SIIM 2017 Scientific Session
Posters & Demonstrations

Automatic Determination of The Need for Intravenous Contrast in Musculoskeletal MRI Examinations Using a Machine Learning Based Natural Language Processing Algorithm

Hari M. Trivedi, MD, University of California, San Francisco; Joseph Mesterhazy; Thienkhai Vu ; MD; Jae Ho Sohn, MD

Background

Efficient and accurate MR protocoling is critical for maintaining radiology workflow and patient care. However, MRI protocoling can be time consuming and vary according to the practice patterns of the protocoling radiologist. Providing a best-practices recommendation for an MRI protocol has the potential to improve efficiency and decrease diagnostic error rates. The purpose of this study was to develop and validate a deep learning-based natural language classifier that uses the free-text clinical indication of an MRI study to appropriately assign the use of intravenous contrast in the protocol.

Case Presentation

1544 free-text clinical indications and their assigned free-text MRI protocols were retrospectively extracted from an in-house radiology communication tool. The data was manually confirmed to be free of any patient identifiers. Study types included all MRIs of the spine (inclusive of all neurologic, traumatic, orthopedic, and oncologic indications) and all musculoskeletal MRIs. Each protocol was preliminarily automatically assigned into the classification “with contrast” (Class 1) or “without contrast” (Class 0) by cross-referencing a list of 79 text strings. Each classification was then manually reviewed by the authors and corrected if necessary. 24 examinations were excluded because the final protocol was ambiguous regarding the use of contrast - for example, protocols specifying that the radiologist be called after the initial sequences to determine whether or not to administer contrast.

The final data set was then divided into training/cross-validation and test sets, containing 1240 and 280 studies, respectively. Supervised machine learning was performed on the training set using a proprietary deep learning-based algorithm from IBM Watson that uses hypothesis generation, string analysis, and word-scoring to generate a prediction for Class 0 or Class 1. Accuracy was validated with the test set. “R: A language and environment for statistical computing” was used for descriptive statistics and other text mining tasks.

Outcome

Of the 1,520 MRI examinations, 650 (42.8%) protocols included the use of intravenous contrast (Class 1) while 870 (57.2%) did not (Class 0). Training time was 46 minutes. In the test set, the algorithm correctly classified 97.9% (137 of 140 cases) in Class 0 and 85.7% (120 of 140 cases) in Class 1, for an overall accuracy of 91.8%. The 3 most common words in the clinical indication were: “pain, weakness, and injury (Figure 1).”

