Automatic extraction of imaging observation and assessment categories from breast magnetic resonance imaging reports with natural language processing

Yi Liu, Li-Na Zhu, Qing Liu, Chao Han, Xiao-Dong Zhang, Xiao-Ying Wang

Department of Radiology, Peking University First Hospital, Beijing 100034, China.

Abstract

Background: Structured reports are not widely used and thus most reports exist in the form of free text. The process of data extraction by experts is time-consuming and error-prone, whereas data extraction by natural language processing (NLP) is a potential solution that could improve diagnosis efficiency and accuracy. The purpose of this study was to evaluate an NLP program that determines American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) descriptors and final assessment categories from breast magnetic resonance imaging (MRI) reports.

Methods: This cross-sectional study involved 2330 breast MRI reports in the electronic medical record from 2009 to 2017. We used 1635 reports for the creation of a revised BI-RADS MRI lexicon and synonyms lists as well as the iterative development of an NLP system. The remaining 695 reports that were not used for developing the system were used as an independent test set for the final evaluation of the NLP system. The recall and precision of an NLP algorithm to detect the revised BI-RADS MRI descriptors and BI-RADS categories from the free-text reports were evaluated against a standard reference of manual human review.

Results: There was a high level of agreement between two manual reviewers, with a κ value of 0.95. For all breast imaging reports, the NLP algorithm demonstrated a recall of 78.5% and a precision of 86.1% for correct identification of the revised BI-RADS MRI descriptors and the BI-RADS categories. NLP generated the total results in <1 s, whereas the manual reviewers averaged 3.38 and 3.23 min per report, respectively.

Conclusions: The NLP algorithm demonstrates high recall and precision for information extraction from free-text reports. This approach will help to narrow the gap between unstructured report text and structured data, which is needed in decision support and other applications.

Keywords: Breast; Natural Language Processing; Magnetic Resonance Imaging; Breast Imaging Reporting and Data System

Introduction

Most recent statistics demonstrate that breast cancer in women ranks as the second most prevalent cancer after skin cancer and the second most lethal cancer after lung cancer.[1] Breast magnetic resonance imaging (MRI) can alter the clinical management of a sizable fraction of women with breast cancer and helps in determining the optimal local treatment.[2-5] In 2007, the American Cancer Society issued a new guideline recommending annual screening for high-risk women using breast MRI.[6-10]

The MRI Breast Imaging Reporting and Data System (BI-RADS) lexicon has begun to be developed by the American College of Radiology (ACR) accordingly. The first edition of the ACR BI-RADS MRI lexicon was published along with the updated ACR BI-RADS for mammography and breast ultrasound in 2003, with the addition of new terminology to improve the description of lesions and the removal of terms that were infrequently used.[11] BI-RADS descriptors can be used by radiologists to reduce variation between radiologists and to standardize the vocabulary used in breast MRI reports.[12]

Currently, the process of data mining from free-text reports mostly depends on manual review, which is time-consuming, error-prone, and costly in studies with large sample sizes. Natural language processing (NLP) programs can convert the free-text form into structured form and have been explored in the last decade and successfully applied to identify positive findings and tumor status from radiology reports.[13-15] For breast imaging, NLP has been used to identify findings suspicious for breast cancer,[16] and determine the composition of BI-RADS breast

tissue.[17] However, few studies have investigated NLP in breast MRI data based on free-form text.

The objective of the current study was to assess the validity of our NLP program to accurately extract MRI descriptors from free-text breast MRI reports, with the ultimate goal of reducing variability in breast MRI interpretations and providing inputs for a clinical decision support system to guide the radiology practice.

Methods

Ethics approval

The study was a retrospective study in accordance with the *Declaration of Helsinki* 1975 (as revised in 2000) and was approved by the responsible Institutional Review Board of Peking University First Hospital (No. 2016[1178]) with waiver of informed consent.

