Identification of Suspected Tuberculosis Patients based on Natural Language Processing of Chest Radiograph Reports

Nilesh L. Jain, D.Sc.,[1] Charles A. Knirsch, M.D., M.P.H.,[2,3] Carol Friedman, Ph.D.[1,4]
George Hripcsak, M.D.,[1]

[1]Department of Medical Informatics, Columbia-Presbyterian Medical Center, New York NY
[2]Division of Infectious Diseases, Columbia-Presbyterian Medical Center, New York NY
[3]Bureau of Tuberculosis Control, New York City Department of Health, New York NY
[4]Department of Computer Science, Queens College, The City University of New York, New York NY

Identification of eligible patients from electronically available patient data is a key difficulty in computerizing clinical practice guidelines because a large amount of the relevant data is stored as free text. We have been using MedLEE (Medical Language Extraction and Encoding System), a natural language processing system, to encode the clinical information in all chest radiograph and mammogram reports. This paper describes a retrospective study to determine if MedLEE can identify patients at risk for having tuberculosis (TB) based on their admission chest radiographs. Reports of 171 adult inpatients with culture-positive TB during 1992 and 1993 were manually coded (by a TB specialist) using seven terms suggestive of TB, and were also encoded by MedLEE. Using manual coding as the gold standard, MedLEE agreed on the classification of 152/171 (88.9%) reports—129/142 (90.8%) suspicious for TB and 23/29 (79.3%) not suspicious for TB; and 1072/1197 (89.6%) terms indicative of TB. Analysis showed that most of the discrepancies were caused by MedLEE not finding the location of the infiltrate. By ignoring the location of the infiltrate, the agreement became 157/171 (91.8%) reports and 946/1026 (92.2%) terms. Thus, natural language processing offers a practical alternative for using free-text reports to determine patient eligibility for computerized clinical practice guidelines.

INTRODUCTION

Automated clinical decision-support systems which generate alerts and reminders have helped in significantly improving the quality and/or reducing the cost of health care delivered to patients.[1-3] Such systems have been used to perform a wide variety of tasks including the selection of appropriate drugs, prevention of adverse drug events, and reduction of unnecessary diagnostic testing. Government health care agencies and professional organizations are taking the lead in developing clinical practice guidelines for various clinical conditions.[4] The focus of this activity is to reduce variation in clinical practice and improve the quality of health care. Various strategies are available for the effective implementation of these guidelines.[5] Among all the strategies, computerization of the practice guidelines may hold most promise.[6] Automated clinical-decision support systems are currently being developed to implement published clinical practice guidelines.[7,8]

A key difficulty in computerizing practice guidelines is the identification of eligible patients, because computerized guidelines depend on electronically available patient data to determine eligibility. However, having patient data electronically available may not be sufficient because the data may be stored as free text and automated systems cannot interpret it. Thus determination of eligibility is restricted to numerical and coded patient data. This excludes a vast amount of patient data which is predominantly available only as free text—radiology reports, discharge summaries, pathology reports, admission histories, and reports of physical examinations. These reports typically contain information which is essential in determining patient eligibility for a practice guideline. 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.[9-11] The second is to use natural language processing (NLP) to encode free-text reports.[12,13]

At Columbia-Presbyterian Medical Center (CPMC), we have been using MedLEE (**Med**ical **L**anguage **E**xtraction and **E**ncoding System),[12,14] an NLP 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.[15] The coded data are used for automated decision-support using our clinical event monitor.[16] The event monitor generates alerts using Medical Logic Modules (MLMs) written using Arden Syntax.[17] 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.[18]

This paper reports a study performed to determine if MedLEE can identify patients at risk of having tuberculosis (TB) based on their admission chest radiographs. Although MedLEE was trained for the overall domain of chest radiographs, it was not specifically trained for TB. The incidence of TB, once believed to be on the decline, is rapidly increasing in the US.[19] It remains the single largest cause of death in the world from an infectious disease.[20] In particular, the number of TB patients in New York City has more than tripled since 1978.[21,22] The Bureau of Tuberculosis Control of the New York City Department of Health has put together a manual of clinical policies and protocols for the prevention, detection, and treatment of TB.[23] Early identification of TB patients is crucial to determine which clinical policy or protocol is most appropriate for them.

METHODS

One hundred seventy six adult inpatients who had culture-positive TB during 1992 and 1993 were the subjects for the study. For each patient, we identified the chest radiograph taken at the time of admission for the episode of care during which they tested positive for tuberculosis. Admission chest radiograph reports were found for 171 patients. The remaining five patients were excluded from the study.

