	17	6	0.74	0.94	0.83
CT scan	48	38	10	0.79	0.88	0.84
PET	56	47	9	0.84	0.92	0.88
No progression	218	200	18	0.92	0.90	0.91
MRI	37	33	4	0.89	0.87	0.88
CT scan	97	88	9	0.91	0.90	0.90
PET	84	79	5	0.94	0.91	0.92
External imaging reports						
Progression	92	77	11	0.84	0.88	0.86
MRI	26	23	3	0.89	0.89	0.89
CT scan	32	27	3	0.84	0.90	0.87
PET	34	27	5	0.79	0.84	0.82
No progression	97	78	13	0.80	0.86	0.83
MRI	36	31	3	0.86	0.91	0.89
CT scan	30	23	4	0.77	0.85	0.81
PET	31	24	6	0.77	0.80	0.79

Abbreviations: CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography.

This visualization approach has the advantage of summarizing on one screen the frequency of the monitoring, the type of imaging used, the classification of tumor response at each monitoring visit, and the anticancer therapies received.

This dynamic tumor response timeline was deployed globally at our institution for all metastatic patients as part of an automatic pipeline to screen reports daily to detect and classify tumor response. After 7 months of deployment, 2,327 imaging records have been screened retrospectively and prospectively by the algorithm: 1,023 CT scans, 759 PET, and 545 MRI records, to create 1,921 timelines in total. Prospectively, it represents on average 332 imaging records screened monthly and the creation of 40 new timelines per month (Fig 4D). This tool provides clinically relevant data organization and visualization to help health care teams in their daily care activities.

DISCUSSION

In this study, we built an automatic and reliable framework for the collection, classification, and reuse of tumor

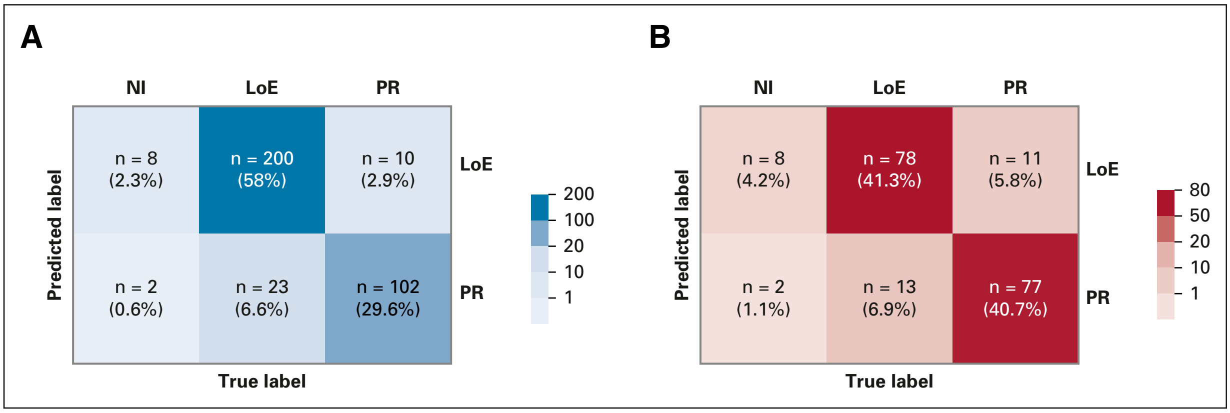

FIG 3. Confusion matrix for the set predicted by the rule-based algorithm in (A) the internal report data set (n = 354) and (B) the external report data set (n = 189). The true label refers to the records manually annotated, while the predicted label represents the classification by the algorithm. The number and percentage of records are indicated in the matrix cells. In 2.9% and 5.3% of the internal and external cases, respectively, the algorithm predicted a response (mostly lack of evolution) while the tumor response was considered not interpretable by the oncologist. LoE, lack of evolution; NI, not interpretable; PR, progression.

