Study uses AI to sort patient messages by complexity
28 August 2020, by Paul Govern

Taking an interest in electronic message threads between surgical patients and their health care teams, a research group at Vanderbilt University Medical Center has tested how well certain commonly used machine learning algorithms can classify such exchanges according to their clinical decision-making complexity. Their report is online ahead of print publication in the Journal of Surgical Research.

The authors note that health care payers such as Medicare include consideration of the complexity of medical decision-making when determining payment for services.

"If effective, automated message analysis might quantify the care delivered online or support billing for online care," the authors write. It could aid staffing decisions and "may aid with [message] triaging."

Two surgeon-researchers independently labeled 500 threads according to their complexity of medical decision-making, and, discussing any disagreements, achieved consensus on labels for each thread: straightforward, low, moderate and no decision. (It turned out there were no highly complex threads in the set.)

The team tested how closely two standard multi-class machine learning algorithms could match this expert classification, one a random forest classifier and the other a multinomial naïve Bayes classifier. Each was trained and validated on 450 of the labeled threads, then tested on the remaining 50. Accuracy was measured in terms of precision, or the ratio of true positives retrieved to the sum of true and false positives retrieved, and recall, or the ratio of true positives retrieved to all positives in the set.

Across their set's four labels of straightforward, low, moderate or no clinical decision-making complexity, with a score of 1.0 signifying perfection, the best performance from the team's two machine learning models was 0.58 for precision, 0.63 for recall.

"Though they did far outperform a third program that graded complexity by simply adding up the number of medical terms in each message thread, neither of the two currently trained machine learning algorithms could be considered adequate for clinical use without more data and further analysis," said the study's lead author, Lina Sulieman, Ph.D., research fellow in the Department of Biomedical Informatics. "Among the details of this study are several findings that can help us improve this type of automated analysis going forward."

Previous studies by Sulieman and others have used machine learning to classify incoming patient messages according to the general types of needs expressed in them—medical, logistical, informational, etc. According to the authors, this looks to be the first attempt to automatically sort message threads according to clinical decision complexity.
According to the study, VUMC's patient portal, My Health at Vanderbilt (the source for the threads used in the study), receives around 30,000 messages from patients and family members in a typical month.

"Secure messaging is one of the most popular features of patient portals, with hospitals seeing exponential growth in the volume of messages," Sulieman said. "Quantifying the complexity of decision-making in patients' messages can facilitate the identification of the right person to manage the thread and reply to the messages based on the level of medical complexity. Currently, this is a manual process and finding a way to automatically perform the triage can save time spent on reading the message and delegate the task to the right person in the team."


Provided by Vanderbilt University