Study population

From March 23, 2009 to June 1, 2017, 2330 imaging reports were consecutively selected from breast MRI study. The mean age of the patients at the time of examination was 50.9 ± 11.2 years (age range: 13–92 years). All these patients met the following inclusion criteria: the biopsy results and/or post-operative pathological results were available at the time of examination or during 3-month follow-up, examination type was identified correctly, and images were available for evaluation. For all breast imaging reports, the institutional standard practice is to assign a BI-RADS final assessment category to each imaged breast. A total of 1635 breast MRI reports containing a total of 2530 lesions were included in the study for the creation of a revised BI-RADS MRI lexicon and synonyms lists as well as the iterative development of an NLP system. The following numbers of lesions were ranked according to BI-RADS: BI-RADS 1 ($n = 10$), BI-RADS 2 ($n = 413$), BI-RADS 3 ($n = 516$), BI-RADS 4 ($n = 266$), BI-RADS 5 ($n = 572$), and BI-RADS 6 ($n = 753$). The 695 breast MRI reports containing a total of 1258 lesions that were not used for developing the system were used as an independent test set for the final evaluation of the NLP system. The following numbers of lesions were ranked according to BI-RADS: BI-RADS 1 ($n = 6$), BI-RADS 2 ($n = 322$), BI-RADS 3 ($n = 335$), BI-RADS 4 ($n = 94$), BI-RADS 5 ($n = 97$), and BI-RADS 6 ($n = 404$).

Image scanning

MRI examination was performed on 3T scanner (GE Discovery 750; GE Healthcare, Milwaukee, WI, USA and Signa Excite, GE Healthcare, Milwaukee, WI, USA) using a dedicated breast surface coil. Standard MRI protocols were used in all subjects: axial T1-weighted fast spin-echo, axial T2-weighted fast spin-echo, axial diffusion-weighted imaging echo planar imaging, axial short time inversion recovery, axial and sagittal dynamic contrast-enhanced T1-weighted fast spin-echo.

Revised BI-RADS MRI lexicon

To standardize terminology, the ACR developed the BI-RADS MRI lexicon. Considering the writing habits of our

department, we revised the ACR BI-RADS MRI lexicon and created a simple ontology structure of this terminology for our system. The revised BI-RADS MRI lexicon includes two major categories of descriptors: overall assessment and lesion assessment. Overall assessment can be separated into fibroglandular tissue and background parenchymal enhancement. Fibroglandular tissue is characterized by almost entirely fat, scattered fibroglandular tissue, heterogeneous fibroglandular tissue, and extreme fibroglandular tissue, whereas background parenchymal enhancement is characterized by symmetry (symmetrical or asymmetrical).

Lesion assessment includes the following descriptors: anatomic locations, morphology, and enhancement kinetics. Anatomic locations are regions in which abnormalities are located, such as the "upper outer quadrant of the right breast." The location and location modifiers are used to associate observations with an anatomic location in the breast. Morphologically, lesions can be separated into focus/foci, mass, and non-mass-like enhancements. A focus is a breast lesion smaller than 5 mm. A mass is a three-dimensional space-occupying lesion and is characterized by shape, size, signal, margin, internal mass enhancement characteristics, and an enhancement kinetic curve. Non-mass-like enhancement is characterized by distribution pattern, scope, signal, internal characteristics, and an enhancement kinetic curve [Table 1]. Other associated findings such as lymphadenopathy and invasion of nipple, skin, chest wall, or pectoralis muscle are reported as well.

Evaluation of the enhancement kinetic curve is based on the temporal enhancement characteristics of the lesion over time. The delayed enhancement phase refers to the signal intensity curve after 2 min or after curve starts to change, and it is described as persistent (continuously increasing enhancement), plateau (signal intensity remains constant after the initial peak), or wash-out (decrease in signal intensity).[11]

BI-RADS final assessment categories includes: 0 (need additional imaging evaluation and/or prior imaging for comparison), 1 (negative), 2 (benign finding), 3 (probably benign finding), 4 (suspicious abnormality-biopsy should be considered), 5 (highly suggestive of malignancy-appropriate action should be taken), and 6 (malignancy confirmed by biopsy-clinically feasible surgical resection).

