



# Framework for Extracting Critical Findings in Radiology Reports

Thusitha Mabotuwana<sup>1,2</sup> · Christopher S. Hall<sup>1,2</sup> · Nathan Cross<sup>2</sup>

Published online: 29 May 2020

© Society for Imaging Informatics in Medicine 2020

## Abstract

Critical results reporting guidelines demand that certain critical findings are communicated to the responsible provider within a specific period of time. In this paper, we discuss a generic report processing pipeline to extract critical findings within the dictated report to allow for automation of quality and compliance oversight using a production dataset containing 1,210,858 radiology exams. Algorithm accuracy on an annotated dataset having 327 sentences was 91.4% (95% CI 87.6–94.2%). Our results show that most critical findings are diagnosed on CT and MR exams and that intracranial hemorrhage and fluid collection are the most prevalent at our institution. 1.6% of the exams were found to have at least one of the ten critical findings we focused on. This methodology can enable detailed analysis of critical results reporting for research, workflow management, compliance, and quality assurance.

**Keywords** Critical results · Radiology reports · Rule-based systems · Radiology informatics

## Introduction

Radiology reports often contain imaging findings that are critical in nature and need to be communicated with a referring physician in a timely manner. Failure to do so could result in mortality or significant morbidity; however, timely communication of critical imaging finding is not uncommon and can lead to delayed treatment, poor patient outcomes, complications, unnecessary testing, lost revenue, and legal liability [1].

Communication of critical findings to referring providers is mandated by the Joint Commission's National Patient Safety Goal to "Improve the Effectiveness of Communication Among Caregivers" [2]. Although critical results communication is mandated, the Joint Commission does not specify an

exhaustive list of critical findings that need to be communicated and this decision is left up to the local treatment facility. As a result, compliance with critical findings reporting requirements is usually determined manually by radiologists performing peer-review to identify any non-compliant reports or manual text searches [3].

In an attempt to support the Joint Commission's mandate for critical results reporting, the American College of Radiology provides practice guidelines on communicating diagnostic imaging findings, emphasizes the timely reporting of critical findings, and recommends documentation directly in the radiology report [4]. The Actionable Reporting Work Group provides further details related to using information technology to provide timely communication of critical findings and provides a fairly exhaustive list of critical findings that would apply in most general hospital settings [5].

With a growing trend toward using deep learning-based technologies to automatically detect the presence of certain findings, including some critical findings (e.g., pneumothorax [6]), having the capability to automatically detect critical findings from radiology reports can also help create annotated datasets at scale. Currently, there are no standard ways to determine the presence of critical findings in reports and having such capabilities can facilitate a host of departmental functions including compliance monitoring and reporting to the Joint Commission. The first step toward improving workflow

✉ Thusitha Mabotuwana  
thusitha.mabotuwana@philips.com

Christopher S. Hall  
christopher.hall@philips.com

Nathan Cross  
nmcross@uw.edu

<sup>1</sup> Precision Diagnosis, Philips Healthcare, Bothell, WA 98021, USA

<sup>2</sup> Department of Radiology, University of Washington, Seattle, WA 98195, USA

efficiency around the resource-intensive, time-consuming nature of manual report review is to have reliable ways to automatically extract critical findings from radiology reports. Therefore, the primary goal of this study was to develop an extensible framework that can be used to detect the presence of critical findings in radiology reports.

## Methods

### Dataset

We extracted data for radiology exams performed between January 1, 2012, and December 31, 2014, from the University of Washington (UW) radiology information system (RIS). The dataset contained 1,210,858 exams performed across the entire multisite tertiary care enterprise, excluding any exams imported from outside of the institution. For each exam, the dataset contained the report text as well as several metadata fields, including exam code, exam date, subspecialty, patient setting (inpatient, outpatient, or emergency), and modality. This quality improvement project was approved by the UW Human Subjects Division as minimal risk and therefore exempt from Institutional Review Board (IRB) review. All data were stored on an encrypted machine within one of the secure data centers at the UW Medical Center with restricted user access.

The most common modalities in the dataset were radiography (XR) 622,367 (51.4%), computed tomography (CT) 229,058 (18.9%), ultrasound (US) 136,177 (11.2%), magnetic resonance imaging (MR) 89,298 (7.4%), mammography (MG) 66,329 (5.5%), nuclear medicine (NM) 22,344 (1.8%), fluoroscopy (FL) 21,051 (1.7%), interventional (XA) 18,118 (1.5%), and positron emission tomography (PT) 6116 (0.5%).

