Exploring Critical Findings and their Communication in Radiology Text Reports using Natural Language Understanding

ACR 2017

Authors and Disclosures

Niketh Bandla, MD - Nothing to disclose

Gilan El Sadaawi, MD PhD - Chief Medical Informatics Officer, M*Modal

Eliot Siegel, MD, FACR, FSIIM - Nothing to disclose

Purpose

There are defined guidelines on critical findings within radiology reports, which specify whether direct communication is recommended within minutes, hours or days.

However from a patient safety and quality monitoring perspective, it is difficult to determine how often these critical findings are detected and documented, and whether these findings are appropriately communicated to the appropriate physician.

A Natural Language Understanding (NLU) based algorithm was proposed to identify these findings and also track communication of findings within the context of the text report.

Methods

1000 text reports from multiple imaging modalities were randomly selected and used to refine a NLU algorithm to categorize reports using 3 categories: (1) ‘No critical findings’, (2) ‘critical findings with communication’, and (3) ‘critical findings without communication’.

Critical findings were selected as defined by The Joint Commission.

Table 1. Suggested critical results for The Joint Commission
Ectopic pregnancy
Intracranial hemorrhage
Pulmonary embolus
Ruptured/leaking aortic aneurysm
Severe spinal cord compression
Significant misplacement of tubes or catheters
Tension pneumothorax
Testicular/ovarian torsion
Unexplained pneumoperitoneum
Unstable spine fracture
Note: Results are listed alphabetically; no priority should be implied.

Methods Continued

The text reports were also classified into (1) ‘Red’, (2) ‘Orange’, and (3) ‘Yellow’ which correspond to actionable findings that need to be directly communicated within minutes, hours, and days as defined by ACR guidelines. Each group was also separated into reports with and without evidence of communication.

The performance of the algorithm was compared with human evaluators using the recommendations in the ACR guidelines.

Results

Within the dataset, the algorithm detected 968 reports without any critical findings, and 30 reports (3%) with critical findings as outlined by The Joint Commission. Two reports could not be assessed.

When assessed against human readers as a reference, initial findings of the algorithm showed 32 false negative reports, and 17 false positive reports containing Critical Findings.

All of the 30 reports that the algorithm found positive for critical findings were accurately assessed as to whether communication was documented in the radiology report.

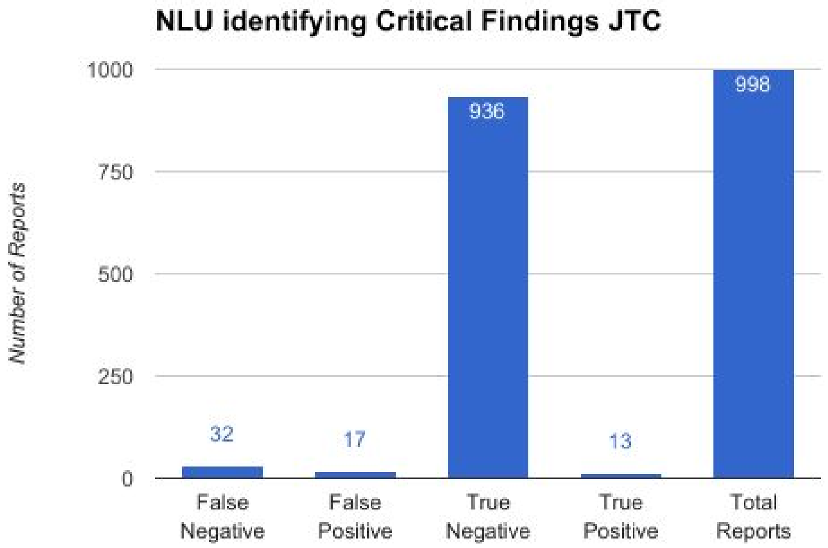
Can NLU find critical findings accurately?