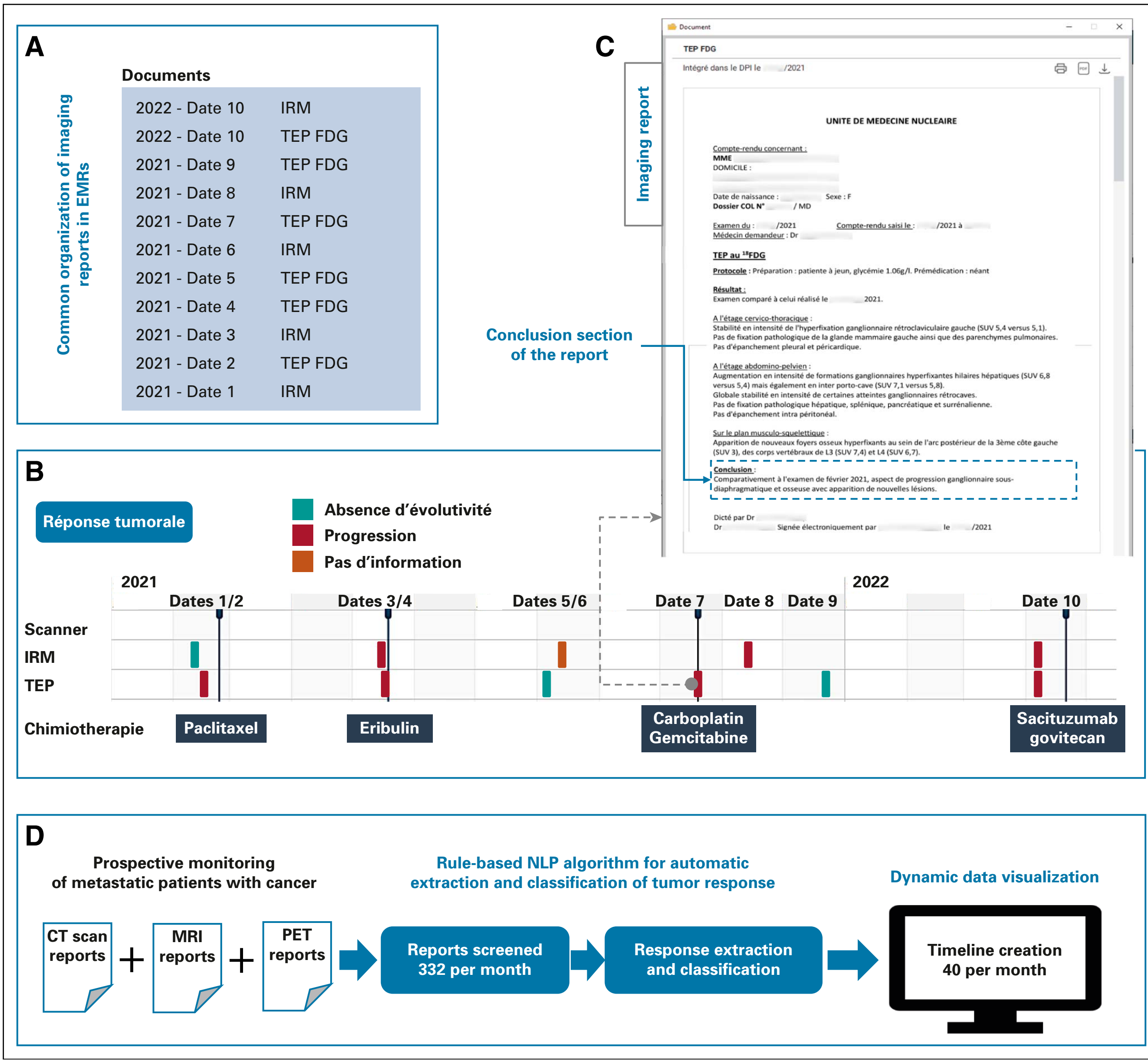

FIG 4. Snapshot of a patient's EMR at our institution to illustrate the deployment of the dynamic tumor response timeline. (A) Our center's imaging reports are organized as a chronological list of documents. (B) After the extraction from the conclusion section of the report and the classification by the rule-based algorithm, tumor response was visualized in a clinical dashboard. In this example, the patient has been treated since 2021 for triple-negative programmed death ligand-1–negative and BRCA-negative metastatic breast cancer. The different types of imaging used for disease monitoring, that is, CT scan (scanner), MRI (IRM), and PET scan (TEP), are represented with sticks. For each record, the tumor response (réponse tumorale) categorized by the algorithm is represented with a color code: progressive disease (progression) with red sticks and no evolution of the disease (absence d'évolutivité) with teal sticks, including the lack of information by the algorithm (pas d'information) with orange sticks. The lines of therapy administered are indicated on the timeline in dark blue boxes. (C) The imaging records indicated in the timeline can be accessed directly by clicking on the sticks. (D) Prospective deployment of the approach at Oscar Lambret Cancer Center. CT, computed tomography; EMR, electronic medical record; FDG, fluorodeoxyglucose; IRM, imagerie par résonance magnétique (MRI); MRI, magnetic resonance imaging; NLP, natural language processing; PET, positron emission tomography; TEP, tomographie par émission de positon (PET scan).

response for cancer patients' care. Our algorithm prototype was aimed at exploiting unstructured tumor response in real time for use by health care teams. An NLP methodology was applied with the creation of an ontology of radiographic terms defining tumor response, mapping text to concepts, and application of decision rules on the basis of clinical practice RECIST guidelines. We deployed the algorithm in the EHR of a large comprehensive cancer center both retrospectively and prospectively to create a tumor dashboard. The use of NLP to create or enrich a clinical dashboard is quite novel and responds to a medical need of making sense of and organizing complex information.

The creation of a glossary on the basis of the main imaging techniques used to assess tumor response makes this prototype particularly robust independently of the cancer type. The rule-based approach developed involved dictionary creation, context and negation checking, and heuristic algorithms to structure results. Our tumorhood evidence entity extraction was based on both anatomic vocabulary and text mentions considered positive for cancer metastases. Despite the variability of vocabulary because of the large number of radiologists