Predicting High Imaging Utilization Based on Initial Radiology Reports: A Feasibility Study of Machine Learning

Saeed Hassanpour, PhD, Curtis P. Langlotz, MD, PhD

Rationale and Objectives: Imaging utilization has significantly increased over the last two decades, and is only recently showing signs of moderating. To help healthcare providers identify patients at risk for high imaging utilization, we developed a prediction model to recognize high imaging utilizers based on their initial imaging reports.

Materials and Methods: The prediction model uses a machine learning text classification framework. In this study, we used radiology reports from 18,384 patients with at least one abdomen computed tomography study in their imaging record at Stanford Health Care as the training set. We modeled the radiology reports in a vector space and trained a support vector machine classifier for this prediction task. We evaluated our model on a separate test set of 4791 patients. In addition to high prediction accuracy, in our method, we aimed at achieving high specificity to identify patients at high risk for high imaging utilization.

Results: Our results (accuracy: 94.0%, sensitivity: 74.4%, specificity: 97.9%, positive predictive value: 87.3%, negative predictive value: 95.1%) show that a prediction model can enable healthcare providers to identify in advance patients who are likely to be high utilizers of imaging services.

Conclusions: Machine learning classifiers developed from narrative radiology reports are feasible methods to predict imaging utilization. Such systems can be used to identify high utilizers, inform future image ordering behavior, and encourage judicious use of imaging.

Key Words: Imaging utilization; natural language processing; prediction modeling; radiology report narrative.

INTRODUCTION

Imaging utilization and cost has increased significantly in the last two decades (1–6), although it has stabilized more recently (7–10). Nevertheless, imaging remains a focus of cost-control efforts. The cost of imaging is currently estimated to be $100 billion annually in the United States (11–13). These issues are especially relevant to high-cost imaging technologies such as computed tomography (CT), magnetic resonance imaging, and positron emission tomography (1,14). A recent study of imaging utilization in large integrated healthcare systems showed a significant increase in the rate of advanced diagnostic imaging and associated radiation exposure between 1996 and 2010 (15). The utilization of CT studies increased by 7.8% annually, and tripled during the 15 years examined by the study. Although national data show the growth rate moderating somewhat, there remain concerns about cost and radiation exposure.

The abdomen is often affected by chronic diseases, such as renal stones and inflammatory bowel disease. Thus, a subset of patients undergoing abdomen-related imaging studies such as abdomen CTs are an appropriate target for the identification of patients at risk for repeated studies, elevated exposure to ionizing radiation, and high costs.

To help our organization and other healthcare providers predict patients at risk for repeated imaging, we developed a prediction model that can recognize high imaging utilizers based on their first 2 days of imaging reports. In this study, we focused on a cohort of patients with abdominal problems. This prediction model is intended to enable healthcare providers to identify patients likely to be high utilizers of imaging services with high specificity.

In a recent related work with a different focus, a readmission prediction model identified patients at high risk of being rehospitalized soon after the discharge for congestive heart failure (16). Unlike our work, which uses narrative text, the readmission prediction model uses structured data for patients at the end of hospitalization, such as diagnosis codes (*International Classification of Diseases, Ninth Revision*), lab results, medications and medications history, and fees and billing, in addition to free-text clinical notes. Given our goal to predict imaging utilization early in a patient's clinical course, many of these clinical data may not be available at early days of imaging utilization. Therefore, to maximize the lead time

Acad Radiol 2016; 23:84–89

From the Department of Radiology, Stanford University, 300 Pasteur Drive, Stanford, CA 94305. Received July 1, 2015; revised August 29, 2015; accepted September 16, 2015. **Address correspondence to:** S.H. e-mail: Saeed.Hassanpour@dartmouth.edu

http://dx.doi.org/10.1016/j.acra.2015.09.014

benefits of our system and because of our focus on imaging, our prediction model requires the first days of imaging utilization reports to provide an accurate prediction.

