ORIGINAL ARTICLE

Comparison Between Manual Auditing and a Natural Language Process With Machine Learning Algorithm to Evaluate Faculty Use of Standardized Reports in Radiology

Carolina V. Guimaraes, MD[a,c], *Robert Grzeszczuk, PhD*[b], *George S. Bisset III, MD*[a], *Lane F. Donnelly, MD*[a,c]

Abstract

Purpose: When implementing or monitoring department-sanctioned standardized radiology reports, feedback about individual faculty performance has been shown to be a useful driver of faculty compliance. Most commonly, these data are derived from manual audit, which can be both time-consuming and subject to sampling error. The purpose of this study was to evaluate whether a software program using natural language processing and machine learning could accurately audit radiologist compliance with the use of standardized reports compared with performed manual audits.

Methods: Radiology reports from a 1-month period were loaded into such a software program, and faculty compliance with use of standardized reports was calculated. For that same period, manual audits were performed (25 reports audited for each of 42 faculty members). The mean compliance rates calculated by automated auditing were then compared with the confidence interval of the mean rate by manual audit.

Results: The mean compliance rate for use of standardized reports as determined by manual audit was 91.2% with a confidence interval between 89.3% and 92.8%. The mean compliance rate calculated by automated auditing was 92.0%, within that confidence interval.

Conclusion: This study shows that by use of natural language processing and machine learning algorithms, an automated analysis can accurately define whether reports are compliant with use of standardized report templates and language, compared with manual audits. This may avoid significant labor costs related to conducting the manual auditing process.

Key Words: Standardized reports, machine learning, natural language processing, compliance

J Am Coll Radiol 2017;■:■-■. Copyright © 2017 American College of Radiology

BACKGROUND

The radiology report has been the time-honored formal form of communication between the radiologist and the referring physician [1]. It is also the most tangible product of the radiologists' efforts [1-5]. Multiple previous publications have shown that standardizing the structure and terminology of radiology reports is associated with improved communication to referring care providers [1-12].

Our department of radiology adopted standardized reports across the entire spectrum of imaging and went live with those standardized templates in late 2015. Standardization included standardized sections and headers ("Exam," "Clinical History," "Technique," "Comparison," "Findings," "Impression"), standardized subsections (where appropriate), standardized language for normal studies, and standardized language for common abnormal examinations. At implementation, there were approximately 230 such standardized reports for various imaging studies. For atypical and complex abnormal examinations, free dictation of text

[a]Department of Radiology, Texas Children's Hospital, Houston, Texas.

[b]InContext, Houston, Texas.

[c]Department of Radiology, Stanford University, Stanford, California.

Corresponding author and reprints: Lane F. Donnelly, MD, Chief Quality Officer, Lucile Packard Children's Hospital at Stanford, Stanford University School of Medicine, Stanford Children's Health 180 El Camino Real Ste M384 Mail Code: 5885, Palo Alto, CA 94304; e-mail: lane.donnelly@stanford.edu.

Robert Grzeszczuk is CEO and founder of InContext. The other authors have no conflicts of interest related to the material discussed in this article.

1546-1440/17/$36.00 ■ https://doi.org/10.1016/j.jacr.2017.10.042

into the appropriate report section or subsection occurs.

To monitor and encourage the use of standardized reports, the department audits radiology reports for use of the standard template headers and use of standardized language. Data regarding compliance of all radiologists with standard templates are sent quarterly. A goal was set for the department as well as individual faculty members to have a mean compliance rate of greater than 95%. This parameter is used as part of the Ongoing Professional Practice Evaluation and serves as a quality metric in the department's pay-for-performance plan.

The audits are performed manually and are labor intensive. Twenty-five reports are audited per faculty member per quarter. There are 42 faculty members within the department. Therefore, approximately 4,200 reports are manually audited per year. Three physicians and two nonphysician staff (quality coaches) are involved in the auditing process.

The purpose of this article was to evaluate whether a software program using natural language processing and machine learning could accurately audit radiologist compliance with the use of standardized reports, compared with performed manual audits.

METHODS

The department had engaged in working with a software developer (InContext, Houston, Texas, USA) to develop a software program to evaluate various aspects of radiology reports. One aspect of the project was to evaluate whether the program could conduct automated reviews of all radiology reports for faculty compliance with use of standardized report templates. If possible, this could relieve the labor burden related to performing manual audits and could also provide comprehensive evaluation of all reports, rather than relying on data based on limited sample audits.

All reports generated from October 2015 were de-identified and loaded into the developed software program. This included 28,615 radiology reports. The developed software program used natural language processing and machine learning algorithms to identify and classify statements within the reports. Identified statements were compared with the template definitions syntactically, and content was compared with sections semantically.

Reports were evaluated for adherence with templates created and maintained with PowerScribe360 dictation software (Nuance Communications, Burlington, Massachusetts, USA). First, structural adherence was evaluated, which included verifying that all components defined within the template (eg, sections, fields) were present and filled with values as prescribed by the template definition (eg, "pick lists" or lists of allowable values for a field). Missing, superfluous, and out-of-order sections were reported as nonconformances, as were field values that did not match pick list values.

It is important to note that the structure of the report template is not static and that presence of certain subsections and fields is contextual. For example, on an ultrasound performed to rule out appendicitis, if the appendix is visualized, the template will include additional fields such as "Appendicolith" and "Hyperemia," but these fields would not be present if the appendix was not seen. As a result, a naive approach where a section or a field was reported as missing based on a membership in a simple list could not be used. Supervised machine learning methodology was used to identify legitimate structural combinations using a subset of correctly formatted reports as a training sample.

Secondly, the content of the report was evaluated for topical coherence to make sure that information was placed appropriately within the report. For example, if cardiac findings were dictated within a "Lungs" section instead of within "Heart" or "Additional findings" section (depending on the structure of the template), a "nonconformance" would be reported. As was the case for structural analysis, the semantic component of the system was trained on how to recognize and categorize clinical statements using statistical classifiers.

The compliance rate for the automated review was calculated for specific imaging examinations (chest radiograph, ultrasound abdomen limited—appendicitis). The confidence interval of the manual audit during that same period (October 2015) was calculated. The manual audit consisted of 25 reports for each of 42 faculty members for the month of the study. The mean compliance rates calculated by automated auditing were then compared with the confidence interval of the mean rate by manual audit.

RESULTS

The mean compliance rate for use of standardized reports as determined by manual audit was 91.2% as calculated at the time of the test period. The confidence interval for that manual audit was between 89.3% and 92.8%. The mean compliance rate calculated by automated auditing