Identification of Findings Suspicious for Breast Cancer Based on Natural Language Processing of Mammogram Reports

Nilesh L. Jain, D.Sc.,[1,2] Carol Friedman, Ph.D.[1,3]
[1]Department of Medical Informatics, Columbia University, New York NY
[2]Clinical Information Services, Presbyterian Hospital in the City of New York, New York NY
[3]Department of Computer Science, Queens College, The City University of New York, New York NY

There is need for encoded data for computerized clinical decision support, but most such data are unavailable as they are in free-text reports. Natural language processing offers one alternative for encoding such data. MedLEE is a natural language processing system which is in routine use for encoding chest radiograph and mammogram reports. In this paper, we study MedLEE's ability to identify mammogram findings suspicious for breast cancer by comparing MedLEE's encoding with a logbook of all suspicious findings maintained by the mammography center. While MedLEE was able to identify all the suspicious findings, it varied in the level of granularity, particularly about the location of the suspicious finding. Thus, natural language processing is a useful technique for encoding mammogram reports in order to detect suspicious findings.

INTRODUCTION

With the increasing computerization of patient data in health care, there is growing interest in the use of automated decision support to improve the quality of health care and/or reduce the cost of health care delivered to the patients. A primary means of providing such decision support has been through the use of alerts and reminders to perform a variety of tasks including the prevention of adverse drug events, implementation of guidelines, and reduction of unnecessary diagnostic testing.[1-3] Studies have shown significant improvements in patient outcomes and/or reduction in health care costs through the routine use of such systems.[4-6]

A key requirement for automated clinical decision-support systems is the availability of patient data electronically and in a format that the systems can understand. While the computerization of patient data is solving the former requirement, the latter requirement can only be fulfilled by having the data available in a coded format. Numeric data such as laboratory results or easily coded data such as pharmacy medications are readily available in a coded format. However, a vast amount of patient data is predominantly available only as free text—radiology reports, discharge summaries, pathology reports, admission histories, and reports of physical examinations. In order to use this data for clinical decision support, it has to be obtained in a coded form. Two primary techniques are available to obtain information in free-text reports in coded form. The first is to use structured data entry to directly create coded reports.[7-9] The second is to use natural language processing to encode free-text reports.[10-12]

At Columbia-Presbyterian Medical Center (CPMC), we have been using MedLEE (Medical Language Extraction and Encoding System),[10-11] a natural language processing system to extract, structure, and encode clinical information in all chest radiograph and mammogram reports since February 1995. On average, MedLEE processes about 650 chest radiographs and mammograms (preliminary and final reports) daily, and the coded data are stored in our clinical database.[13] The coded data are used for automated decision-support using our clinical event monitor.[14] The event monitor generates alerts using Medical Logic Modules written in Arden Syntax.[15]

We have studied MedLEE extensively in the domain of chest-radiograph encoding. An evaluation to detect the presence or absence of 6 clinical conditions in 200 admission chest radiograph reports showed that MedLEE was not distinguishable from 6 internists and 6 radiologists, and was superior to 6 lay persons and 3 other computer methods.[16] Another study to determine if MedLEE can identify patients at risk of having tuberculosis based on their admission chest radiographs showed that MedLEE agreed with an infectious diseases expert on 92% of the reports.[17] Encoded chest radiograph results from MedLEE are being used to automatically detect tuberculosis patients for respiratory isolation[18] and for reporting to the Department of Health.[19]

Having achieved success with chest radiograph reports, we are now exploring the use of MedLEE for encoding mammogram reports and discharge sum-

1091—8280/97/$5.00 © 1997 AMIA, Inc.

maries. In this paper, we will focus on mammogram reports. Breast cancer is the most common form of cancer for women in the United States, affecting 32% of women diagnosed with cancer.[20] The most common use of mammograms is for breast cancer screening. Recently, studies have demonstrated the benefits of routine mammograms in terms of early detection of cancer and the subsequent reduction in mortality.[21] This has led to the recommendation of an annual mammogram as a preventive health measure for women over the age of 50.[22]

One direct consequence of the recommended annual screening is that mammography clinics are seeing large patient loads. This has led to the need for automated techniques to process the mammogram reports and screen for suspicious cases. This paper reports on a study to determine if MedLEE can identify suspicious mammogram findings that have been identified by radiologists.

