Journal of the American Medical Informatics Association, 25(5), 2018, 568–571
doi: 10.1093/jamia/ocx125
Advance Access Publication Date: 27 October 2017
Brief Communication

Brief Communication

Using machine learning for sequence-level automated MRI protocol selection in neuroradiology

Andrew D Brown[1,2] **and Thomas R Marotta**[1,2]

[1]Department of Medical Imaging, St Michael's Hospital, Toronto, ON, Canada and [2]Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author: Andrew D Brown, St Michael's Hospital, Department of Medical Imaging, University of Toronto, 30 Bond Street, 3rd Floor, Cardinal Carter, Toronto, ON, Canada M5B 1W8. E-mail: andrew.brown@sloan.mit.edu. Phone: 647-687-2814, Fax: 416-864-5380

Received 20 June 2017; Revised 15 September 2017; Editorial Decision 4 October 2017; Accepted 7 October 2017

ABSTRACT

Incorrect imaging protocol selection can lead to important clinical findings being missed, contributing to both wasted health care resources and patient harm. We present a machine learning method for analyzing the unstructured text of clinical indications and patient demographics from magnetic resonance imaging (MRI) orders to automatically protocol MRI procedures at the sequence level. We compared 3 machine learning models – support vector machine, gradient boosting machine, and random forest – to a baseline model that predicted the most common protocol for all observations in our test set. The gradient boosting machine model significantly outperformed the baseline and demonstrated the best performance of the 3 models in terms of accuracy (95%), precision (86%), recall (80%), and Hamming loss (0.0487). This demonstrates the feasibility of automating sequence selection by applying machine learning to MRI orders. Automated sequence selection has important safety, quality, and financial implications and may facilitate improvements in the quality and safety of medical imaging service delivery.

Key words: radiology, MRI, machine learning, quality improvement, patient safety, clinical decision support

INTRODUCTION

The ultimate goal of medical imaging is to deliver results that convey the most meaningful clinical information.[1] However, even after determining the most appropriate imaging test, a myriad of complicated choices must be navigated during the course of image acquisition. The choices that determine how the most effective set of medical images should be acquired can vary according to factors as diverse as the institution in which the images are acquired, the body part being tested, and the manufacturer of the imaging equipment used. Imaging protocols are the precise instructions that define how a set of medical images should be acquired. If images are acquired incorrectly, important clinical findings may be missed, contributing to wasted health care resources and patient harm.

Computerized clinical decision support systems have been shown to improve clinical decision-making.[2–5] In this work, we demonstrate how machine learning techniques could be used as the basis of a clinical decision support system designed to predict appropriate magnetic resonance imaging (MRI) sequence selection.

Use of protocols in clinical care

In medical imaging, the term "protocol" is used to describe the conditions, parameters, and settings for image acquisition and reconstruction for a requested procedure.[6] The imaging protocol may include instructions for patient preparation, such as administration of oral or intravenous contrast media, the area of the body to be imaged, and the specific imaging modality. In the nonemergent clinical setting, a referring physician (primary care or specialist) orders a radiological imaging procedure for a patient (Figure 1). This order may be completed on a handwritten order form or entered into the patient's electronic medical record (EMR). In more sophisticated

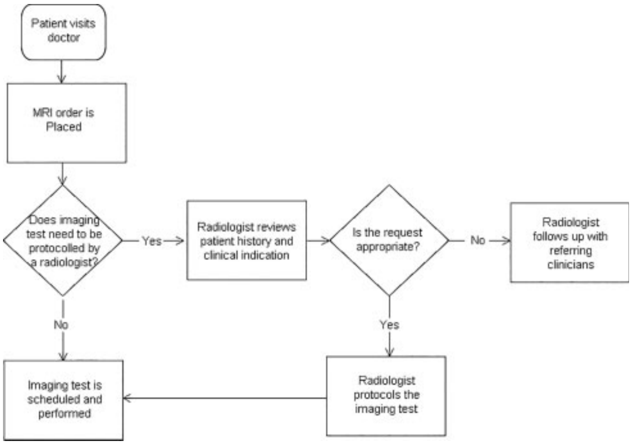

Figure 1. Process of medical imaging exam protocoling.

electronic order entry systems, the referring physician may be assisted by decision support software to ensure that the order aligns with evidence-based appropriateness guidelines.[7,8]