working at our center, the trained algorithm achieved excellent accuracy, which was confirmed in the independent validation set. Good results were also observed on the test set from external radiologists. Only a few reports had insufficient information for tumor status classification. A greater proportion of reports could be classified when the algorithm was applied to the whole report instead of the conclusion section. However, a sharp drop in accuracy was observed because imaging reports contain rich descriptive information in the results section, including nonsignificant lesions or lesions not related to cancer. Moreover, the redundant information in the results section and the conclusion section could also induce errors.

Several NLP approaches have been developed to automate free text processing in EHR[24] and used to analyze radiology reports. Cheng et al[25] in a study showed that NLP was an accurate tool for the automated classification of tumor status from electronic databases. This is a relatively simple annotation approach that requires few involvements. Although the oncologic vocabulary to describe tumor evolution can be expressed in various nuances and formats, this type of unstructured data follow international reporting norms, which makes it relatively accessible to capture with an NLP model compared with the general human language used in the media. In this study, the choice of developing a rule-based algorithm was guided by the type of NLP tasks to perform and the possibility to use rule-based NLP independently of the programming language. The rules were tailored to build an effective, fast, interpretable, and shareable model for a health care application. The precision (positive predictive value) was an important metric of performance. Another overarching goal was to achieve a low false-negative rate, which was $< 3\%$ for internal reports and 6% on the external test set. The dynamic visualization tool provides access to the original imaging report along with its classification by the algorithm, which allows health care workers to easily check the disease status.

The use of machine learning or deep learning techniques has been shown to automatically classify tumor responses from French radiology reports.[26] A support vector machines technique, linear support vector machine, the naive Bayes method, the logistic regression, feed-forward neural network, and convolutional network yielded an accuracy ranging from 0.83 to 0.90 for binary classification (progression and no progression), and 0.82 for a four-class classification (complete response, partial response, stable disease, and progressive disease) using a logistic regression approach. Our rule-based NLP approach achieved similar results for binary classification with precision ranging from 0.87 to 0.94.

