Clinician-Driven Automated Classification of Limb Fractures from Free-Text Radiology Reports

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
{amol.wagholikar,guido.zuccon,anthony.nguyen}@csiro.au
[2] Department of Emergency Medicine, RBWH, Brisbane, Queensland Health
{kevin_chu,shane_martin}@health.qld.gov.au
{kim_lai,jaimi_greenslade}@health.qld.gov.au

Abstract. The aim of this research is to report initial experimental results and evaluation of a clinician-driven automated method that can address the issue of misdiagnosis from unstructured radiology reports. Timely diagnosis and reporting of patient symptoms in hospital emergency departments (ED) is a critical component of health services delivery. However, due to disperse information resources and vast amounts of manual processing of unstructured information, a point-of-care accurate diagnosis is often difficult. A rule-based method that considers the occurrence of clinician specified keywords related to radiological findings was developed to identify limb abnormalities, such as fractures. A dataset containing 99 narrative reports of radiological findings was sourced from a tertiary hospital. The rule-based method achieved an F-measure of 0.80 and an accuracy of 0.80. While our method achieves promising performance, a number of avenues for improvement were identified using advanced natural language processing (NLP) techniques.

Keywords: limb fractures, emergency department, radiology reports, classification, rule-based method, machine learning.

1 Introduction

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 on 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 was discharged home. This is not an uncommon event because the reporting did not occur in real-time. As a result, there are potential delays in the diagnosis of subtle fractures missed by the treating doctor until the receipt of the radiologist's report. 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 the reading of the free-text report. Large "batches" of x-rays are reviewed often days after the patient's ED presentation. This is a labour intensive process which 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 [1-2]. This research aims to address these issues through the application of a gazetteer rule-based 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 [3-5]. Studies have shown that rule-based methods can be applied for identifying clinical conditions from radiology reports such as acute cholecystitis, acute pulmonary embolism and other conditions [6]. The purpose of these methods is to simulate human reasoning for any given information processing task to achieve full or partial automation.

2 Related Work

Previous studies that focused on the problem of identification of subtle limb fractures during the diagnosis of ED patients showed that about 2.1% of all fractures were not identified during initial presentation to the Emergency Department [7]. A similar study about radiological evidence for fracture reports that 1.5% of all x-rays had abnormalities that were not identified in the Emergency Department records [8]. Further research also reported that 5% and 2% of the x-rays of the hand/fingers and ankle/foot from a pediatric Emergency Department had fractures missed by the treating ED doctor [9]. These small percentages of incidences may have significant impact on the overall patient healthcare as these missed fractures may 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 using support vector machine (SVM) and machine learning techniques[10-11]. Even though the results of machine learning based classifiers show high effectiveness, their applicability in clinical settings may be limited. Machine learning methods are data–driven, and as a result, if the training sample is not a representative selection of the problem domain, then the resulting model will not generalise. In addition, machine learning approaches are required to be retrained on new corpora and tasks and collating training data to build new classifier models can be a timely and labour intensive process. These issues provide the motivation for the investigation of rule-based methods which have the ability to model expert knowledge as easily implementable rules.

3 Methods

A set of 99 de-identified free-text descriptions of patient's limb x-rays reported by radiologists were extracted from a tertiary hospital's picture archiving and communication system (PACS). An ethics approval was granted by the Human

Research Ethics Committee at Queensland Health to use this data. The average length of free-text reports is about 52 words with total 930 unique words in the vocabulary. Some reports are semi-structured, with section headings such as "History", "Clinical Details", "Findings", appearing in the text.

3.1 Ground Truth Development

One ED visiting medical officer and one ED Registrar were engaged as assessors to manually classify the patient findings. Findings were assigned to either one of the following two classes: (1) "Normal", means identifying no fractures or dislocations and (2) "Abnormal", identifying the presence of a reportable abnormality such as fracture, dislocation, displacement etc., which requires further follow-up. To gather ground truth labels about the data, an in-house annotation tool was developed. This tool allowed the assessors to manually annotate and classify the free-text reports into one of the two target categories. The two assessors initially agreed on the annotations of 77 of the 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.

3.2 Rule-based classifier

A rule-based classifier was developed and implemented with rules as a set of key-words extracted from the x-ray reports assessment criteria as documented by the clinicians prior to the ground truth annotation task. The classifier was implemented to classify the text into "Normal" and "Abnormal" categories as shown in Table 1.