Figure 1

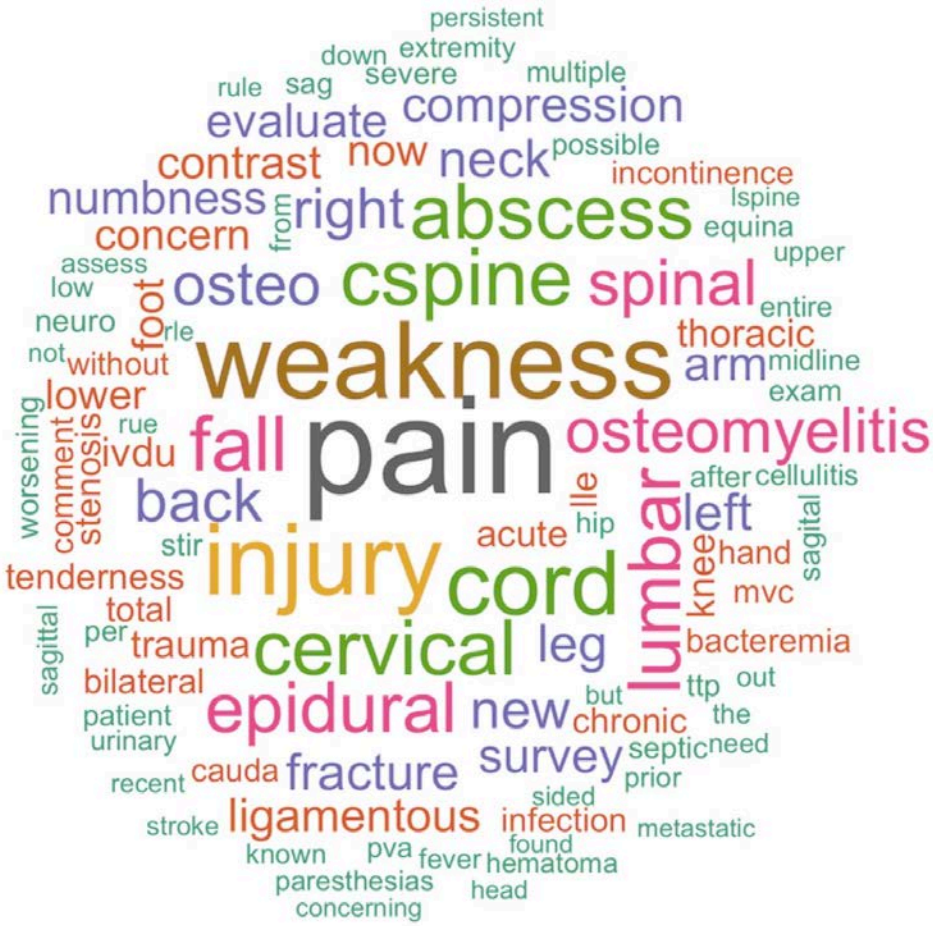

Figure 1. Wordcloud of most commonly mentioned words in indications.

Incorrectly classified cases where no contrast should have been given include epidural abscess evaluation in a dialysis patient and brachial plexus evaluation in an IV drug user. The algorithm also identified 9 cases where contrast was indicated, but was not given by the protocoling radiologist.

Incorrectly classified cases where contrast should have been given include patients with clinical indications that included the terms sag survey, lumbar spine pain, TB, and neurocysticercosis, among others. The algorithm identified 7 cases where contrast was not indicated, but was given by the protocoling radiologist.

Discussion

Our classifier achieved excellent accuracy, particularly for studies in Class 0 not requiring intravenous contrast. One overarching source of error was cases for which there was no consistent pattern in contrast administration, thereby limiting our algorithm's performance given the lack of a true 'gold standard.' In the future, such errors could be ameliorated by requiring concordance among multiple radiologists for any protocol used in the data set.

It is also worth noting that the above results were achieved without including the ordered study type (i.e. MRI lumbar spine without contrast) in the training data. We expect that the inclusion of this data would increase overall accuracy but at the expense of misclassifying cases where the ordered study was erroneous, for example a non-contrast MRI ordered to evaluate for osteomyelitis.

Finally, an intrinsic limitation of our study was that the radiologist-assigned protocol was free-text rather than selected from a pre-defined set, as is more commonly done in modern systems. For future work, we propose to train a similar algorithm on a large data set with pre-defined protocols across multiple subspecialties. Given a large enough training set and inclusion of the ordered study type in the training data, it is conceivable that a global algorithm for MRI protocoling across all subspecialties could be created. Such a tool could also be modified for use by the referring clinician to determine the correct study to order, a frequent question fielded by radiologists at our institution.

Conclusion

We demonstrate that supervised machine learning using natural language processing can be an effective tool for determining the need for intravenous contrast in musculoskeletal MRI studies. The ability of deep learning-based classification is not limited to two classes (contrast vs. no contrast), but rather can be trained to distinguish between several dozen classes given sufficient input data for training.

References

1. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
2. Brown, Andrew D, and Thomas R Marotta. "A Natural Language Processing-Based Model to Automate MRI Brain Protocol Selection and Prioritization.." Academic Radiology, November 23, 2016. doi:10.1016/j.acra.2016.09.013.
3. Ginat, Daniel Thomas, Pranay Uppuluri, Gregory Christoforidis, Gregory Katzman, and Seon-Kyu Lee. "Identification of Neuroradiology MRI Protocol Errors via a Quality-Driven Categorization Approach.." Journal of the American College of Radiology : JACR 13, no. 5 (May 2016): 545–48. doi:10.1016/j.jacr.2015.08.027.
4. Meystre, S M, G K Savova, K C Kipper-Schuler, and J F Hurdle. "Extracting Information From Textual Documents in the Electronic Health Record: a Review of Recent Research.." Yearbook of Medical Informatics, 2008, 128–44.
5. Pons, Ewoud, Loes M M Braun, M G Myriam Hunink, and Jan A Kors. "Natural Language Processing in Radiology: a Systematic Review.." Radiology 279, no. 2 (May 2016): 329–43. doi:10.1148/radiol.16142770.

Keywords

artificial intelligence, machine learning, deep learning, natural language processing, NLP