Preprocessing

An internally developed NLP program (Smartree Clinical Information System, Beijing, China) was used to detect the revised BI-RADS MRI descriptors and BI-RADS final assessment categories for each imaged breast from text reports [Figure 1]. Each breast MRI report spanned multiple paragraphs and occasionally split at various locations [Table 2]. The steps were as previously reported.[18] Firstly, section boundaries were noted. In our text reports, each section began with an anatomical location that was followed by another anatomical location, such as "left breast" and "the upper outer quadrant of the right breast." After this step, each lesion in the reports could be distinguished by associating anatomical location with corresponding descriptors. Secondly, the

Table 1: All standard descriptors included in the breast magnetic resonance imaging reports.

Entity type	Descriptors	Detailed descriptors
Overall assessment		
Fibroglandular tissue	Almost entirely fat	
	Scattered fibroglandular tissue	
	Heterogeneous fibroglandular tissue	
	Extreme fibroglandular tissue	
Background parenchymal enhancement	Symmetrical or asymmetrical	Symmetric, Asymmetric
Lesion assessment		
Focus		
Mass	Shape	Oval, Round, Irregular
	Size	
	Signal	
	T1-weighted imaging	Low signal, Isointensity, High signal
	T2-weighted imaging	Low signal, Isointensity, High signal
	Diffusion weighted imaging	Low signal, Isointensity, High signal
	Margin	Circumscribed, Not circumscribed, Irregular, Spiculated
	Internal enhancement	Homogeneous, Heterogeneous, Rim enhancement, Dark internal septations, Enhancing internal septations, Central enhancement
	Enhancement kinetic curve	Persistent, Plateau, Wash-out
Non-mass enhancement	Distribution pattern	Focal area, Linear, Ductal, Segmental, Regional, Multiple regions, Diffuse
	Scope	
	Signal	Low signal, Isointensity, High signal
	T1-weighted imaging	Low signal, Isointensity, High signal
	T2-weighted imaging	Low signal, Isointensity, High signal
	Diffusion weighted imaging	Homogeneous, Heterogeneous, Strippled/punctate, Clumped, Reticular/dendritic
	Internal enhancement	
	Enhancement kinetic curve	Persistent, Plateau, Wash-out
Other associated findings	Lymphadenopathy, Invasion of nipple, skin, chest wall, pectoralis muscle	
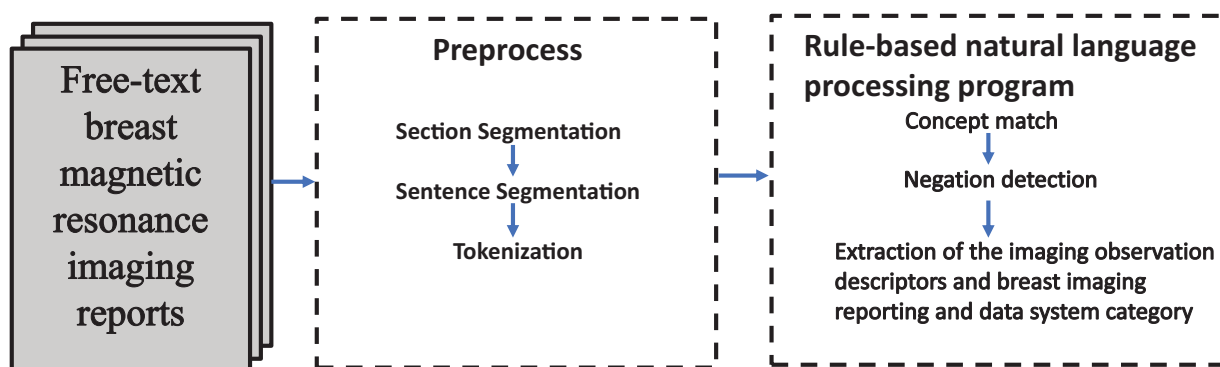
Figure 1: An overview of our rule-based natural language processing system for extracting breast reports. After the text was input, we processed the text including section segmentation, sentence segmentation, tokenization, concept match, and negation detection, then we extracted the imaging observation descriptor and breast imaging reporting and data system category.

