Automated classification of limb fractures from free-text radiology reports using a clinician-informed gazetteer methodology

Amol Wagholikar[1], Guido Zuccon[1], Anthony Nguyen[1],
Kevin Chu[2], Shane Martin[2], Kim Lai[2], Jaimi Greenslade[2]

1. The Australian e-Health Research Centre, Brisbane, CSIRO

2. Department of Emergency Medicine, Royal Brisbane & Women's Hospital, Brisbane, Queensland Health

RESEARCH

Please cite this paper as: Wagholikar A, Zuccon G, Nguyen A, Chu K, Martin S, Lai K, Greenslade J. Automated classification of limb fractures from free-text radiology reports using a clinician-informed gazetteer methodology. AMJ 2013, 6, 5, 301-307. http://doi.org/10.21767/AMJ.2013.1651

Corresponding Author:
Name: Dr Amol Wagholikar
Address: The Australian e-Health Research Centre (AEHRC) Level 5, UQ Health Sciences Building, Royal Brisbane and Women's Hospital, Herston, Queensland, Australia, 4029
Email: amol.wagholikar@csiro.au

Abstract

Background

Timely diagnosis and reporting of patient symptoms in hospital emergency departments (ED) is a critical component of health services delivery. However, due to dispersed information resources and a vast amount of manual processing of unstructured information, accurate point-of-care diagnosis is often difficult.

Aims

The aim of this research is to report initial experimental evaluation of a clinician-informed automated method for the issue of initial misdiagnoses associated with delayed receipt of unstructured radiology reports.

Method

A method was developed that resembles clinical reasoning for identifying limb abnormalities. The method consists of a gazetteer of keywords related to radiological findings; the method classifies an X-ray report as abnormal if it contains evidence contained in the gazetteer. A set of 99 narrative reports of radiological findings was sourced from a tertiary hospital. Reports were manually assessed by two clinicians and discrepancies were validated by a third expert ED clinician; the final manual classification generated by the expert ED clinician was used as ground truth to empirically evaluate the approach.

Results

The automated method that attempts to individuate limb abnormalities by searching for keywords expressed by clinicians achieved an F-measure of 0.80 and an accuracy of 0.80.

Conclusion

While the automated clinician-driven method achieved promising performances, a number of avenues for improvement were identified using advanced natural language processing (NLP) and machine learning techniques.

Key Words

Limb fractures, emergency department, radiology reports, classification, rule-based method, machine learning.

What this study adds:

1. This study reports the evaluation of a clinician-driven rule-based gazetteer method for classifying radiological evidence.
2. The outcome of our study delineates avenues for the improvement and use of a clinician-driven rule-based classifier in the context of free-text radiology report classifications.

Background

The analysis of X-rays is an essential step in the diagnostic work-up of many conditions, including fractures, in injured Emergency Department (ED) patients. X-rays are initially interpreted by the treating ED doctor, and if necessary patients are appropriately treated. X-rays are eventually reported by the specialist in radiology and these findings are relayed to the treating doctor in a formal written report. The ED, however, may not receive the report until after the patient has been discharged home. This is not an uncommon event because the reporting did not occur in real-time. Diagnosis delays can occur between discharge and report receipt for subtle fractures missed during presentation.

The review of X-ray reports is a necessary practice to ensure fractures and other conditions identified by the radiologist were not missed by the treating doctor. The review requires reading of a free-text report. Large "batches" of X-rays are reviewed often days after the patient's ED presentation. This is a labour intensive process that adds to the diagnostic delay. The process may be streamlined if it can be automated with clinical text processing solutions. These solutions will minimise delays in diagnosis and prevent complications arising from diagnostic errors. [1-3]