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*

4 Results and Discussion

Results obtained by our gazetteer rule-based approach on the dataset containing 99 radiology reports are reported in Table 2, along with the performance of a Naïve Bayes classifier that was used to classify on the same dataset [12]. The Naïve Bayes classifier was trained and evaluated using a 10-fold cross validation approach. This approach used 90% of reports for training and subsequently evaluated on the remaining 10% within each cross validation fold. The average of the evaluation results across the 10 folds was reported as the classifier's performance. A set of stemmed tokens in combination with high order semantic features such as SNOMED CT concepts related to morphological abnormalities and disorders generated by the Medtex system [13] were used to represent the reports. 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 by rule-based and NB classification

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

The rule-based system classified 49 reports as "Normal". Thirty-three of these were classified as "normal" due to the "no + fracture" rule. The remaining 16 reports did not match any rule, and thus were classified as "normal" (i.e. "no rule fired"). The high false negative count from the rule-based system suggests that the keywords that were used to characterise "Abnormal" cases by the clinician were not complete or adequate to capture all possible cases of abnormalities. Although the proposed keyword rule-based approach is simplistic but shows promise, advanced Natural Language Processing techniques such as those adopted in Medtex [14] can be used to improve classification performances. More keywords can also be learnt using computational linguistic methods, such as the Basilisk bootstrapping algorithm [15].

5 Conclusion and Future Research

This work has described an initial investigation of a clinician-driven rule-based method for automatic classification of free-text limb fracture x-ray findings. We described a simple keyword spotting approach where keywords were derived from classification criteria provided by clinicians. The rule-based classification method achieved promising results with F-measure performances of 0.80 and an accuracy of 0.80. As future work, the research will aim to improve the simple keyword approach with more advanced clinical text processing techniques to complement the proposed rule-based classification method. The possible integration of our method in real-life workflow of hospital emergency departments will also be considered.

Acknowledgements. The authors are thankful to Bevan Koopman for feedbacks on earlier draft of this paper. This research was supported by the Queensland Emergency Medicine Research Foundation Grant, EMPJ-11-158-Chu-Radiology.

References

1. James M. R., Bracegirdle A. and Yates D. W. X-ray reporting in accident and emergency departments – an area for improvements in efficiency. Arch Emerg Med, 8:266–270, 1991.
2. Siegel E., Groleau G., Reiner B. and Stair T. Computerized follow-up of discrepancies in image interpretation between emergency and radiology departments. J Digit Imaging, 11:18–20, 1998.
3. Long W.J, et al. Reasoning requirements for diagnosis of heart disease. Artificial Intelligence in Medicine, 10(1), pp. 5–24, 1997.
4. Harleen K., Siri Krishan W.Empirical Study on Applications of Data Mining Techniques in Healthcare, Journal of Computer Science 2 (2): 194-200, pp.1549-3636, 2006.
5. 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, Vol 2, Issue 3,933-937, March 2012.
6. Lakhani P, Kim W, Langlotz CP. Automated detection of critical results in radiology reports. J Digit Imaging 25(1):30–36, 2012.
7. Cameron MG. Missed fractures in the emergency department. Emerg Med (Fremantle), 6:3, 1994.
8. Sprivulis P. and Frazer A. Same-day x-ray reporting is not needed in well supervised emergency departments. Emerg Med (Fremantle), 13:194–197, 2001.
9. Mounts J., Clingenpeel J., and E. Byers E. McGuire and Kireeva Y. Most frequently missed fractures in the emergency department. Clin Pediatr (Phila), 50:183–186, 2011.
10. De Bruijn B., Cranney A., O'Donnell S., Martin J.D. and Forster A.J. Identifying wrist fracture patients with high accuracy by automatic categorization of x-ray reports. Journal of the American Medical Informatics Association (JAMIA), 13(6):696–698, 2006.
11. Thomas B.J., Ouellette H., Halpern E.F. and Rosenthal D.I. Automated computer-assisted categorization of radiology reports. American Journal of Roentgenology, 184(2):687–690, 2005.
12. Zuccon G, Wagholikar A, Nguyen A, Chu, K, Martin S, Greenslade J., Identifying Limb Fractures from Free-Text Radiology Reports using Machine Learning, Technical Report, CSIRO, 2012.
13. Nguyen AN, Lawley MJ, Hansen DP, et al. A simple pipeline application for identifying and negating SNOMED clinical terminology in free text. Proceedings of the Health Informatics Conference; August 2009, Canberra, Australia; 188–93, 2009.
14. 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), vol. 17, no. 4, pp. 440-445, July/August 2010.
15. Thelen, M., Riloff, E. A bootstrapping method for learning semantic lexicons using extraction pattern contexts, Proceedings of the ACL-02 conference on Empirical methods in natural language processing, p.214-221, July 06, 2002