In some instances, the imaging protocol is obvious from the order. For example, in the case of a chest X-ray, 2 standard views of the chest are obtained. However, for more advanced imaging modalities, no one-to-one correspondence exists between the imaging procedure ordered and the protocol. Advanced imaging studies, such as computed tomography (CT) and MRI, can have multiple protocols associated with a single procedure. Selection of the appropriate protocol then depends on the indication for the exam and the patient history. For example, a physician may order a brain MRI for a patient, but the radiology practice may have multiple protocols for imaging the brain, each designed to address a particular clinical question. In MR imaging, a radiologist must review the clinical history, the indications for the test, and the requested exam to select the correct imaging sequences that optimize the amount of information that can be obtained from each imaging test. An MRI sequence is a computer program that controls the soft-tissue contrast as well as the spatial orientation and resolution of the images acquired.[9] This step of assigning the appropriate imaging parameters is referred to as protocoling and ensures that the correct protocol with the correct parameters is used.[10]

Limitations of the current system

One of the biggest limitations of the current protocoling workflow is that it is time-consuming. In a recent study of the duration and quantity of tasks performed in the radiology reading room, Schemmel et al.[11] found that protocoling and tasks related to image acquisition make up a significant portion of the workday for most radiologists.

In addition, protocoling tasks disrupt the radiologist's primary responsibility – image interpretation. For example, radiologists may be interrupted while interpreting imaging studies by phone calls from CT and MRI technologists regarding imaging protocols. A study by Yu and colleagues[12] found that an on-call radiologist could be interrupted as many as 2 or 3 times during the interpretation of a single CT or MRI scan to attend to responsibilities such as protocoling, injections, and communicating with referring clinicians. These types of disruptions divert attention from image interpretation and have been correlated with interpretative errors and patient harm.[13,14]

Automated protocol selection

Manual protocoling is inherently inefficient, time-consuming, and cumbersome. In this study, we extend methods for automated protocol selection to specific MRI sequences. The primary objective of this study was to determine whether machine learning techniques can be used to prescribe specific MRI sequences from data found in MRI orders. The secondary aim was to employ a distance metric to compare model results in the application of a multilabel classification problem.

Our machine learning methods performed MRI sequence selection with a high degree of accuracy and significantly outperformed the baseline model. We hope this work will contribute to the development of systems and policies that reduce the potential for error and improve the quality and safety of medical imaging service delivery.

METHODS

Data description

Data were supplied by St Michael's Hospital Department of Medical Imaging. St Michael's Hospital is an academic teaching hospital and level I trauma center in downtown Toronto, Ontario, Canada. Permission to use the dataset was granted by the hospital's institutional review board. Using the Montage Search and Analytics platform (Montage Healthcare Solutions, Philadelphia, PA, USA), radiology information system (RIS) data were extracted from all MRI brain examinations performed during an 18-month period from January 1, 2014, to June 30, 2015. The RIS dataset, which contained 7487 observations in .csv format, documented patient demographics, study type, and the study's clinical indication. Data from the EMR and interpretations of prior medical imaging reports were not used in the present work, but may be included in a future study. All data analysis was performed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria).

Features

A training dataset for the models was constructed using traditional natural language processing techniques.[15] An example of the clinical indication provided by the ordering physician would be the following: "Clinical History: Assess acute ischemic stroke versus Todds paralysis." Text from the clinical indications stated in MRI orders was converted to lowercase and stop words were removed. Additional stop words specific to our narrative data included words such as "clinical," "indication," "history," "please," and "assess."

The processed sentences produced a term-document matrix, with each word representing a separate column and each row representing an MRI order. Each word served as a single independent variable in the model. Additional independent variables included patient age and sex, location, and ordering service.

Training strategy

All models were trained using the dataset from the 18-month observation period. We randomly divided the dataset into a training set (70% of the total dataset) and a test set (30% of the total dataset). In the dataset, each MRI had an associated set of MRI sequences, which were selected at the time of protocoling by a radiologist. Radiologists at our institution can choose from 41 different MRI sequences. The prediction task represents a multilabel classification problem with 41 classes, each corresponding to an MRI sequence.

Algorithms

To make meaningful comparisons, we defined a baseline method against which we compared the results of the prediction models. As our baseline method, we predicted the most commonly occurring outcome in our dataset for all observations in our test set.