METHODS

The Breast Cancer Center at Columbia-Presbyterian Medical Center performs mammograms for both screening and diagnosis. As part of its routine operation, the Breast Cancer Center maintains a logbook containing all suspicious findings found on screening and diagnostic mammograms, along with the date of the finding, name and medical record number of the patient, initials of the technician and attending radiologist, name of referring physician, and action taken. For example, a typical entry in the log book could include finding *new mass* on *3/19/97* for patient *Jane Doe* (medical record number *1234567*), initials *ABC/XYZ*, referring physician *Dr. Smith*, and action *left message for Dr. Smith*.

We extracted one year of data from the logbook. For each of those patients, we obtained the original mammogram report from our clinical data repository.[13] These reports were then encoded using MedLEE. For each report, the MedLEE encoding consists of terms classified into many categories including problems (e.g., calcification), procedures (e.g., mammogram), recommendations (e.g., biopsy), and findings (e.g., spiculated). Each term also has several modifiers including section of examination (e.g., impression), body location (e.g., breast), position (e.g., anterolateral), region (e.g., bilateral), quantity, certainty and size.

For each finding in the logbook, we determined the nearest term (and associated modifiers) in the MedLEE encoding from the impression section of the mammogram report and from the description section of the mammogram report. We then compared the logbook findings to each of these terms separately. For the comparison, all the terms within each of following sets were considered to be equivalent to each other:

- mass, cyst, density, lesion, lump, nodule
- calcification, microcalcification
- architectural distortion, architectural irregularity

Each comparison was classified into one of the four categories specified in Table 1.

Table 1. Categories and their explanation for comparing logbook findings and encoding terms

Category	Explanation
Same	both logbook finding and encoding term are the same
Encoding	encoding term has more information than the logbook finding
Logbook	logbook finding has more information than the encoding term
Different	both logbook finding and encoding term are different (though they may overlap)

Comparisons which were classified as *encoding*, *logbook*, and *different* were further analyzed to determine the cause of the disagreement.

RESULTS

The study period chosen was from March 1, 1996 to February 28, 1997. During this period, the Breast Cancer Center entered information on 160 patients into the logbook. Of these, 9 patients had to be excluded for the following reasons—for 5 patients, the medical record number was not recorded or incorrectly recorded in the logbook, and the correct medical record number could not be found by name lookup since it was a common name (the medical record number is required to obtain the mammogram report from the clinical information system); for 3 patients, MedLEE could not parse and encode the mammogram report; and for 1 patient, the mammogram report was not available in the clinical information system. For the remaining 151 patients, the logbook recorded 173 suspicious findings.

Table 2 contains the results of comparing the logbook findings with the encoded terms from the impression and description sections of the mammogram reports respectively. The level of complete agreement was very low. However, for each suspicious finding, the MedLEE encoding also contained the suspicious term. The main difference was the level of granularity of the findings, with MedLEE usually having more detail.

Table 2. Comparison of logbook findings and encoding terms from different sections of mammogram

Category	Mammogram section	
	Impression	Description
Same	24 (13.9%)	16 (9.2%)
Encoding	56 (32.4%)	61 (35.3%)
Logbook	40(23.1%)	23 (13.3%)
Different	53 (30.6%)	73 (42.2%)

The most common form of additional information in *Encoding* related to location of the logbook finding. The impression section encoding also listed laterality (e.g., left) 17 times, quadrant (e.g., upper inner) 10 times, region (e.g., subareolar) 7 times, and clock location (e.g., 8:00) 9 times. The description section encoding also listed laterality 16 times, quadrant 20 times, region 12 times, and clock location 11 times. Note that some encodings had both laterality and quadrant which were not in the logbook finding. Other sources of additional information usually were modifiers such as *asymmetric*, *clustered*, *diffuse*, *indeterminate*, *malignant*, *new*, *palpable*, *shadow* which provided more information about the suspicious finding.

The category *Logbook* included logbook findings for which there was no corresponding term in the encoding. Eleven logbook findings had no corresponding term in the impression section encoding, and 2 logbook findings had no corresponding term in the description section encoding. Another source of additional information was location of the logbook finding. The impression section encoding did not have laterality 19 times, quadrant 5 times, region twice, and clock location 8 times. The description section encoding did not have laterality 8 times, region twice, and clock location 5 times. Other sources of additional information usually were modifiers such as those listed for *Encoding*.