### Critical Findings

Using the list of critical findings put together by the Actionable Reporting Work Group as a guide, a board-certified radiologist (author NC) identified an initial list of important critical findings important to the radiology department. For this proof-of-concept research, we identified a list of 10 critical findings to focus on based on clinical significance and interest for the department. The list of selected findings was acute appendicitis, acute cholecystitis, acute ischemic stroke, brain herniation, fluid collection, acute intracranial hemorrhage, visceral laceration, pneumoperitoneum, pneumothorax, and acute pulmonary embolism.

As shown in Fig. 1, radiology reports typically contain several *sections*. Although the section names are not standardized across all institutions and dictation macros, the reports typically contain separate sections related to (1)

“Examination” to describe the type of radiology exam performed, (2) “Indication” to represent patient’s presenting signs, symptoms, and reason for examination, (3) “Comparison” to mention relevant prior studies that were used by the radiologist for comparison, (4) “Technique” to include the scanning methodology used, (5) “Findings” to describe any imaging findings observed by the radiologist, and an (6) “Impression” section to summarize the diagnostic interpretation of the observations. These sections could be combined (e.g., “Findings and Impression”) or omitted, and/or other section headers could be included depending on the nature of the exam and observed findings. Sometimes, there would be an “Addendum” section too, which includes information added after a finalized report, for a wide variety of reasons. A “Critical Result” section is an institution-specific addition that clearly identifies the case for future analysis and follow-up and documents the details of the communication of the critical findings.

The “Findings”, “Impression”, “Addendum”, and the “Critical Result” section, if present, were searched for the 10 identified critical findings. The University of Washington follows a Critical Red/Orange/Yellow nomenclature to identify findings that, per policy, require documented provider notification within 1 h, 12 h, or 48 h, respectively, according to the medical urgency of the finding [7]. The “Critical Result” section is added as a macro to the report by the dictating radiologist if one of these findings is identified. Current work does not rely on having a “Critical Result” section to ensure generalizability across hospitals.

### Negation, Index Cases, and Acuity

Although a radiology report may refer to a finding that can be critical, not all these occurrences are related to the first occurrence (i.e., not the index case), and if they are stable, they are not considered critical anymore and should not require additional radiologist effort. As a result, it is important to be able to distinguish between new findings versus references to existing findings (e.g., prior critical findings that are no longer critical). Also, the threshold at which a radiologist may decide a finding needs communication can vary, for instance, depending on the degree of enlargement of hemorrhage. To simplify analysis and ensure catching cases did change in a detrimental fashion, all interval increases in bleeding, pneumothorax, pneumoperitoneum, and herniation were classified as critical.

To identify index cases, we identified a list of terms commonly used to describe acuity of a finding (such as “new” and “acute”). Similarly, we also identified a list of terms routinely used to describe findings that are already known (e.g., “stable” and “resolved”). A list of negation-related terms was also identified to determine if a finding has been explicitly negated. Table 1 shows a subset of the acuity and negation-related terms that were identified.

CTA OF THE CHEST: TRAUMA AORTIC ARCH PROTOCOL

INDICATION:  
blunt trauma, abdominal tenderness.

COMPARISON:  
Same day CT abdomen and pelvis CT

TECHNIQUE:  
Reconstruction thickness: 1.25mm. Reconstruction interval: 1.25mm  
Superior Extent: Thoracic inlet. Inferior Extent: Adrenal glands  
...

FINDINGS:  
CHEST  
Heart: Heart size is upper limits of normal. No coronary artery calcifications.  
Aorta and great vessels: Normal.  
...

Right lung: Normal  
Left lung: Contusion in the lingula with multiple pulmonary lacerations (5/176) adjacent to multiple rib fractures (3/10). There also is mild groundglass opacity in posterior left lower lobe, which may represent atelectasis or contusion.

Right pneumothorax: Absent  
Left pneumothorax: Small left hemopneumothorax with pleural defect adjacent to fractures in the anterolateral seventh and eighth ribs. Average attenuation is 33 Hounsfield unit suggesting blood products in the pleural space.  
...

ABDOMEN:  
See abdominal CT for details.

IMPRESSION:

- Small left hemopneumothorax with adjacent rib fractures and lingular pulmonary laceration and contusion.
- No splenic laceration or intra-abdominal injury.

CRITICAL RESULT (ORANGE)

The critical result left hemopneumothorax was conveyed to Dr. xxx by xxx, M.D. at mm/dd/yyyy hh:mm AM.