These chest radiograph reports were manually coded by an infectious diseases specialist (CAK–he specializes in the care of TB patients) using seven terms suggestive of TB (Table 1). Manual coding was restricted to the impression and description sections of the report. For each term, a 1 was recorded if it appeared positively in the report and 0 if it appeared negatively or did not appear in the report. If any term was coded 1 for a report, the report was considered suspicious for TB based on manual coding.

The chest radiograph reports were then coded using MedLEE. MedLEE used the text from the clinical information, impression, and description sections of the report. Encodings of the clinical information section were ignored. For each report, MedLEE produced findings (clinical terms) along with modifiers indicating body location, region, certainty, degree, change, status, descriptor, quantity, section of the chest radiograph report, and parse mode. The recognized terms came from MedLEE's own lexicon.

Table 1. Terms suggestive of TB used in the manual coding of chest radiograph reports.

Cavitation
Upper lobe infiltrate
Middle or lower lobe infiltrate
Pleural effusion
Miliary tuberculosis
Hilar adenopathy
Tuberculosis

MedLEE's output was used to code the reports for the terms in Table 1. Additional terms were used because MedLEE did not always find the lobar location of an infiltrate and MedLEE identified negative findings (Table 2). Thus MedLEE's output was coded for 14 terms. A clinical finding of density, opacity, mass, consolidation, lesion, or pneumonia was considered as an infiltrate with appropriate location. The term *infiltrate* from Table 2 includes infiltrate findings with and without lobar location. Positive terms were coded similar to manual coding. Negative terms were coded 1 if the term appeared negatively in the report, and 0 otherwise. A report was considered suspicious for TB if any term appearing in Table 1 was coded 1. By ignoring the location modifiers for infiltrate, a report was considered suspicious for TB if any of the 6 terms (5 from Table 1 and *infiltrate* from Table 2) was coded 1.

Table 2. Additional terms suggestive of TB used in coding chest radiographs based on MedLEE output.

Infiltrate
No cavitation
No infiltrate
No pleural effusion
No miliary tuberculosis
No hilar adenopathy
No tuberculosis

Three comparisons were made between manual coding and MedLEE—positive terms suggestive of TB (Table 1 and *infiltrate* from Table 2); negative terms (Table 2) only when they were present in MedLEE's output; and reports considered suspicious for TB (with and without infiltrate location). All discrepancies were analyzed to determine their origin. In all these comparisons, manual coding was assumed to be the gold standard.

RESULTS

The first two comparisons were at the level of individual terms used in manual coding and in coding MedLEE's output. MedLEE agreed on 1218/1368 (89.0%) positive terms, and agreement ranged from 130/171 (76.0%) to 170/171 (99.4%) (Table 3).

Table 3. Comparison of positive terms identified by MedLEE and manual coding.

	Agree		Disagree	
MedLEE	Yes	No	No	Yes
Manual coding	Yes	No	Yes	No
Cavitation	29	128	13	1
Upper lobe infiltrate	51	91	21	8
Mid/lower lobe infiltrate	59	71	27	14
Pleural effusion	33	130	5	3
Miliary tuberculosis	3	167	1	0
Hilar adenopathy	11	145	6	9
Tuberculosis	36	118	10	7
Infiltrate	113	33	18	7
Total	335	883	101	49
	(89.0%)		(11.0%)	

Three negative terms (no cavitation, no miliary tuberculosis, no tuberculosis) did not appear in any report. For the other three terms, agreement with manual coding was 85/95 (89.5%) and ranged from 28/38 (73.7%) to 25/25 and 32/32 (100%) (Table 4).

Table 4. Comparison of negative terms identified by MedLEE and manual coding.

	Agree	Disagree
MedLEE	Yes	Yes
Manual coding	No	Yes
No infiltrate	28	10
No pleural effusion	25	0
No hilar adenopathy	32	0
Total	85	10
	(89.5%)	(10.5%)

The circumstances of disagreement were analyzed for the discrepant term agreements. During the analysis, some errors were noted in manual coding and in the coding of MedLEE's output. Of the 160 term discrepancies, 32 were resolved by correcting the coding errors. Initially 77 reports contained no discrepant terms, and discrepant terms in another 19 reports were resolved by the correction. Thus 96/171 (56.1%) reports contained no discrepant terms. The remaining 75 reports contained the 128 term discrepancies, ranging from 1 to 4 discrepant terms per report (average–1.71 terms per report).