Table 1. Dataset characteristics

Characteristics	Overall (7487)	Training (5239)	Testing (2248)
Age (mean [SD])	50.4 (16.4)	50.4 (16.4)	50.2 (16.3)
Female (%)	60.3	59.2	62.6
No. of words[a] (mean [SD])	3.9 (2.4)	3.9 (2.4)	3.9 (2.3)
Location (%)			
Emergency	1.1	1.1	1.2
Inpatient	19.2	19.3	19.0
Outpatient	79.4	79.4	80.0
Ordering service (%)			
Neurology	28.3	28.7	27.4
Neurosurgery	21.0	20.3	22.5
Family medicine	17.6	17.6	17.7
Most common protocols			
Fast brain with gadolinium[b]	1513	1059	454
Fast brain[c]	1319	923	396
Sella with gadolinium[d]	908	636	272

SD, standard deviation.

[a] Represents the mean number of words in the clinical indication per examination.

[b] Consists of the following sequences: sagittal T1 3D, axial T2 turbo spin echo (TSE), axial T2 fluid-attenuated inversion recovery (FLAIR), axial T2 fast field echo (FFE), diffusion-weighted imaging (DWI), axial T1, axial T1 with gadolinium, coronal T1 3D with gadolinium.

[c] Consists of the following sequences: sagittal T1, axial T2 TSE, axial T2 FLAIR, axial T2 FFE, DWI.

[d] Consists of the following sequences: sagittal T1, coronal T1, coronal T2, sagittal T1 with gadolinium, coronal T1 with gadolinium.

Using the caret package in R, support vector machine, gradient boosting machine (GBM), and random forest models were trained with the term-document matrix and patient demographic information to predict the specific MRI sequences for observations in the test set.

Model evaluation

For each machine learning algorithm, we created 41 different binary classifiers to predict whether or not a specific sequence should be prescribed to an observation in the test set. We then compared the algorithm's results to sequence selections made by our radiologists. To measure the quality of the model predictions, we calculated accuracy, precision, and recall. We also calculated Hamming loss,[16] which is a multilabel distance metric that measures the proportion of misclassifications. The lower the Hamming loss, the lower the rate of misclassification.

RESULTS

Descriptive statistics of our dataset are presented in Table 1. Overall, our dataset consisted of 7487 observations that represent a single record of a single study request in the RIS and does not include the final report. Exploratory data analysis demonstrated that the dataset was relatively sparse, with label density of 0.147 and cardinality of 6.02. Label density of 0.147 indicates that the average percentage of occurrence of each sequence in the dataset is 14.7%. Label cardinality of 6.02 signifies that each observation has an average of >6 labels, that is, MRI sequences, associated with it.

The performance of the machine learning models relative to the baseline is shown in Table 2. All 3 machine learning models outperformed the baseline. Of these models, GBM demonstrated the best performance on all metrics.

DISCUSSION

Selection of MRI sequences is an important task in advanced medical imaging. A radiologist selects sequences that will provide the greatest amount of information and enable the interpreter of the images to answer the clinical question posed by the referring clinician. However, protocoling workflow can be inefficient and error-prone.[1]

Automated sequence selection therefore has important safety, quality, and financial implications. In the area of safety, these techniques may free radiologists from distraction, allowing them to focus on image interpretation, and reduce variation in sequence selection, which often introduces error and waste in radiology workflows.[17] In the quality and financial areas, this approach may facilitate greater efficiency, reducing wait times and improving accessibility to medical imaging services while improving patient throughput and revenue for imaging providers.

Here we demonstrate an automated approach to protocoling MRI brain procedures at the sequence level. Performing predictions at the level of the MRI sequence takes the problem of protocoling to a more granular level than previous studies.[18] This allows for greater standardization and a more robust clinical application.[19]

Our results demonstrate that the baseline model was relatively accurate (Table 2). This is likely because the fast brain protocol is commonly used at our institution (Table 1) and the sequences used in this protocol are shared by other protocols. However, despite this favorable level of accuracy, the Hamming loss of our baseline is poor, suggesting that the baseline struggles with edge cases. This potential limitation of the baseline is alluded to by the sparse nature of the dataset. The machine learning approaches significantly outperform the baseline, with GBM demonstrating the best performance.

Despite the high level of accuracy of our machine learning approach, several limitations are noted. The model was constructed using data from a single institution. Our hospital has robust neurology and neurosurgery services, a factor that has implications for the language used in imaging orders and the mix of procedures. This may limit the direct translation of these models to other academic hospitals or community settings that may have less advanced MRI capabilities. In addition, these models did not make use of EMR data. In practice, radiologists may supplement the history provided in the

Table 2. Model performance metrics compared to baseline.