**Fig. 1** A radiology report containing a critical result. Underlining is added for emphasis and is not present in the original report, “...” is used to shorten the report length where normal findings are discussed,

and “xxx” is used to mask physician names. Dates are output in a standard format as part of the macro used in order to allow easy parsing of results later

**Study Design**

We processed all 1,210,858 reports to identify the individual report sections (while also accounting for known variations in section titles—e.g., some reports may contain “Summary” instead of “Impression”, spacing, and punctuation), and

individual sentences within each section. The critical findings extraction algorithm then attempts to identify the critical finding, if any, at the sentence level.

We randomly sampled the larger dataset to create a working dataset containing 50,000 exams having the same modality distribution. This was used to perform various free text

**Table 1** Subset of the identified terms related to acuity and negation

Type	Term(s)
Critical finding	New, acute, prominent, increase, enlarged, critical, large, severe, moderate, unstable
Resolved finding	Unchanged, absent, prior, reduction, redemonstrated, stable, improved, old, chronic, shrinking, benign, resolved
Negated finding	No, without, absent, negative, none, rather than

queries to understand text variations, refine the list of critical findings, and develop algorithms. From this dataset, we created a smaller development set for manual analysis to establish ground truth so that we can determine algorithm accuracy on the development set. This was created by using keyword searches (due to low prevalence) to sample the 50,000 exam dataset until all critical finding types were represented with sufficient variability in the text. During the process, we developed the heuristic that the word “critical” or “acute” should appear anywhere in the report as initial filter prior to extracting any critical findings. To manage the annotation effort, we randomly selected sentences containing a critical finding until each of the 10 findings was represented with a minimum of 30 sentences. This positive training set contained 381 sentences. Another dataset having 912 sentences was created that referenced the findings of interest, but were not critical (e.g., “No pneumothorax or acute bony abnormality.”), in order to identify any cases that the algorithm may have missed, if any (i.e., false-negatives). Author NC ensured that the positive dataset contained true critical findings only, while the non-critical dataset contained only true-negatives.

To validate our algorithm’s ability to correctly identify critical findings mentioned in radiology reports, we created a validation set containing 327 sentences (giving a rough 80–20 split between annotated train and test sets) using the same approach that was used to create the development set, but using exams from the larger dataset that were not part of the 50,000 training dataset. Test set contained 96 positive findings and 231 negative findings. An overview of the data processing pipeline is shown in Fig. 2.

### Algorithm Development and Processing Pipeline

For the identified list of critical findings, we developed algorithms using a rule-based approach to detect the presence of a critical finding. The framework was developed to be easily extensible and allows a rule to contain synonyms, relevant anatomies, and (radiology) subspecialties as optional parameters. If these are specified, the rule is applied only to exams that have matching metadata. For instance, Table 2 shows the list of terms identified as relevant for intracranial hemorrhage. Implementation accounts for all variations in text as well, including special characters (e.g., *extra-axial* is the most common form, but this could be written with a space, or without a separation at all; similarly, variants of *increase* could be increased and increasing).

A sentence identified as having a critical finding is then evaluated for the presence of acuity and negation per phrases shown Table 1. If a sentence matches a rule defining a critical finding per Table 2, it is assumed to be critical unless a “Negated” or “Resolved” term per Table 1 appears in the sentence (e.g., sentence “There is a new large right-sided pneumothorax with mediastinal shift to the left.” will be

identified as having a critical finding and “new, large” will be included in matched metadata identifying criticality).

Each report contains multiple sentences, and establishing ground truth for a corpus containing 50,000 exams is not feasible. Therefore, we randomly “spot checked” the algorithm output against sentences from the training set having 50,000 exams and improved the algorithm accuracy iteratively, only implementing rules that could be generalizable to avoid potential overfitting issues.

Algorithm output on the training set contained 343 true-positives, 866 true-negatives, 38 false-negatives, and 46 false-positives, resulting in 93.5% (95% CI 92.0–94.8%) accuracy, 90.0% (95% CI 87.4–92.2%) sensitivity, and 95.0% (95% CI 93.9–95.9%) specificity. Algorithm output by finding on the training set is shown in Table 3.