The primary cause of discrepancies was the inability of MedLEE to identify the lobar location of an infiltrate (26 times). The reason for this is that MedLEE was not trained to have such a fine granularity for body locations. Also, 9 times the location was inferred as upper and lower lobe because the report stated that infiltrates were throughout the lung. For 21 discrepancies, MedLEE was unable to parse a part of a sentence or did not recognize words which were not in its lexicon or were misspelt. On 8 occasions, MedLEE interpreted *no definite infiltrate* as absence of an infiltrate which was not considered right. For 8 discrepancies, MedLEE was unable to use the right context from the rest of the sentence, or from previous sentences. In 4 cases, MedLEE failed to create a finding containing the term. In 4 other cases, MedLEE identified an adenopathy but could not find its location. Twice, TB was not mentioned directly but as PPD positive or known infectious disease which MedLEE did not identify as TB. On the other hand, manual coding was also unable to identify the location of an infiltrate or an infiltrate itself 32 times. In 14 other instances, manual coding missed terms that MedLEE identified successfully.

The third comparison was at the level of the entire report about suspicion for TB. MedLEE agreed with manual coding on 152/171 (88.9%) reports. When using infiltrate without lobar location, the agreement became 157/171 (91.8%) (Table 5). MedLEE correctly identified patients to be suspicious for TB not identified by manual coding. The overall sensitivity of MedLEE was 135/171 (78.9%) with infiltrate location and 146/171 (85.4%) without infiltrate location, while manual coding was 142/171 (83.0%).

Table 5. Number of reports considered suspicious for TB by manual coding and MedLEE.

Suspicious for TB	Manual coding	MedLEE	
		Infiltrate with location	Infiltrate w/o location
Yes	142	129 (90.8%)	137 (96.5%)
No	29	23 (79.3%)	20 (69.0%)
Total	171	152 (88.9%)	157 (91.8%)

Of the 19 discrepancies, MedLEE was unable to find infiltrate (or its location) 9 times, failed to create a finding twice, incorrectly identified terms twice, and could not identify a term once due to misspelling. Manual coding was unable to identify hilar adenopathy thrice and infiltrate twice. Five discrepancies were resolved if the infiltrate location was ignored.

DISCUSSION

Researchers have long been interested in using NLP for encoding information in different free-text reports including radiography reports, pathology reports, and discharge summaries.[12,13] This study focuses on encoding the clinical information in chest radiograph reports. Such encoded information has been used to identify reports that describe new or expanding neoplasms for monitoring and follow-up of these patients.[24] More commonly, it is used to increase the amount of electronic patient data that can be used by computers for decision support.[14,25]

Many published reports have also described studies conducted to evaluate the performance of NLP systems.[13,14,18,24,26] However, most of these studies are trying to validate the technology and demonstrate that NLP is feasible; only one study addressed a specific clinical issue.[24] Various evaluations conducted on MedLEE have shown that NLP is practical and feasible,[12,14,18] and MedLEE has now been put into routine clinical use.

This paper reports one of the earliest clinical studies of an NLP system, rather than an experimental study conducted under laboratory conditions. While the study was retrospective because we had more information on that patient population, MedLEE is being used routinely to identify patients suspected of having TB. An agreement rate of 89% on positive terms surpassed our expectations because MedLEE has been trained for the domain of chest radiographs in general, and not specifically for TB. One possible reason for the high agreement rate is that the study focuses on few selected clinical terms, and it does not report the agreement rate of other terms.

One of MedLEE's weaknesses demonstrated by this study was its inability to determine the location of lung infiltrates, even though such information was present in the report. This is because MedLEE has not been specifically trained to look for this. In fact, when location information was ignored, the agreement of the infiltrate terms went from 272/342 (79.5%) to 146/171 (85.4%). By fine-tuning MedLEE for these findings, it may be possible to achieve a better agreement rate.

The motivation of the study was to demonstrate that NLP can be used to encode free-text reports in order to determine patient eligibility for clinical practice guidelines. This was attempted in one other system where admission reports of patients with coronary artery disease and chronic stable angina were encoded to determine if they were eligible for coronary-artery bypass grafting surgery.[27] In our study, 90.8% of the patients with suspected TB were correctly identified by MedLEE. Manually, the same task would require trained personnel to read and interpret every chest radiograph report, and determine if the patient is eligible for TB guidelines. This is a difficult task because we have 250 final chest radiograph reports produced daily, but is necessary due to the high incidence of TB in our patient population. Thus NLP offers a practical alternative to determine patient eligibility for computerized practice guidelines.