sentence boundaries were noted, as each sentence ended with a period. Third, the NLP program preprocessed the text reports for tokenization of imaging features based on our reviewed BI-RADS, which included the BI-RADS MRI lexicon and synonyms. At the word level, we also corrected spelling mistakes and expanded abbreviations to the full form. Finally, each report was converted into a list of vocabulary flagged with its section, sentence, and feature number [Table 3].

Extracting imaging observation descriptors and BI-RADS categories

After the preprocessing steps above, the next step was the concept match. We created lists for each class in the reviewed BI-RADS lexicon that has synonyms so that this processing module could match input text to BI-RADS terms in the reviewed BI-RADS lexicon by both exact match and synonym match. Then, the negation detection

Table 2: A representative original breast magnetic resonance imaging report in our clinical work.

Imaging description

Bilateral breasts are scattered fibroglandular tissue. Background parenchymal enhancement are symmetric.

There is a 0.8 cm × 1.8cm × 1.2 cm oval mass with a not circumscribed margin in the upper outer quadrant of the left breast. Isointensity in T1-weighted imaging, high signal in T2-weighted imaging and high signal in diffusion weighted imaging. Internal enhancement is heterogeneous and kinetic curve is washout. There is a non-mass enhancement in the lower inner quadrant of the right breast. Isointensity in diffusion weighted imaging, internal enhancement is heterogeneous.

There is no invasion of skin, nipple, chest wall and pectoralis muscle in bilateral breast. There is no enlarged lymph node in bilateral armpit.

Diagnostic impression

Breast imaging reporting and data system category of the mass in upper outer quadrant of the right breast: 6.

Breast imaging reporting and data system category of the non-mass enhancement in the lower inner quadrant of the left breast: 3.

Table 3: Annotated text with different section, sentence and descriptor number.

Tokenized Text	Report Number	Section Number	Sentence Number	Descriptor Number
Fibroglandular tissue	1	1	1	1
Background parenchymal enhancement	1	1	2	1
Mass	1	2	1	1
The upper outer quadrant of the left breast	1	2	1	2
Oval	1	2	1	3
Size	1	2	1	4
Margin	1	2	1	5
T1-weighted imaging	1	2	2	6
T2-weighted imaging	1	2	2	7
Diffusion weighted imaging	1	2	2	8
Internal enhancement	1	2	3	9
Kinetic curve	1	2	4	10
Non-mass enhancement	1	3	1	1
The lower inner quadrant of the right breast	1	3	1	2
Diffusion weighted imaging	1	3	2	3
Internal enhancement	1	3	2	4
Skin	1	4	1	1
Nipple	1	4	1	2
Chest wall	1	4	1	3
Pectoralis muscle	1	4	1	4
Lymph node	1	4	2	1

step was conducted to check whether concepts in the sentence were negated. The text strings that denoted negation or uncertainty were detected and added to our revised lexicon. We used distance and direction to restrict the association between corresponding features and negation words. We imposed the eight-word rule as the default distance for negative words and only limited the direction after phrase modification after reviewing 1635 reports. The next step was to extract imaging observation descriptors, anatomic locations, and BI-RADS assessment categories for each lesion. The final results are shown in Table 4.

Establishment of a standard reference

The remaining 695 breast MRI reports were reviewed by a board-certified diagnostic radiologist with 6 years of post-graduate experience. The reviews were then evaluated by a second radiologist with 3 years of post-graduate experience. The two manual reviewers were blinded from the NLP results and extracted information independently and any discrepancies between data found were evaluated by a third independent reviewer. Thus, this data set of 695 reports comprised the gold standard for our evaluation, and comparative analyses were performed.