This paper investigates the use of an automated gazetteer approach where keywords that may suggest the presence or absence of an abnormality were provided by expert ED clinicians. Rule-based methods are commonly used in Artificial Intelligence.[4-6] Studies have shown that gazetteer rule-based methods can be applied for identifying clinical conditions from radiology reports such as acute cholecystitis, acute pulmonary embolism and other conditions.[7] The purpose of these methods is to simulate human reasoning for any given information processing task to achieve full or partial automation. Gazetteer-based (or keyword-based) classifiers are one of three common types of classifiers, others being statistical and linguistic classifiers.[8]

In this paper, an application of a gazetteer rule-based method for limb fracture identification is reported. The paper describes related works that justify the importance of this research and discusses a brief review of state-of-the-art computational approaches in this domain. The methods section describes the development of the clinician-informed gazetteer approach including data collection, ground truth development and implementation. The results section shows outcomes of the approach. Finally, we discuss initial results and explain avenues for improving the applicability of the rule-based method for automatic classification of limb fractures. We also discuss the application of statistical machine learning based classification (an alternative class of classifiers according to the taxonomy in)[8] to the problem of identifying limb abnormalities.

Related work

Previous studies that focused on the problem of identification of subtle limb fractures during the diagnosis of ED patients showed that 2.1% of all fractures were not identified during initial presentation to the Emergency Department.[9] A similar study about radiological evidence for fracture reported that 1.5% of all X-rays had abnormalities that were not identified in the Emergency Department records.[10] Further research also reported that 5% and 2% of the X-rays of the hand/fingers and ankle/foot, respectively, from a paediatric Emergency Department had fractures missed by the treating ED doctor.[11] These small incidences might have significant impact on overall patient health as these missed fractures can develop into more complex conditions. Timely recognition of fractures is therefore important.

There have been efforts to automatically detect fractures and other abnormalities from free-text radiology reports. De Bruijn et al. (2006) reported an algorithm based on support vector machine (SVM) was able to identify acute wrist fractures in free-text radiology reports with an overall F-measure of 0.91.[12] Thomas et al. (2005) reported a method that used a text search algorithm for classifying radiology reports into the categories "fracture", "normal" and "neither normal nor fracture". [13] Their approach is similar to that described in this paper, however, their text search algorithm was refined *iteratively,* while the system investigated here was directly developed using a clinician-driven gazetteer without further modifications. While Thomas et al. showed high sensitivities and specificities across the classes, their findings might not scale beyond the data they considered, because of the iterative development process. It is therefore difficult to assess the generalisability of their system.

Methods

Data Collection

De-identified free-text descriptions of patients' limb X-rays reported by radiologists were extracted from a tertiary hospital's picture archiving and communication system (PACS). Ethics approval was granted by the Human Research Ethics Committee at Queensland Health to use this data. The data included 99 free-text X-ray reports of ED patients. The average length of free-text reports was about 52 words. There were in total 930 unique words in the vocabulary. Some reports were semi-structured, with section headings such as "History", "Clinical Details", "Findings", appearing in the text.

Ground Truth Development

One ED visiting medical officer and one ED Registrar were engaged as assessors to manually classify patient findings. Findings were assigned to either one of the following three classes: (1) "Normal", no fractures or dislocations were found, (2) "Abnormal", X-ray examination exposed a reportable abnormality (e.g., a fracture, dislocation, displacement) that required follow-up, and (3) "Unsure", indicating that the clinician was unsure whether to classify the report as either normal or abnormal. To gather ground truth labels about the data, an in-house annotation tool was developed. This tool allowed the assessors to manually annotate the free-text reports, allowing classification into one of three target categories. Figure 1 shows a screenshot of the annotator tool used by the assessors during their manual annotation process. The highlight within the report indicated the span of evidence that the clinician highlighted in support of their decision. The two assessors initially agreed on the annotations of 77/99 reports and disagreed on the remaining 22 reports. The disagreed reports were resolved and validated by a senior staff specialist in Emergency Medicine, who acted as a third assessor. The senior staff specialist independently applied own expertise in an unbiased manner to resolve the disagreements between the two assessors. The disagreements were resolved without any influence from the two assessors. Twenty of the 22 reports with disagreements were annotated as normal by one assessor and abnormal by the other. The main cause of this disagreement was the difference between the subjective judgments of the two assessors. These 20 reports were related to scheduled or unscheduled visits to review previously known abnormal cases. The third assessor conveyed that these reports should be treated as abnormal cases. The disagreement about the remaining two reports was due to true judgment errors made by one of the two initial assessors. The classification of reports validated by the ED staff specialist was used as the final ground truth for comparative analysis and for evaluating the rule-based classifier.