Although the number of publications regarding NLP applications to medical imaging is increasing fast, their actual adoption in clinical practice seems more limited.[27] This study showed the deployment process to integrate the automated real-time classification of tumor response in the EHR at our institution. The rapid and easy-to-use visualization tool allows sharing of the tumor response status with the health care team and multidisciplinary boards as soon as the imaging report is available. This visualization and navigation tool is intended to assist the decision process in busy workflows for cancer patients' monitoring and care. The tool does not discharge health care workers from reading the full radiology report. The choice of a rule-based NLP method was also driven by liability issues. Any artificial intelligence–based method is imperfect and could induce errors with potentially dire consequences.[28] Because of the lack of clarity on AI liability for health care workers and hospitals,[29] both workers and hospitals prefer to stay on the safe side.[28] An approach analyzing the interpretation of the radiologists instead of interpreting images may be more acceptable and less threatening for radiologists.

Implementing structured reporting in our radiology department was the alternative to our approach to structure tumor response. The experience reported by Olthof et al[30] to improve radiology reporting was successful but took 3 years and called for a strong engagement of radiologists. Implementing a change in a multidisciplinary workflow is a challenge. Radiologists must reach a consensus on a structured reporting method that fits their criteria/requirements for each type of imaging report and clinical questions to explore. This consensus should also be shared with the end user of the report (oncologists) because it may modify the information needed to make clinical decisions. The information technology infrastructure of the institution needs to be adapted, which is costly. Structured reporting should be tested to demonstrate it does not lower the clarity, accuracy, and completeness of the report. The vocabulary and the way radiologists work must be standardized. Extra resources in the radiology department will be needed for several months because the workload will increase with the practice change, with a potential rereading of the reports if the information needed by clinicians is not in the structured template. Extra resources will also be needed to check the quality of the data collected and to set up regular training for new radiologists and students. Finally, radiologists and/or oncologists should have the willingness to implement such a change. Engaging clinicians in coding is a challenge as exemplified by clinical coding[31] or sharing experience in United Kingdom.[32] The present NLP approach was developed because it was a faster and less costly method for our institution to automate the task compared with implementing structured reporting for radiology reports. Moreover, few high-level studies have investigated the benefit of implementing structured reporting in radiology to ascertain whether it improves the quality of the reports or benefits patients.[33]

NLP applications using medical text in languages other than English are increasing.[34] Our rule-based method was developed for non-English language texts with a concept extractor and classifier dedicated to oncology and is aligned with the numerical transition of the health care system in France.[35] It provides resources and a general NLP-based framework for a fast and scalable automated analysis of free-text imaging records rather than a ready-to-use tool for clinical informatics researchers and practitioners looking to apply NLP techniques and tools to clinical practice.

The single-institution data set is a limitation of this study, and further analysis with external datasets is needed to address the algorithm capabilities to be generalized. Nevertheless, our algorithm performance was high, despite a significant within-site linguistic variability, and the preliminary test on imaging reports from 46 external radiology centers showed encouraging results. Our algorithm did not integrate irRECIST/iRECIST criteria, which may limit its performance in patients treated with immunotherapy. Integrating iRECIST criteria into the model is possible without new development. This would require additional keywords and rules. This limitation was mitigated in our study by the infrequency of pseudoprogression in our cohort and by the choice to leave to oncologists the interpretation of this atypical tumor response to immunotherapy. In fact, additional factors are used to differentiate atypical tumor response to immunotherapy such as the number of metastatic lesions, the lesions kinetics, RECIST-defined criteria, or the patient clinical state. Given that our approach relies on OCR quality, the algorithm may have a lower performance if the OCR quality or if the quality of the scanned reports from other hospitals is low. Although some words and expressions mitigating response have been added to the model, the lack of standardized terminology or controlled vocabulary may affect its performance. A prospective evaluation of the model after the deployment has not been performed yet.