Statistical analysis

All statistical analysis was performed using SPSS version 20.0 (IBM Corp., Armonk, NY, USA). We used the data set of 695 manually reviewed breast MRI reports to evaluate our NLP program. We compared the information extraction frames produced by our program with the information determined by the two radiologists who reviewed the reports. In doing these assessments, we

Table 4: The final result of the original report after natural language processing extraction.

Output of the information extracted by natural language processing

Fibroglandular tissue: scattered fibroglandular tissue
Background parenchymal enhancement: symmetric
Lesion 1: Mass
    Location: upper outer quadrant of the left breast
    Shape: oval
    Signal
       T1-weighted imaging: isointensity
       T2-weighted imaging: high signal
       Diffusion weighted imaging: high signal
    Margin: not circumscribed
    Internal enhancement: heterogeneous
    Kinetic curve: washout
    Breast imaging reporting and data system category: 6
Lesion 2: non-mass enhancement
    Location: lower inner quadrant of the right breast
    Signal
       Diffusion weighted imaging: isointensity
    Internal enhancement: heterogeneous
    Breast imaging reporting and data system category: 3
Invasion of skin: no
Invasion of nipple: no
Invasion of chest wall: no
Invasion of pectoralis muscle: no
Enlarged lymph node: no

counted complete matches (all data in the information frame is correct), partial matches (at least one descriptor of imaging observation describing a lesion is incorrect or absent), and non-matches (description of a lesion is completely missed). Recall and precision were determined for the information collected by the NLP algorithm, compared against the standard reference. Recall, also referred to as sensitivity, measures the proportion of reports from which all the desired information is extracted. Precision, also referred to as positive predictive value, measures the proportion of reports with extracted data that are correct.[19]

Results

Manual extraction performance

The first reviewer detected 1258 lesions and all of them were true positive. There were 1220 completely matched cases and 38 partially matched cases among the detected lesions. The first reviewer demonstrated a recall of 97.8% and a precision of 98.1% for correct identification of the revised BI-RADS MRI descriptors and BI-RADS final assessment categories. The second reviewer detected 1260 lesions, of which 1258 were true positive and two were false positive. There were 1213 completely matched cases and 45 partially matched cases among the detected lesions. The second reviewer demonstrated a recall of 97.1% and a precision of 97.7% for correct identification of the revised BI-RADS MRI descriptors and the BI-RADS final assessment categories. There was a high level of agreement, with a κ value of 0.95.

NLP algorithm performance

Our NLP program detected 1279 lesions, of which 1258 lesions were true positive and 21 were false positive. There were 1141 completely matched cases and 107 partially matched cases among the detected lesions. For all breast imaging reports, the NLP algorithm demonstrated a recall of 78.5% and a precision of 86.1% for correct identification of the revised BI-RADS MRI descriptors and BI-RADS final assessment categories. The performance of the NLP algorithm's for individual study types is presented in Table 5.

Efficiency

We compared the time to data acquisition between the NLP program and the manual review. NLP generated the total results in <1 s, whereas the average time for each manual reviewer to extract key information was 3.38 and 3.23 min per report, respectively.

Discussion

In this study, we describe an NLP program to extract our revised BI-RADS descriptors from breast MRI reports producing an output comprising each observation with its associated characteristics and anatomic locations. This information is necessary to drive the decision support system to predict the probability of malignancy of given suspicious findings in breast MRI. The recall and precision of our system for achieving information extraction are acceptable, and if generalized to other reports, could make our approach useful for information extraction for driving decision support in breast MRI with narrative reporting.