Figure 1: Manual annotation of radiology reports

.

Rule-base classifier

A set of keywords was extracted from the criteria for the assessment of X-ray reports as documented by the clinicians prior to the ground truth annotation task. The keywords were used to identify either the presence or absence of abnormalities such as fractures in a radiology report. A rule-base was then developed around the appearance of the collected keywords among the free-text of the radiology reports. Specifically, the rules specified that if a keyword related to the presence of abnormalities was found in the text, then the report should be classified as *"Abnormal"*. Keywords that negate the presence of abnormalities were also identified in the text; if found, then the report should be classified as *"Normal"*. If no such keywords were found, then the system would default to a "Normal" classification. Note that the set of keywords was developed in the context of radiology reports that involve limb structures and may not contain other keywords that indicate abnormalities in other contexts. Table 1 reports the keywords associated with the classification task.

Classifier implementation

The steps taken to implement the clinician-based gazetteer approach are presented here. The free-text of radiology reports was first pre-processed to remove punctuation and to reduce it to a normalized form. Case folding was implemented by reducing all letters within the text to lower case. Regular expressions were formed to individuate keywords belonging to the rule-base within the text. A script was developed to remove special characters (such as ‘?’,’\’,’\\’,’-‘,” “, “,”, ‘(‘ ,’)’) from the report text.

Table 1: Keywords used for building the rule-base

Keywords	Suggested Classification
no + fracture	*Normal*
old + fracture	*Abnormal*
fracture	*Abnormal*
x ray + follow up	*Abnormal*
dislocation	*Abnormal*
FB	*Abnormal*
osteomyelitis	*Abnormal*
osteoly	*Abnormal*
displacement	*Abnormal*
intraarticular extension	*Abnormal*
foreign body	*Abnormal*
articular effusion	*Abnormal*
avulsion	*Abnormal*
septic arthritis	*Abnormal*
subluxation	*Abnormal*
osteotomy	*Abnormal*
callus	*Abnormal*

Clinical advisors provided a list of keywords that would identify normal and abnormal findings in limb X-ray reports; keywords are reported in Table 1. Resembling the reasoning of the clinicians, the system was developed such that keywords were prioritised in order to identify negated "Abnormal" cases first (i.e. "Normal"). The other rules were fired sequentially to identify "Abnormal" cases by processing the text of the report. The main reason to prioritise the rules in this sequence was driven by the clinician's needs to focus only on the "Abnormal" cases. The clinicians had specified the main bottleneck in the real-time reporting was a large volume of "normal" cases that need to be identified manually.

The implementation of regular expressions considered word boundaries, where appropriate. For example, to individuate the keyword "fb" (e.g., fractured bone), the Java regular expression "\\bfb\\b" was used; this surrounded the token "fb" with word boundaries. Word boundaries were not applied to the beginning and end of all keywords. For example, the regular expression "\\bfracture" did not use the trailing word boundary so that both the occurrence of "fracture" and "fractures" could be captured. Non-consideration of word boundaries at the end of the keyword is a rudimentary implementation of word stemming. Hyphens were substituted by a space at pre-processing time. As a result, regular expressions that capture keywords that may be spelt with or without hyphens were developed for this purpose. For example, "x ray" and "follow up" were encoded as the following regular expressions "\\bx[\\s]*ray\\b" and "\\bfollow[\\s]*up\\b".