Sensitivity: 29%

Specificity: 98%

PPV: 43%

NPV: 97%



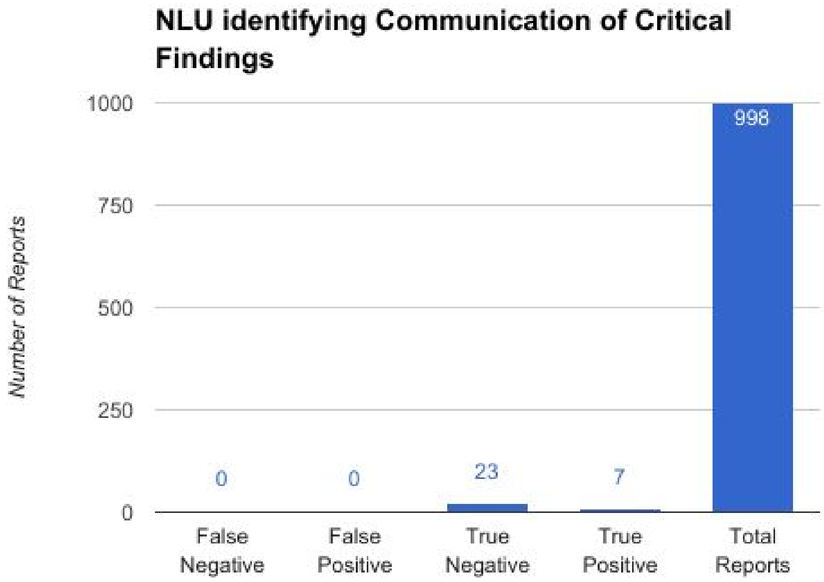
Can NLU identify if CF were communicated?