To date, only a few studies have addressed information extraction from breast MRI reports.[17,20] Similar to our study, Sippo et al[17] used the BI-RADS observation kit, which is an open source, to detect BI-RADS final assessment categories for each imaged breast free-text report; however, they did not detect all the imaging observations. The results of the BI-RADS category with BROK were superior to ours, most likely because their NLP program was more targeted and specifically designed for category extraction. Our NLP program identified both BI-RADS descriptors and BI-RADS categories, which added a level of complexity to the information extraction process. The BI-RADS observation kit processed breast MRI reports strictly by removing measurements numbers (e.g., 3 cm × 4 cm × 5 cm) that could be mistaken for BI-RADS final assessment categories and only extracted data from the report impression. Such a process is impossible for our research, as our NLP program needs to detect all the imaging characteristics (including size, location, and so on) of each report. In another study, a different NLP system based on MedLEE extracted findings and anatomic location; however, it extracted imaging characteristics from mammography reports not breast MRI reports.[13] Our method recognized named entities pertinent to breast MRI and also the BI-RADS final assessment category. Our approach produced a concise, complete information frame

Table 5: The performance of natural language processing system for each descriptor.

Entity type	Descriptors	Recall	Precision
Fibroglandular tissue		76.1	89.1
Background parenchymal enhancement		98.6	89.2
Mass	Shape	80.2	94.0
	Size	80.5	94.3
	Signal		
	T1-weighted imaging	84.2	98.6
	T2-weighted imaging	82.3	96.4
	Diffusion weighted imaging	82.6	96.7
	Margin	70.0	82.0
	Internal enhancement	81.2	95.1
	Enhancement kinetic curve	76.6	89.7
Non-mass enhancement	Distribution Pattern	85.4	92.1
	Scope	82.1	93.2
	Signal		
	T1-weighted imaging	84.9	97.5
	T2-weighted imaging	83.1	96.8
	Diffusion weighted imaging	81.9	97.1
	Internal enhancement	82.3	94.4
	Enhancement kinetic curve	77.2	88.7
Lymphadenopathy		99.7	85.1
Invasion	Nipple	99.8	85.2
	Skin	98.3	83.9
	Chest wall	98.4	84.0
	Pectoralis muscle	99.1	85.9
Breast imaging reporting and data system category		83.9	98.2
Overall		86.1	78.5

for each lesion with its location, morphology, and enhancement kinetics in the breast MRI.

The ACR BI-RADS MRI lexicon was not directly selected as the gold standard for our work as our department has its own writing habits. Thus, we revised the ACR BI-RADS MRI lexicon and created a simple ontology structure of this terminology for our system by analyzing a large sample of breast MRI reports. For all breast imaging reports, the NLP algorithm demonstrated a recall of 78.5% and a precision of 86.1% for correct identification of the revised BI-RADS MRI descriptors and the BI-RADS final assessment categories. In our study, the error in data extraction by NLP mainly resulted from non-standard phrases and formatting. The performance of NLP for each image feature extraction was also different. In addition to "Fibroglandular tissue," "Internal mammary gland location," "Margin," and "Enhancement kinetic curve," the precision of other features was over 90% and the recall was over 80%, which was similar to the results of previous studies on information extraction of image reports.[17,21,22] Some errors in data extraction by the NLP algorithm resulted from the fact that the original image report was not standardized. In the traditional work mode, the writing of image reports depended on the personal knowledge and work habits of each radiologist as well as the regulation of report writing and habits of the department. Without unified standards, most reports did not meet the standardization requirement, which had a negative impact

on the application of NLP.[23] For example, as to the description of "lesion," a various expression such as "bulk abnormal enhancement" and "strip abnormal enhancement" often appear in the report. However, NLP technology cannot extract it as "focus/mass" or "non-mass enhancement." We created synonym lists for each class in the reviewed BI-RADS lexicon and this processing module was capable of matching input text to BI-RADS terms in the reviewed BI-RADS lexicon by both exact match and synonym match; however, the NLP program could not match all the MRI descriptors. Another reason for errors in data extraction was the absence of some descriptors, which was related to the use of free-text reports, and thus the integrity of the description content could not be guaranteed. In the future, data from structured reports should be used for NLP research and the efficiency of NLP regarding data quality should be further improved.