To identify the occurrence of keywords formed by two non-consecutive tokens in the text, two regular expressions (one for each keyword token) were evaluated concurrently. For example, to identify occurrences of "x ray + follow up", two regular expressions were formed: one for "x ray" ("\\bx[\\s]*ray\\b"), the other for "follow up" ("\\bfollow[\\s]*up\\b"). Note that in these cases, the two tokens that form the keyword can appear at any point of the report.

A modification to the non-consecutive tokens to preserve word ordering was also implemented to identify occurrences of "no + fracture" to capture, for example, the text spans "no fracture" and "no evidence of fracture" (e.g., "\\bno\\b[a-z\\s]*\\bfractrure"). However, the occurrence of the first token may actually not refer to the occurrence of the second token. For example, the token "no" may refer to a token other than "fracture" due to the inability to restrict the scope of the application of the negation. In effect, the scope of search for the keywords is constrained to a paragraph, as a newline character will be a terminating token. This is a limitation of the current implementation of the method and could be resolved by considering the scope of search between the two keywords, or more generally, co-reference information extraction algorithms. Note that "no + fracture" was the only keyword used to identify normal cases. Furthermore, this keyword contained a single token (e.g., "no") to identify negations. While a wider list of negation tokens and expressions need to be considered, this is beyond the scope of this initial investigation. This improvement will be achieved in future developments.

The classifier was implemented such that if the text of the reports matches one of the keywords of Table 1, then the associated rule was applied to classify the report. Rules are implemented as "if-else" statements; if a regular expression produces a match for a report, then the report is classified according to the corresponding rule. Rules are applied sequentially according to the order of their keywords in Table 1. The ordering aimed to mimic the classification logic of the clinicians as well as place keywords with higher confidence in their classification higher up in the list. If no rule is matched, then the report is classified as being "normal"; this is recorded as a "no rule fired" event in the implementation.

Results

Results obtained by the gazetteer rule-based approach on the dataset containing 99 radiology reports are reported in Table 2. Classification results were evaluated in terms of F-measure and accuracy (see Table 2). The number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) instances were also reported.

Table 2: Classification results obtained on a set of 99 free-text radiology reports by the clinican-driven gazetteer method

Method	*F-measure*	*Accuracy*	*TP*	*TN*	*FP*	*FN*
Rule-based	0.80	0.80	39	40	11	9

Discussion

Figure 2 reports the frequency distribution of the rules used for the classifications of the reports. A large number of rules developed from the clinician guidelines were not used in the classification process over the evaluation dataset. This was because the reports containing keywords associated with the unused rules were not present in the reports, or because rules with higher priorities triggered the classification of the reports. The rules "no + fracture" and "fracture" accounted for the majority of the classifications. Table 3 reports the breakdown of false negative and false positive errors for each rule, along with the percentage of errors with respect to the times each rule was used. The number of errors caused by each rule is also reported in Figure 2. Overall, the rule-based system classified 50 reports as "Abnormal". Of these, 11 reports were assessed as normal cases and thus constitute false positive classifications. The statistics reported in Table 3 suggest that false positives were mainly due to the use of the rules "subluxation" and "fracture" (63.6% of all FP cases). A manual revision of the false positive cases generated by the "subluxation" rule revealed that the reports mentioned the presence of subluxation, although the reports were classified as "Normal" in the ground truth.