Models	Baseline	Support vector machine	Random forest	Gradient boosting machine
Accuracy	0.889	0.949	0.948	0.951
Precision	0.647	0.834	0.840	0.856
Recall	0.538	0.815	0.795	0.804
Hamming loss	0.1110	0.0510	0.0523	0.0487

procedure order with details from the EMR.[20] The inclusion of patient history from the EMR has the potential to significantly improve the results. In future studies we may be able to improve the performance of our models by employing additional preprocessing steps, including negation detection, part-of-speech processing, and assigning to clinical ontologies.[21]

CONCLUSION

Our results demonstrate the feasibility of using machine learning techniques to automatically process unstructured, free-text clinical indications and patient demographics to reliably perform sequence-level protocol selection in MRI brain procedures. Of the models tested, the GBM model achieved the highest accuracy (>95%) on the test set. This method demonstrates the potential application of machine learning as a foundation for clinical decision support tools to guide sequence acquisition decisions and potentially improve efficiency, quality, and cost.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.

COMPETING INTERESTS

The authors have no competing interests to declare.

REFERENCES

1. Boland GW, Duszak R, Kalra M. Protocol design and optimization. *J Am Coll Radiol*. 2014;11:440–41.
2. Garg AX, Adhikari NKJ, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*. 2005;293(10):1223–38.
3. Roshanov PS, Misra S, Gerstein HC, *et al*. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. *Implement Sci*. 2011;6(1):92.
4. Bright TJ, Wong A, Dhurjati R, *et al*. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med*. 2012;157(1):29–43.
5. Nuckols TK, Smith-Spangler C, Morton SC, *et al*. The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis. *Syst Rev*. 2014;3(1):56.
6. Supanich MP. Computed tomography imaging operation. In: Haaga JR, Boll D, eds. *CT and MRI of the Whole Body, 2-Volume Set*. 6th ed. Philadelphia: Elsevier; 2017: 3–30.
7. Rosenthal DI, Weilburg JB, Schultz T, *et al*. Radiology order entry with decision support: initial clinical experience. *J Am Coll Radiol*. 2006;3(10):799–806.
8. Sistrom CL, Dang PA, Weilburg JB, *et al*. Effect of computerized order entry with integrated decision support on the growth of outpatient procedure volumes: seven-year time series analysis. *Radiology*. 2009;251(1):147–55.
9. Brown MA, Semelka RC. MR imaging abbreviations, definitions, and descriptions: a review. *Radiology*. 1999;213:647–62.
10. Swensen SJ, Johnson CD. Radiologic quality and safety: mapping value into radiology. *J Am Coll Radiol*. 2005;2(12):992–1000.
11. Schemmel A, Lee M, Hanley T, *et al*. Radiology workflow disruptors: a detailed analysis. *J Am Coll Radiol*. 2016;13(10):1210–14.
12. Yu JP, Kansagra AP, Mongan J. The radiologist's workflow environment: evaluation of disruptors and potential implications. *J Am Coll Radiol*. 2014;11(6):589–93.
13. Balint BJ, Steenburg SD, Lin H, *et al*. Do telephone call interruptions have an impact on radiology resident diagnostic accuracy? *Acad Radiol*. 2014;21(12):1623–28.
14. Grundgeiger T, Sanderson P. Interruptions in healthcare: theoretical views. *Int J Med Inform*. 2009;78:293–307.
15. Manning CD, Raghavan P, Schutze H. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press; 2008:19–34.
16. Schapire RE, Singer Y. BoosTexter: a boosting-based system for text categorization. *Machine Learning*. 2000;39(2–3):135–68.
17. Boland GW, Duszak R. Protocol management and design: current and future best practices. *J Am Coll Radiol*. 2015;12(8):833–35.
18. Brown AD, Marotta TR. A natural language processing-based model to automate MRI brain protocol selection and prioritization. *Acad Radiol*. 2017;24(2):160–66.
19. Pijl ME, Doornbos J, Wasser MN, *et al*. Quantitative analysis of focal masses at MR imaging: a plea for standardization. *Radiology*. 2004;231(3):737–44.
20. Waite S, Scott J, Gale B, *et al*. Interpretive error in radiology. *Am J Roentgenol*. 2017;208(4):739–49.
21. Friedman C, Shagina L, Lussier Y, *et al*. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc*. 2004;11(5):392–402.