One of the most notable advantages of NLP is that it can generate total results in a matter of seconds compared to the hours it would take for a manual reviewer to obtain the same information. This difference is more apparent when extrapolated over many cases. The NLP program could significantly reduce the time and effort required to update a clinical database or assist clinical decisions and achieve this with a high degree of accuracy. In contrast, a manual reviewer may be prone to fatigue and reduced efficiency over time.

The literature suggests that health information technology could improve the efficiency and quality of medical care.[24] NLP is an informatics tool that can help to fulfill this promise. More specifically, NLP may improve and accelerate data extraction that drives the decision support system, which predicts the probability of malignancy of given suspicious findings in breast MRI. Our research demonstrated the utility of the combined use of an established lexicon with an NLP program. The error in data extraction by NLP mainly results from non-standard phrases and formatting in our study. Although the recall and precision of our system for achieving information extraction are acceptable, we also need to optimize our program by adding more synonyms and increasing recognition capability. Furthermore, the most effective improvement would be a greater standardization of our reports and, fortunately, there is a growing trend in radiology toward the use of a standardized lexicon for all radiology information resources. As standardized lexicons are adopted by radiology disciplines, there will be greater opportunities for the NLP program to be applied to radiology reports to extract information with a high degree of accuracy.

There were limitations in our study. Firstly, a large amount of the breast MRI reports were selected from our department and thus the segmentation rules and the reviewed BI-RADS lexicon were developed based on the writing habits of our department. Thus, the accuracy we observed with NLP may not be generalizable, and the feasibility of its extension needs to be discussed. With continued fine-tuning and validation of the program, we expect to improve its accuracy and verify its efficiency, and by increasing the amount of data from more different centers, as large amounts of data from other centers can effectively help us expand our synonyms dictionary. In addition, there are often multiple lesions in breast MRI reports and the index lesion is most crucial to clinicians in determining the management and prognosis of patients. However, our study extracted information from all the lesions in breast MRI reports not just from the index lesions. In the future, the NLP system can be used to extract the index lesion and their corresponding imaging features.

In conclusion, the main contribution of this study is the evaluation of the NLP program that extracts imaging observations and their attributes from breast MRI reports. The focus is on multiple breast lesions, and the NLP program characterizes each lesion by using their anatomic locations to differentiate lesions and their associated characteristics. We believe our approach will help to narrow the gap between unstructured text reports and structured data that is needed for decision support and other applications. In future work, we will incorporate our NLP system into a decision support system.