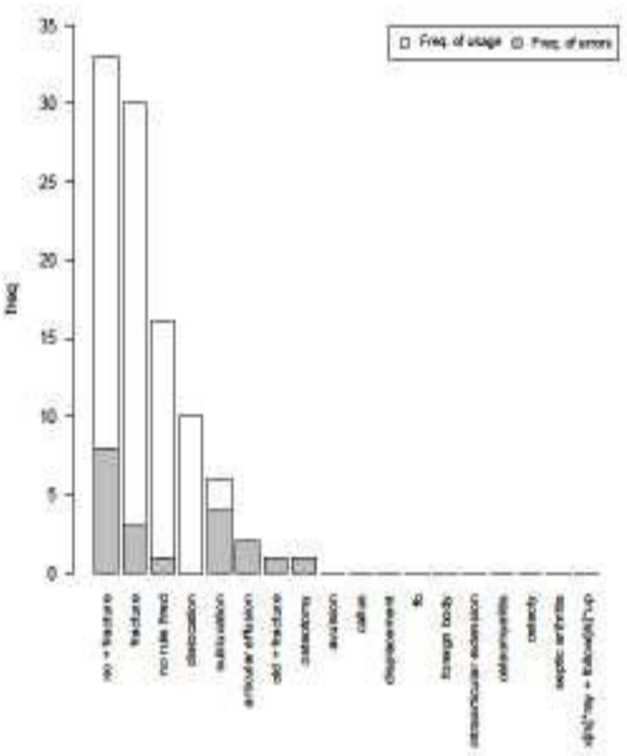
Table 3: Rule breakdown for false positive and false negative cases.

Rule	*FP(% error)*	*FN(% error)*
no + fracture	-	8 (24.24%)
subluxation	4(66.67%)	-
articular effusion	2(100%)	-
fracture	3(10%)	-
osteotomy	1(100%)	-
old+fracture	1(100%)	-
No rule fired	-	1(62.5%)
Total	11	9

In review of these reports, the clinicians agreed that these subluxation cases were likely to be a part of a chronic degenerative change, required no follow up from the Emergency Department, and thus were assessed as normal cases. Although these cases are prone to errors or disagreement due to subjective interpretation, rules could be improved by identification and exclusion from notifying cases related to degenerative changes. An analysis of the false positive cases obtained from the rule "fracture" suggested that, while the word fracture was present in these reports, it was not used to notify the presence of an abnormality, but rather to indicate the reason for the X-ray (e.g., "? Fracture"), or to state the absence of fractures. Other notable false positive cases were those produced by the "articular effusion" rule (two false positives). A manual review of such reports identified that they contained no fracture or dislocation; however the reports identified articular effusions. Effusions can follow trauma, infection or inflammation and should be diagnosed clinically. However, the clinicians agreed that these cases required clinical correlation to determine their significance. Specifically, these cases were not considered "Abnormal" because they were judged as *not significant* enough to require follow up.

Of the 49 reports classified "Normal", 9 of them were false negatives (i.e. referring to cases with abnormalities). Eight of the nine false negative cases were produced by the "no + fracture" rule. This was because limited constraints were applied to the keyword "no" in the rule "no + fracture" so as to test whether the negation referred to the presence of a fracture or to other entities in the report. It was not tested if this keyword occurred within a limited distance from the word "fracture". Thus, reports containing sentences such as "no previous imaging" or "no films", and the word "fracture" were erroneously considered as reporting abnormalities.

Figure 2: Distribution of rules used for classifying the radiology reports; they grey shades refer to the number of incorrect classification produced by each rule.



Comparison with machine learning classification approaches

Classification systems based on the explicit encoding of rules, as presented in this article, are complementary to machine learning algorithms, which instead implicitly learn classification patterns from the occurrence of features in positive and negative examples.

Zuccon et al. (2013) have evaluated a number of machine learning algorithms for the identification of abnormalities from free-text radiology reports.[14] Specifically, the Naïve Bayes and Support Vector Machine classifiers were tested along with features such as tokens, punctuation, token stems, token negations, token stem bi-grams, token stem tri-grams, as well as higher order semantic features (e.g., SNOMED CT medical concepts). Features were extracted from free-text radiology reports using the Medtex system.[15] They evaluated the machine learning approaches on the same dataset on which we reported. In the following, we provide a retrospective discussion of the findings. The classifiers were trained and evaluated using 10-fold cross validation (e.g., in each iteration, 90% of reports were used for training and the remaining 10% were used for testing). The best F-measure was achieved by the Naïve Bayes classifier (F-measure=0.92) when using stemmed token bigrams, negation features and SNOMED CT concepts related to morphological abnormalities and disorders. The comparative analysis of results obtained by the rule-based approach and the Naïve Bayes classifier is shown in Table 4.