In conclusion, our approach provides a smart methodology to exploit and organize tumor response through automated extraction and classification from unstructured imaging reports. It can be integrated into the EHR into an interactive clinical dashboard to assist health care workers. Moreover, this initiative can also help populating the hospital clinical data repository with highly relevant structured information that can be reused with machine learning methods[16] for real-world studies, clinical research, or medicoeconomics research.

AFFILIATIONS

[1]Department of Information Systems, Oscar Lambret Cancer Center, Lille, France
[2]Department of Medical Oncology, Oscar Lambret Cancer Center, Lille, France
[3]Inserm, U1192, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), University of Lille, Lille, France

CORRESPONDING AUTHOR

Nawale Hajjaji, MD, PhD, Oscar Lambret Cancer Center, 3 rue Frederic Combemale, 59020 Lille, France; e-mail: n-hajjaji@o-lambret.fr.

AUTHOR CONTRIBUTIONS

Conception and design: Gery Laurent, Franck Craynest, Nawale Hajjaji
Collection and assembly of data: Gery Laurent, Franck Craynest, Maxime Thobois, Nawale Hajjaji
Data analysis and interpretation: Gery Laurent, Franck Craynest, Nawale Hajjaji
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Nawale Hajjaji
Consulting or Advisory Role: Daiichi Sankyo/AstraZeneca, Lilly, Pfizer, Novartis
Research Funding: Pfizer (Inst)
Travel, Accommodations, Expenses: AstraZeneca

No other potential conflicts of interest were reported.

REFERENCES

1. Pastorino R, De Vito C, Migliara G, et al: Benefits and challenges of Big Data in healthcare: An overview of the European initiatives. Eur J Public Health 29:23-27, 2019 (suppl 3)
2. Willems SM, Abeln S, Feenstra KA, et al: The potential use of big data in oncology. Oral Oncol 98:8-12, 2019
3. Ross MK, Wei W, Ohno-Machado L: "Big Data" and the electronic health record. Yearb Med Inform 23:97-104, 2014
4. Savoy A, Patel H, Murphy DR, et al: Electronic health records' support for primary care physicians' situation awareness: A metanarrative review. Hum Factors 10.1177/00187208211014300 [epub ahead of print on May 25, 2022]
5. Quinn M, Forman J, Harrod M, et al: Electronic health records, communication, and data sharing: Challenges and opportunities for improving the diagnostic process. Diagnosis (Berl) 6:241-248, 2019