Sensitivity: 100%

Specificity: 100%

PPV: 100%

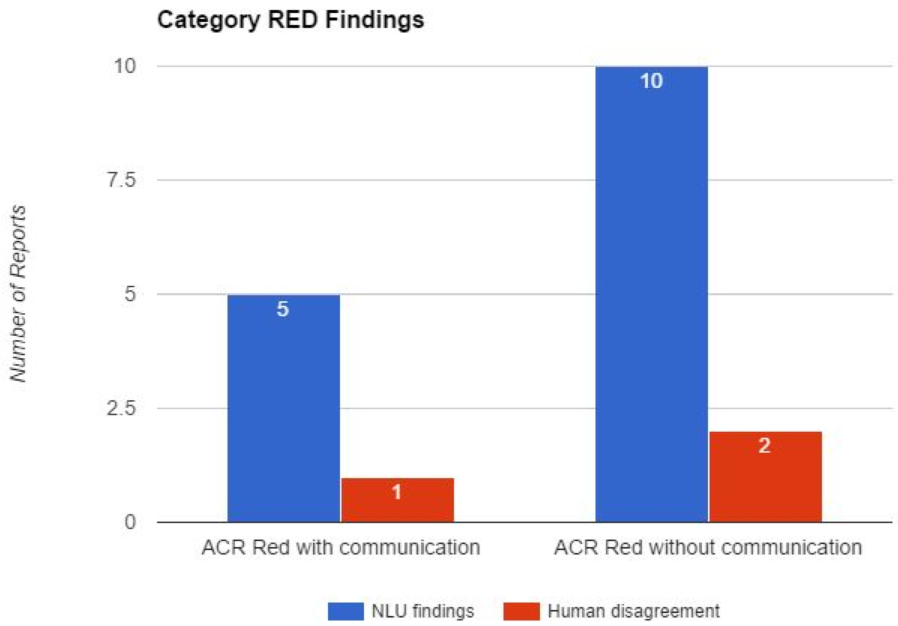
NPV: 100%

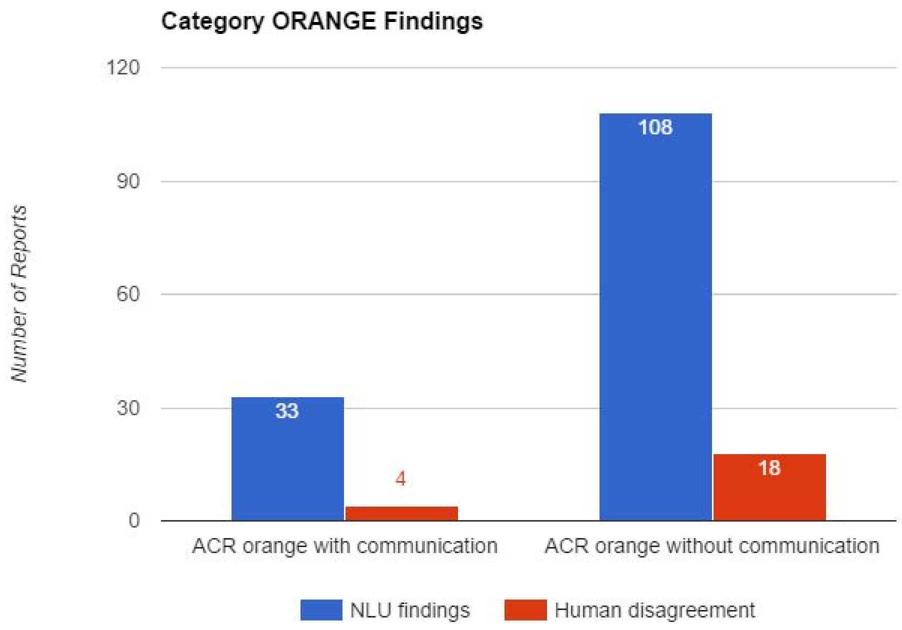


Results continued

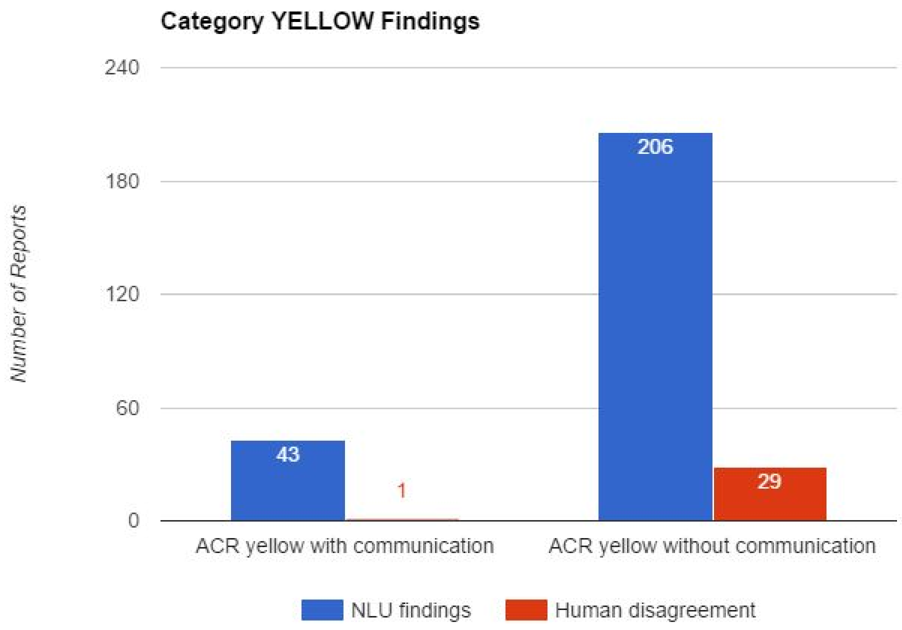
Upon further refinement, the NLU algorithm was able to identify 405 reports with actionable findings. These reports were categorized into Red, Orange, and Yellow for which it found a total of 15, 141, and 249 reports in those categories respectively. Red, Orange, and yellow represent categories for which findings should be communicated within minutes, hours, and days respectively.

Each category was also separated based on documented evidence of communication. In the Red category, 5/15 reports had communication. In the Orange category, 33/141 reports had communication. In the yellow category 43/249 reports had communication. These results were then referenced against human evaluators.


Category RED Findings
10
7.5
5
2.5
0
Number of Reports
5
1
10
2
ACR Red with communication
ACR Red without communication
NLU findings
Human disagreement



Category ORANGE Findings
120
90
60
30
0
Number of Reports
33
4
108
18
ACR orange with communication
ACR orange without communication
NLU findings
Human disagreement


Category YELLOW Findings
240
180
120
60
0
Number of Reports
206
43
29
1
ACR yellow with communication
ACR yellow without communication
NLU findings
Human disagreement

Strengths and Limitations

Strengths: Able to identify the presence or absence of documented communication in a report with high degree of accuracy. Able to analyze a large volume of reports rapidly.

Limitations: Inconsistency in sometimes interpreting context in a given report. For example, qualifiers such as ‘healing’, ‘no interval change’, ‘no evidence of’, or ‘no evidence to suggest’ would sometimes lead the NLU algorithm to falsely categorize a report as containing actionable findings. Further validation and human input can refine the algorithm to better interpret clinical context.

Conclusion

Making a report machine intelligible can facilitate algorithmic assessment of report quality and automated tracking of specific variables such as critical findings within a large dataset of reports.

NLU technology can be used to identify both the actionable findings in imaging reports as well evidence of communication of those findings. This can be a potential quality control measure and could possibly result in improved application of ACR recommendations while ensuring more consistent documentation of actionable findings.

References

1. Larson, Paul A., Lincoln L. Berland, Brent Griffith, Charles E. Kahn Jr, and Lawrence A. Liebscher. 2014. “Actionable Findings and the Role of IT Support: Report of the ACR Actionable Reporting Work Group.” *Journal of the American College of Radiology: JACR* 11 (6): 552–58.