Table 4: Classification results obtained on a set of 99 free-text radiology reports by the clinician-informed gazetteer rule-based approach and a Naive Bayes classifier[14]

Method	*F-measure*	*Accuracy*	*TP*	*TN*	*FP*	*FN*
Rule-based	0.80	0.80	39	40	11	9
Naive Bayes	0.92	0.92	44	47	4	4

While the empirical effectiveness of the gazetteer rule-based approach investigated in this paper is inferior to that obtained by the Naïve Bayes classifier evaluated on the same dataset, it is noteworthy that the machine learning approach used 90% of the dataset to train its classification models compared to the use of no development set for the rule-based classifier. In addition, even though the results of machine learning based classifiers show high effectiveness, their applicability in clinical settings may be limited. In fact, machine learning methods are data–driven and, as a result, the model lacks generalisability if the training sample is not a representative selection of the problem domain. Machine learning approaches need to be retrained on new corpora and tasks. Collating training data to build new classifier models can be a timely and labour intensive process. Furthermore, a direct mechanism to trace the evidence behind a machine learner's classification decision may not be possible. This largely limits their effectiveness as a clinical decision support tool. These issues provide the motivation for the investigation of rule-based methods which have the ability to model expert knowledge as easily implementable rules.

The gazetteer approach offers a number of advantages intrinsic of the rule-based methodology, which are lacking in the machine learning approach. Specifically, the approach investigated here allows the traceability of the automatic classification decisions as the rules were driven by the specific keywords deemed important by the clinicians. Such a rule-based methodology can be easily integrated in the workflow where clinicians can specify and update their own rules.

Although the investigated keyword rule-based approach is simplistic and obtains only suboptimal performance, it does show promise as advanced Natural Language Processing techniques such as those adopted in Medtex[16] can be used to improve classification performances. More keywords can also be learned using computational linguistic methods, such as the Basilisk bootstrapping algorithm.[17] The use of a gazetteer approach allows clinicians to define their own set of keywords used for classification, resembling part of the current manual processing and reasoning carried out for finding misdiagnosis in X-ray reports.

Advanced NLP methods include normalisation of text such as stemming and mapping to clinical terminology concepts (e.g., SNOMED CT), identification and application of negation terms to concepts, and the use of the clinical terminology semantics to exploit the relationships between concepts to allow for more complex inference and reasoning. These techniques can be integrated with the proposed clinician-informed keyword based approach to enhance the reliability and real-life usability of rule-based systems. The potential benefit of the gazetteer method is its adaptation and portability to other clinical domains. In fact, this approach can be easily adopted as clinicians can specify keywords as per the underlying clinical domain. Our approach simplifies the manual tasks of classification and improves efficiency in the clinical practice by a semi-automated process.

Conclusion

This work has described an initial investigation of a clinician-informed rule-based method for automatic classification of limb fractures from radiology reports. We described a gazetteer approach where keywords were derived from classification criteria provided by clinicians. The rule-based classification method achieved promising results. The investigated method has the potential to improve healthcare workflow by alleviating the tedious manual process associated with the revision of ED radiology reports. As future work, the research will aim to improve the simple keyword approach with more advanced clinical text processing techniques to complement the described rule-based classification method. The use of text normalisation techniques such as stemming and clinical terminology mapping will be investigated. In addition, we plan to investigate the robustness of the method across a larger dataset from multiple hospitals. The possible integration of the gazetteer-based method in real-life workflow of hospital emergency departments will also be considered and evaluated in future research.