6. Jensen K, Soguero-Ruiz C, Oyvind Mikalsen K, et al: Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci Rep 7:46226, 2017
7. West VL, Borland D, Hammond WE: Innovative information visualization of electronic health record data: A systematic review. J Am Med Inform Assoc 22:330-339, 2015
8. Senathirajah Y, Kaufman DR, Bakken SR: User-composable electronic health record improves efficiency of clinician data viewing for patient case appraisal: A mixed-methods study. EGEMS (Wash DC) 4:1176, 2016
9. Committee on Patient Safety and Health Information Technology; Institute of Medicine: Evaluating the Current State of Patient Safety and Health IT. Washington, DC, National Academies Press (US), 2011
10. Pai VM, Rodgers M, Conroy R, et al: Workshop on using natural language processing applications for enhancing clinical decision making: An executive summary. J Am Med Inform Assoc 21:e2-e5, 2014
11. Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 45:228-247, 2009
12. Henkel M, Horn T, Leboutte F, et al: Initial experience with AI Pathway Companion: Evaluation of dashboard-enhanced clinical decision making in prostate cancer screening. PLoS One 17:e0271183, 2022
13. Sager N, Lyman M, Bucknall C, et al: Natural language processing and the representation of clinical data. J Am Med Inform Assoc 1:142-160, 1994
14. Spasić I, Livsey J, Keane JA, et al: Text mining of cancer-related information: Review of current status and future directions. Int J Med Inform 83:605-623, 2014
15. Yim WW, Denman T, Kwan SW, et al: Tumor information extraction in radiology reports for hepatocellular carcinoma patients. AMIA Jt Summits Transl Sci Proc 2016:455-464, 2016
16. Chen P-H, Zafar H, Galperin-Aizenberg M, et al: Integrating natural language processing and machine learning algorithms to categorize oncologic response in radiology reports. J Digit Imaging 31:178-184, 2018
17. WINDEV. https://doc.windev.com/
18. ABBYY FlexiCapture 12 release 3 release notes. https://www.abbyy.com/media/23578/flexicapture-12-release-3.pdf
19. Mulkey F, Theoret MR, Keegan P, et al: Comparison of iRECIST versus RECIST V.1.1 in patients treated with an anti-PD-1 or PD-L1 antibody: Pooled FDA analysis. J Immunother Cancer 8:e000146, 2020
20. Ramon-Patino JL, Schmid S, Lau S, et al: iRECIST and atypical patterns of response to immuno-oncology drugs. J Immunother Cancer 10:e004849, 2022
21. Référentiel LABELIX de certification de service en imagerie médicale. https://labelix.fr/wp-content/uploads/2021/02/LABELIX-Referentiel-de-labellisation-en-imagerie-medicale-V-2.4-Decembre-2016.pdf
22. Jade T, Belamaric-Wilsey B, Wallis M: SAS Text Analytics for Business Applications: Concept Rules for Information Extraction Models. Cary, NC, SAS Institute, 2019
23. visjs. https://visjs.org/
24. Senders JT, Karhade AV, Cote DJ, et al: Natural language processing for automated quantification of brain metastases reported in free-text radiology reports. JCO Clin Cancer Inform 3:1-9, 2019
25. Cheng LTE, Zheng J, Savova GK, et al: Discerning tumor status from unstructured MRI reports—Completeness of information in existing reports and utility of automated natural language processing. J Digit Imaging 23:119-132, 2010
26. Goldman J-P, Mottin L, Zaghir J, et al: Classification of oncology treatment responses from French radiology reports with supervised machine learning. Stud Health Technol Inform 294:849-853, 2022
27. Casey A, Davidson E, Poon M, et al: A systematic review of natural language processing applied to radiology reports. BMC Med Inform Decis Mak 21:179, 2021
28. Price WN, Gerke S, Cohen IG: Potential liability for physicians using artificial intelligence. JAMA 322:1765-1766, 2019
29. Ethics and Governance of Artificial Intelligence for Health. https://www.who.int/publications-detail-redirect/9789240029200
30. Olthof AW, Borstlap J, Roeloffzen WW, et al: Improvement of radiology reporting in a clinical cancer network: Impact of an optimised multidisciplinary workflow. Eur Radiol 28:4274-4280, 2018
31. Brummell Z, Owen N, Walker D: Data, clinical coding and clinicians: Lost in translation. Br J Hosp Med (Lond) 80:364-365, 2019
32. Millares Martin P: Consultation analysis: Use of free text versus coded text. Health Technol 11:349-357, 2021
33. Nobel JM, van Geel K, Robben SGF: Structured reporting in radiology: A systematic review to explore its potential. Eur Radiol 32:2837-2854, 2022
34. Névéol A, Dalianis H, Velupillai S, et al: Clinical natural language processing in languages other than English: Opportunities and challenges. J Biomed Semantics 9:12, 2018
35. Plantier M, Havet N, Durand T, et al: Does adoption of electronic health records improve the quality of care management in France? Results from the French e-SI (PREPS-SIPS) study. Int J Med Inform 102:156-165, 2017

APPENDIX

TABLE A1. Confusion Matrix for the Internal Records Set When the Whole Imaging Report Is Analyzed (results section and conclusion section)

Tumor Response	True Label NI (%)	True Label LoE (%)	True Label PR (%)
Predicted label LoE	3.7	31.0	2.0
Predicted label PR	2.4	30.3	30.6

NOTE. The true label refers to the records manually annotated, while the predicted label represents the classification by the algorithm. The percentage of records is indicated in the matrix cells.

Abbreviations: LoE, lack of evolution; NI, not interpretable; PR, progression.