The category *Different* includes cases where there was some overlap as well as cases with no overlap at all. In one case, the logbook recorded *palpable red hot tender* whereas the impression section encoding had *mastitis* from the mammogram. Once again, location was the most common source of difference between the logbook findings and MedLEE encodings. The impression section encoding had laterality 15 times, quadrant 4 times, region twice, and clock location 4 times. The description section encoding had laterality 12 times, quadrant 9 times, region 10 times, and clock location 8 times. On the other hand, the impression section encoding did not have laterality 10 times, quadrant 5 times, region twice, and clock location 8 times. And the description section encoding did not have laterality 21 times, quadrant 3 times, region once, and clock location 5 times. Other sources of difference usually were modifiers such as those listed for *Encoding*.

Table 3 contains the pairing of comparison categories for the two sections of the mammogram report.

Table 3. Comparison categories for impression and description sections compared

	Description			
Impression	*Same*	*Encoding*	*Logbook*	*Different*
Same	2	10	4	8
Encoding	4	32	4	16
Logbook	7	10	8	15
Different	3	9	7	34

This shows that even though some impression section terms and modifiers match with the logbook findings, there is usually more or other information in the description section terms. If the impression section encoding contains more information, then the description section encoding is also likely to contain more information. Nothing conclusive can be stated if the impression section encoding contains less information than the logbook. Finally, if the impression section encoding is different from the logbook finding, then the description section encoding is also likely to be different from the logbook.

DISCUSSION

Researchers have long been interested in using natural language processing for encoding information in different free-text reports including radiography reports, pathology reports, and discharge summaries.[10-12] This study focuses on encoding the clinical information in mammogram reports. To our knowledge, this is the first study of natural language processing specifically for mammograms. Since mammography is a very constrained domain, we expect natural language processing to be successful within this domain.

Many published reports have also described studies conducted to evaluate the performance of natural language processing systems.[11,12,16,17,23-26] However, most of these studies are trying to validate the technology and demonstrate that natural language processing is feasible; only one study addressed a specific clinical issue.[24] Various evaluations conducted on MedLEE have shown that natural language processing is practical and feasible,[10,11,16,17,27] and MedLEE has now been put into routine clinical use. This paper represents an example of a potential clinical use of natural language processing in a routine se tting.

One major difference noted between the logbook findings and MedLEE encodings was the location modifiers. In many mammograms, location information such as laterality is usually inferred from the context that has been set up in the earlier sentences. However, MedLEE currently processes reports one sentence at a time, and does not retain context from previous sentences. This makes it difficult for MedLEE to make the inferences that humans make while reading free-text reports. An obvious enhancement to MedLEE would be to try and retain location information and use it as a modifier for the terms.

For the purposes for identifying suspicious masses, MedLEE performed well because it was able to identify all the suspicious findings, either through the impression section or the description section of the mammogram. However, as noted above, the location information was often lacking or incomplete. While this does not impact the flagging of mammograms as potentially abnormal, it does have an effect on other possible automated tasks. One such task could be the ordering of follow-up examinations such as a biopsy where precise location of the suspicious mass is required.

This study has a few limitations. It uses the mammography logbook as a reference standard. However, while there was some structure to the information in the logbook, the level of detail was different for different radiologists. This became evident when we noticed some radiologists noting the laterality as well as quadrant information of a finding, and other radiologists only noting the finding with no location information. This difference in individual styles influenced the comparisons made in this study.

This study only measures the sensitivity (or true-positive rate) of MedLEE because all the encoded cases that were examined were known to have suspicious findings. Since we did not study the encoding of a set of normal mammograms to see if they also contained any suspicious terms, we do not know the specificity (or 1 – false-positive rate) of MedLEE for detecting suspicious findings in mammograms.

In conclusion, this study demonstrates that natural language processing is a useful and practical technique for encoding mammogram reports to detect suspicious findings. It can serve as a valuable assistant in dealing with large volumes of mammograms that are being generated by the current preventive health maintenance guidelines recommending annual mammograms for women over age 50.

Acknowledgments

Supported in part by Contract N01-LM63542 and Grant R29-LM05397 from National Library of Medicine; Columbia University Center for Advanced Technology in High Performance Computing and Communication in Health Care, a New York State Center for Advanced Technology supported by the New York State Science and Technology Foundation. We would also like to thank Lyudmila Shagina of the Department of Medical Informatics for her technical assistance with running MedLEE, Suzanne J. Smith, MD, of the Breast Cancer Center for providing the logbook with the suspicious mammogram findings, and Justin Starren, MD, PhD, of the Department of Medical Informatics for all his help. More information about MedLEE can be found at http://www.cpmc.columbia.edu/MedLEE/.