References

1. James MR, Bracegirdle A, Yates DW. X-ray reporting in accident and emergency departments – an area for improvements in efficiency. Arch Emerg Med. 1991 Dec; 8(4):266-70.
2. Saab M, Stuart J, Randall P, Southworth S. X-ray reporting in accident and emergency departments – reducing errors. Eur J Emerg Med. 1997 Dec; 4(4):213-6.
3. Siegel E, Groleau G, Reiner B, Stair T. Computerized follow-up of discrepancies in image interpretation between emergency and radiology departments. J Digit Imaging. 1998 Aug; 11(3 Suppl 1):18-20.
4. Long WJ, Fraser H, Naimi S. Reasoning requirements for diagnosis of heart disease. Artif Intell Med. 1997 May; 10(1):5-24.
5. Harleen K, Siri Krishan W. Empirical Study on Applications of Data Mining Techniques in Healthcare, Journal of Computer Science. 2006; 2(2):194-200, pp.1549-3636.
6. Subhash Chandra N, Uppalaiah B, Charles Babu G, Naresh Kumar K, Raja Shekar P. General Approach to Classification: Various Methods can be used to classify X-ray images. IJCSET. 2012 Mar; 2(3):933-937.
7. Lakhani P, Kim W, Langlotz CP. Automated detection of critical results in radiology reports. J Digit Imaging. 2012 Feb; 25(1):30-6.
8. Conway M, Dowling J, Chapman W. Using chief complaints for syndronic surveillance: A review of chief complaint based classifiers in North America. J Biomed Inform. 2013 Apr 17. pii: S1532-0464(13)00046-4
9. Cameron MG. Missed fractures in the emergency department. Emerg Med (Fremantle). 1994; 6:3.
10. Sprivulis P, Frazer A. Same-day x-ray reporting is not needed in well supervised emergency departments. Emerg Med (Fremantle). 2001; 13:194–197.
11. Mounts J, Clingenpeel J, Byers E, McGuire, Kireeva Y. Most frequently missed fractures in the emergency department. Clin Pediatr (Phila).2011; 50:183–186.
12. De Bruijn B, Cranney A, O'Donnell S, Martin JD, Forster AJ. Identifying wrist fracture patients with high accuracy by automatic categorization of x-ray reports. Journal of the American Medical Informatics Association (JAMIA).2006; 13(6):696–698.
13. Thomas BJ, Ouellette H, Halpern EF, Rosenthal DI. Automated computer-assisted categorization of radiology reports. AJR Am J Roentgenol. 2005 Feb; 184(2):687-90.
14. Zuccon G, Wagholikar A, Nguyen A, Chu, K, Martin S, Greenslade J. Automatic Classification of Free-Text Radiology Reports to Identify Limb Fractures using Machine Learning Algorithms and SNOMED CT. Proceedings AMIA Clinical Research Informatics Symposium, San Francisco, USA, March 2013, 300-304.
15. Nguyen AN, Lawley MJ, Hansen DP, Colquist S. A simple pipeline application for identifying and negating SNOMED clinical terminology in free text. Proceedings of the Health Informatics Conference; Canberra, Australia. August 2009, 188–93.
16. Nguyen A, Lawley M, Hansen D, Bowman R, Clarke B, Duhig E, Colquist, S. Symbolic Rule-based Classification of Lung Cancer Stages from Free-Text Pathology Reports. Journal of the American Medical Informatics Association (JAMIA). 2010; 17(4):440-445.
17. Thelen M, Riloff E. A bootstrapping method for learning semantic lexicons using extraction pattern contexts, Proceedings of the 2002 Conference on Empirical Methods in NLP, Philadelphia, PA (2002), 214-221.

ACKNOWLEDGEMENTS

The authors are thankful to Bevan Koopman for feedback on this paper. This research is a collaborative work between Department of Emergency Medicine at Royal Brisbane & Women's Hospital and the Australian e-Health Research Centre. The authors sincerely thank the funding institution and collaborators for their support.

PEER REVIEW

Not commissioned. Externally peer reviewed

FUNDING

This research was supported by the Queensland Emergency Medicine Research Foundation Grant, EMPJ-11-158-Chu-Radiology.

ETHICS COMMITTEE APPROVAL

Ethics approval was granted by the Human Research Ethics Committee at Queensland Health to use de-identified radiology reports.