This study has couple of limitations. The results may be biased because the manual coding of one physician was being used as the gold standard. We have previously shown that physicians disagree among themselves about the presence of clinical conditions in a report.[18] We also noted that our gold standard was not infallible and contained errors. Different results may have been obtained if the gold standard manual coding was done by a group of physicians. A simple rule was used to identify eligible patients—presence of at least one positive term suggestive of TB. Since the study subjects were limited to TB patients, it gave us an accurate estimate of the sensitivity of the rule in TB patients. However we do not know the specificity of the rule because it was not tested on chest radiographs from a non-TB population. We suspect that the specificity will not be as high because the selected terms (Table 1) occur often in abnormal chest radiograph reports where the abnormality may not be related to TB.

In conclusion, this study shows that MedLEE agrees very well (nearly 90%) with manual coding performed by a human expert on chest radiograph reports of TB patients. Thus NLP can be used in order to encode free-text reports to determine patient eligibility for clinical practice guidelines.

Acknowledgments

Supported by Grant 36-40-94065 from Telecommunications and Information Infrastructure Assistance Program, National Telecommunications and Information Administration, US Department of Commerce; Grants T15-LM07079, R29-LM05627, and 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.

References

1. 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.
2. 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.
3. Tierney WM, Miller ME, Overhage JM, McDonald CJ. Physician inpatient order writing on microcomputer workstations. Effects on resource utilization. JAMA 1993; 269: 379-83.
4. Institute of Medicine. Guidelines for Clinical Practice: From Development to Use. Washington DC: National Academy Press; 1992.
5. Conroy M, Shannon W. Clinical guidelines: their implementation in general practice. Br J Gen Pract 1995; 45: 371-5.
6. Elson RB, Connelly DP. Computerized patient records in primary care. Their role in mediating guideline-driven physician behavior change. Arch Fam Med 1995; 4: 698-705.
7. Safran C, Rind DM, Davis RB, et al. Guidelines for management of HIV infection with computer-based patient's record. Lancet 1995; 346: 341-6.
8. Tierney WM, Overhage JM, Takesue BY, et al. Computerizing guidelines to improve care and patient outcomes: the example of heart failure. J Am Med Informatics Assoc 1995; 2: 316-22.
9. Bell DS, Pattison-Gordon E, Greenes RA. Experiments in concept modeling for radiographic image reports. J Am Med Informatics Assoc 1994; 1: 249-62.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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 15th SCAMC. New York: McGraw-Hill; 1991: 381-5.
16. Hripcsak G, Clayton PD, Jenders RA, Cimino JJ, Johnson SB. Design of a clinical event monitor. Comp Biomed Res 1996; 29: 194-221.
17. Hripcsak G. Writing Arden Syntax medical logic modules. Comp Biol Med 1994; 24: 331-63.
18. 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.
19. Barnes PF, Barrows SA. Tuberculosis in the 1990s. Ann Intern Med 1993; 119: 400-10.
20. Bloom BR, Murray CJ. Tuberculosis: commentary on a reemergent killer. Science 1992; 257: 1055-64.
21. Sepkowitz KA, Telzak EE, Recalde S, Armstrong D. Trends in the susceptibility of tuberculosis in New York City, 1987-1991. New York City Area Tuberculosis Working Group. Clin Infect Dis 1994; 18: 755-9.
22. Frieden TR, Fujiwara PI, Washko RM, Hamburg MA. Tuberculosis in New York City—turning the tide. N Engl J Med 1995; 333: 229-33.
23. Fujiwara PI, Harris HW, Dorsinville M, Frieden TR. Tuberculosis Clinical Policies and Protocols. New York: New York City Department of Health; 1994.
24. Zingmond D, Lenert LA. Monitoring free-text data using medical language processing. Comp Biomed Res 1993; 26: 467-81.
25. 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.
26. do Amaral MB, Satomura Y. Structuring medical information into a language-independent database. Med Informatics 1994; 19: 269-82.
27. Lenert LA, Tovar M. Automated linkage of free-text descriptions of patients with a practice guideline. In: Safran C, ed. Proceedings of 17th SCAMC. New York: McGraw-Hill; 1993: 274-8.