MATERIALS AND METHODS

Data Set Construction

The Stanford Translational Research Integrated Database Environment was used to build training and testing data sets. Through this database, we had access to 4,056,227 radiology reports from 564,210 patients since 1998. The use of these data in our project was approved by Stanford's institutional review board. To prepare the data, we grouped the radiology reports by their corresponding patient identifiers and extracted the imaging study dates through regular expressions (17). For this work, we excluded patients with one or more missing study dates. We required a minimum 6-month follow-up period for patients in our data set to exclude the patients whose imaging utilization may still be in progress at the point of our study. As a result, our data set covers patients who started using Stanford imaging facility from 1998 through 2014. We excluded patients whose first imaging studies were obtained in 2015.

To focus on patients with abdominal problems, we narrowed our data set to patients who had at least one abdomen CT study in their imaging history at our organization. The considered abdomen CTs for assembling the cohort include all abdomen CT variants, such as abdomen and pelvis CTs. After these data cleaning steps, our data set contained 23,175 patients who had undergone a total of 111,247 imaging studies. We randomly divided the corresponding imaging reports into a training set and a test set. The training set contained 18,384 patients (~80% of entire data set), and was used to build the prediction model. Our test set containing 4791 patients (~20% of entire data set) was used to evaluate the prediction model.

Study Design

We constructed a prediction system that uses initial radiology report contents to identify patients likely to be high utilizers of imaging studies. Because our system is a binary classifier, we needed to select a cutoff point for patients who were high versus low imaging utilizers. Using the clinical expertise of one of the authors and consulting with other domain expert radiologists at our organization (see Acknowledgments), we devised a clinically reasonable number of imaging days to serve as a cutoff point. We considered patients with more than 5 days of imaging utilization as high imaging utilizers (positive) and the patients with 5 or less days of imaging utilization as low imaging utilizers (negative). The imaging utilization covers all types of imaging studies such as abdomen and pelvis CTs and chest X-rays. This 5-day cutoff gives us a clinically plausible measure to recognize heavy utilizers of our imaging facility and classifies one in six patients in our data set as high utilizer. Because the level of temporal granularity for imaging studies in our data set is the calendar day and all performed studies on a single day share the same time stamp, we used days of imaging utilization, which is highly correlated with the number of imaging studies, as the dependent variable. We extracted the radiology reports according to this criterion to build a training set. In the training set, we combined the radiology reports for studies that are performed on a same day as a single day of reports. Because there is one radiology report for each performed study, there may be one or more individual study reports in a single day of imaging utilization.

To build the prediction system, we used a support vector machine (SVM) (18) learning framework that utilizes radiology reports from patients with high and low imaging utilization to train a text classifier. Among different classification approaches, SVM is one of the most effective and widely used methods. In this work, we modeled the training set's radiology report text as vectors and used an SVM framework to build a classifier that determines if a patient will be a high utilizer based on the radiology report text from the patient's first days of imaging utilization.

Our preliminary work also determined the optimal number of days of input reports required by the model to make an accurate prediction. A separate test set of patients was used to evaluate the performance of the prediction system. Standard classification evaluation metrics such as accuracy, sensitivity (recall), specificity, positive predictive value (precision), and negative predictive value were used to measure the quality of the prediction results.

Radiology Report Modeling

To use radiology report content in an SVM classification framework, we needed to model radiology report text quantitatively. We therefore modeled radiology reports as vectors in Euclidian space, where each vector dimension corresponds to an n-gram, which is a contiguous sequence of one, two, or three tokens (i.e. words) in a report. If a report contains a specific n-gram, that n-gram has a nonzero weight in the report's vector representation. The weight of each n-gram was computed using term frequency–inverse document frequency, a common weighting scheme in text mining (19). A term frequency–inverse document frequency weight increases proportionally by the n-gram frequency in the report, and is scaled down by the commonality of the n-gram among all reports in the data set.

SVM Prediction Model

SVM is a maximum-margin classifier, which finds the decision boundary with the largest separation between positive and negative training examples (18). In this work, we used LIBSVM, a widely used open source machine learning library, to train our SVM classifier (20). In our model training, we wanted to provide enough context for each training example to manage sparse information in the training set. We experimented with incorporating the first few days of imaging