References

1. Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. JAMA 1991; 266: 2847-51.
2. Overhage JM, Tierney WM, McDonald CJ. Computer reminders to implement preventive care guidelines for hospitalized patients. Arch Int Med 1996; 156: 1551-6.
3. Tierney WM, McDonald CJ, Hui SL, Martin DK. Computer predictions of abnormal test results. Effects on outpatient testing. JAMA 1988; 259: 1194-8.
4. Tierney WM, McDonald CJ, Martin DK, Rogers MP. Computerized display of past test results. Effect on outpatient testing. Ann Intern Med 1987; 107: 569-74.
5. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of

the charges for outpatient diagnostic tests. N Engl J Med 1990; 322: 1499-504.

6. Tierney WM, Miller ME, Overhage JM, McDonald CJ. Physician inpatient order writing on microcomputer workstations. Effects on resource utilization. JAMA 1993; 269: 379-83.
7. Bell DS, Pattison-Gordon E, Greenes RA. Experiments in concept modeling for radiographic image reports. J Am Med Informatics Assoc 1994; 1: 249-62.
8. Kuhn K, Zemmler T, Reichert M, Rosner D, Baumiller O, Knapp H. An integrated knowledge-based system to guide the physician during structured reporting. Meth Inform Med 1994; 33: 417-22.
9. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Meth Inform Med 1994; 33: 454-63.
10. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Informatics Assoc 1994; 1: 161-74.
11. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Nat Lang Eng 1995; 1: 83-108.
12. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Informatics Assoc 1994; 1: 142-60.
13. Johnson S, Friedman C, Cimino JJ, Clark T, Hripcsak G, Clayton PD. Conceptual data model for a central patient database. In: Clayton PD, ed. Proceedings of Fifteenth Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill; 1991: 381-5.
14. Hripcsak G, Clayton PD, Jenders RA, Cimino JJ, Johnson SB. Design of a clinical event monitor. Comput Biomed Res 1996; 29: 194-221.
15. Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden Syntax. Comput Biomed Res 1994; 27: 291-324.
16. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 1995; 122: 681-8.
17. Jain NL, Knirsch CA, Friedman C, Hripcsak G. Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. J Am Med Informatics Assoc 1996; 3 (Suppl.): 542-6.
18. Knirsch CA, Jain NL, Hripcsak G, Friedman C, Pablos-Mendez A. Isolation of suspected tuberculosis patients using manual and automated clinical protocols. Infect Cont Hosp Epidemiol 1997; in press.
19. Hripcsak G, Knirsch C, Jain NL, Pablos-Mendez A. Automated tuberculosis detection. J Am Med Informatics Assoc 1997; in press.
20. Kelsey JL, Bernstein L. Epidemiology and prevention of breast cancer. Ann Rev Pub Health 1996; 17: 47-67.
21. Smart CR, Hendrick RE, Rutledge JH III, Smith RA. Benefit of mammography screening in women age 40–49: current evidence from randomized clinical trials. Cancer 1995; 75: 1619-26.
22. Smigel K. NCI proposes new breast cancer screening guidelines. JNCI 1993; 85: 1626-8.
23. Zingmond D, Lenert LA. Monitoring free-text data using medical language processing. Comput Biomed Res 1993; 26: 467-81.
24. Haug PJ, Koehler S, Lau LM, Wang P, Rocha R, Huff SM. Experience with a mixed semantic/syntactic parser. J Am Med Informatics Assoc 1995; 2 (Suppl.): 284-8.
25. do Amaral MB, Satomura Y. Structuring medical information into a language-independent database. Med Informatics 1994; 19: 269-82.
26. Lenert LA, Tovar M. Automated linkage of free-text descriptions of patients with a practice guideline. In: Safran C, ed. Proceedings of Seventeenth Annual Symposium on Computer Applications in Medical Care. New York: McGraw-Hill; 1993: 274-8.
27. Johnson SB, Friedman C. Integrating data from natural language processing into a clinical information system. J Am Med Informatics Assoc 1996; 3 (Suppl.): 537-41.