Conflicts of interest

None.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90. doi: 10.3322/caac.20107.
2. Tillman GF, Orel SG, Schnall MD, Schultz DJ, Tan JE, Solin LJ. Effect of breast magnetic resonance imaging on the clinical management of women with early-stage breast carcinoma. J Clin Oncol 2002;20:3413–3423. doi: 10.1200/JCO.2002.08.600.
3. Bedrosian I, Mick R, Orel SG, Schnall M, Reynolds C, Spitz FR, et al. Changes in the surgical management of patients with breast carcinoma based on preoperative magnetic resonance imaging. Cancer 2003;98:468–473. doi: 10.1002/cncr.11490.
4. Hylton N. Magnetic resonance imaging of the breast: opportunities to improve breast cancer management. J Clin Oncol 2005;23:1678–1684. doi: 10.1200/JCO.2005.12.002.
5. Braun M, Pölcher M, Schrading S, Zivanovic O, Kowalski T, Flucke U, et al. Influence of preoperative MRI on the surgical management of patients with operable breast cancer. Breast Cancer Res Treat 2008;111:179–187. doi: 10.1007/s10549-007-9767-5.
6. Schelfout K, Van Goethem M, Kersschot E, Colpaert C, Schelfhout AM, Leyman P, et al. Contrast-enhanced MR imaging of breast lesions and effect on treatment. Eur J Surg Oncol 2004;30:501–507. doi: 10.1016/j.ejso.2004.02.003.
7. Zhang Y, Fukatsu H, Naganawa S, Satake H, Sato Y, Ohiwa M, et al. The role of contrast-enhanced MR mammography for determining candidates for breast conservation surgery. Breast Cancer 2002;9:231–239. PMID:12185335.
8. Esserman L, Hylton N, Yassa L, Barclay J, Frankel S, Sickles E. Utility of magnetic resonance imaging in the management of breast cancer: evidence for improved preoperative staging. J Clin Oncol 1999;17:110–119. doi: 10.1200/JCO.1999.17.1.110.
9. Beatty JD, Porter BA. Contrast-enhanced breast magnetic resonance imaging: the surgical perspective. Am J Surg 2007;193:600–605. doi: 10.1016/j.amjsurg.2007.01.015.
10. Pediconi F, Catalano C, Padula S, Roselli A, Moriconi E, Dominelli V, et al. Contrast-enhanced magnetic resonance mammography: does it affect surgical decision-making in patients with breast cancer? Breast Cancer Res Treat 2007;106:65–74. doi: 10.1007/s10549-006-9472-9.
11. Mercado. BI-RADS update. Radiol Clin North Am 2014;52:481–487. doi: 10.1016/j.rcl.2014.02.008.
12. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, et al. The ACR BI-RADS experience: learning from history. J Am Coll Radiol 2009;6:851–860. doi: 10.1016/j.jacr.2009.07.023.
13. Sevenster M, van Ommering R, Qian Y. Automatically correlating clinical findings and body location in radiology reports using MedLEE. J Digit Imaging 2012;25:240–249. doi: 10.1007/s10278-011-9411-0.
14. Ip IK, Mortele KJ, Prevedello LM, Khorasani R. Repeat abdominal imaging examinations in a tertiary care hospital. Am J Med 2012;125:155–161. doi: 10.1016/j.amjmed.2011.03.031.
15. Cheng LT, 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 2012;23:119–132. doi: 10.1007/s10278-009-9215-7.
16. Jain NL, Friedman C. Identification of findings suspicious for breast cancer based on natural language processing of mammogram reports. Proc AMIA Annu Fall Symp 1997;829–833. PMID:9357741.
17. Sippo DA, Warden GI, Andriole KP, Lacson R, Ikuta I, Birdwell RL, et al. Automated extraction of BI-RADS final assessment categories from radiology reports with natural language processing. J Digit Imaging 2013;26:989–994. doi: 10.1007/s10278-013-9616-5.
18. Gao H, Aiello Bowles EJ, Carrel D, Buist DS. Using natural language processing to extract mammographic findings. J Biomed Inform 2015;54:77–84. PMID:25661260.
19. Pierre A, Juan P, Ilya Z. Sense induction in Folksonomies. The Twenty-second International Joint Conference on Artificial Intelligence. Barcelona, Spain. Available at: https://www.researchgate.net/publication/220813362_RULIE_Rule_Unification_for_Learning_Information_Extraction. Accessed July 16, 2011.

20. Hripcsak G, Austin JH, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. Radiology 2002;224:157–163. doi: 10.1148/radiol.2241011118.
21. Dreyer KJ, Kalra MK, Maher MM, Hurier AM, Asfaw BA, Schultz T, et al. Application of recently developed computer algorithm for automatic classification of unstructured radiology reports: validation study. Radiology 2005;234:323–329. doi: 10.1148/radiol.2341040049.
22. Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural language processing technologies in radiology research and clinical applications. Radiographics 2016;36:176–191. doi: 10.1148/radiol.2341040049.
23. Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. Radiology 2016; 279:329–343. doi: 10.1148/radiol.16142770.
24. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2016;144:742–752.