### Algorithm Validation

Algorithm performance on the test set containing 327 sentences was 91.4% (95% CI 87.6–94.2%) accuracy, 84.4% (95% CI 77.9–89.1%) sensitivity, and 94.4% (95% CI 91.7–96.4%) specificity. There were 15 false-negatives related to incorrect negation detection, and 13 false-positives that were primarily due to long, comma-separated complex sentences that were referencing prior critical findings. Algorithm output by finding on the test set is shown in Table 4.

### Results

There were 19,966 (1.3%) out of 1,210,858 reports that contained at least one type of a critical finding. Table 5 shows the distribution of critical findings by scanned modality.

Table 6 shows the distribution of the findings in our dataset. Note that an exam can contain multiple critical findings, although this is not so common (resulting in sum of exams by finding being greater than total number of exams).

Based on Table 5, most of the critical findings are observed on CT and MR exams. Out of the findings we considered, intracranial hemorrhage and fluid collection were the most prevalent.

We also explored the distribution of critical findings by modality. For this, we considered only findings that occurred a minimum of 20 times for a given modality. Table 7 shows the top two finding for each of the scanning modalities. The percentages are based off the total number of occurrences of the given condition. For instance, 1436 pneumothorax findings on CR exams account for 52.5% of all pneumothorax cases (2736 per Table 6). Similarly, we can see that 44.1% of the acute strokes are diagnosed on MR exams.

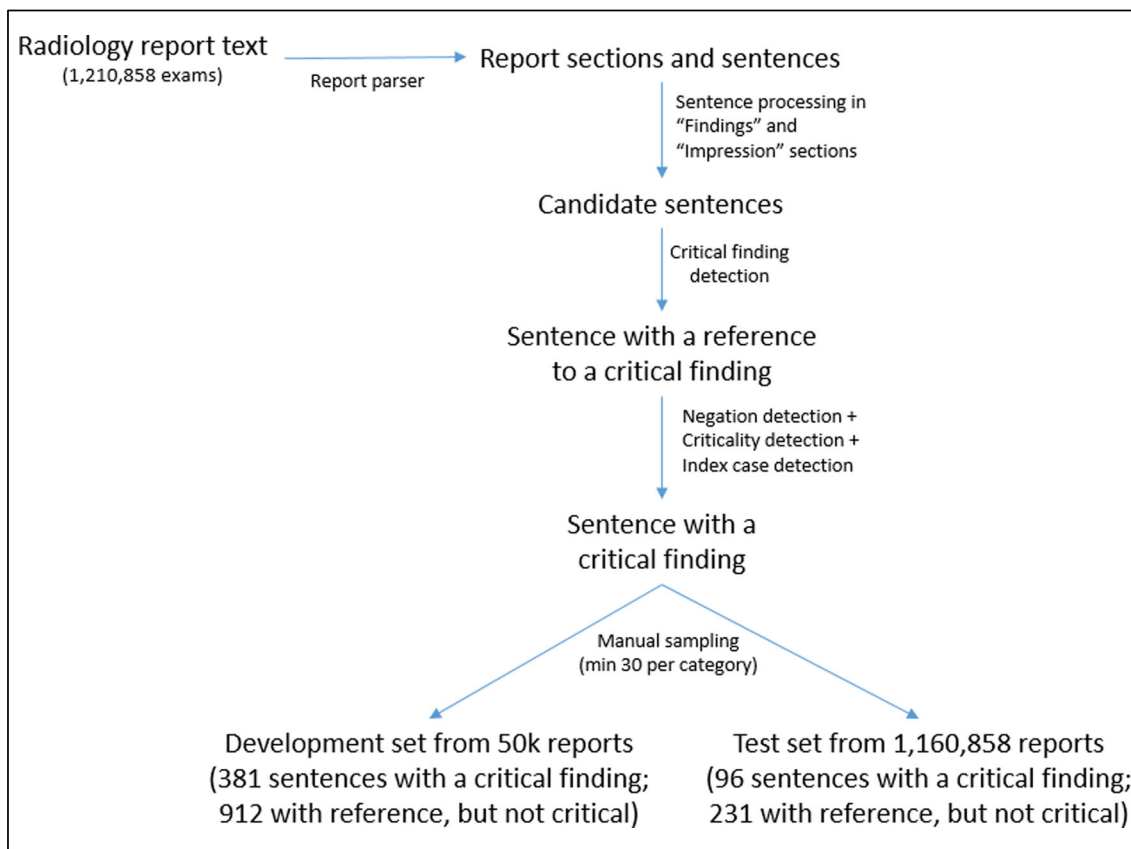


Fig. 2 Overview of data processing pipeline

### Discussion

Using a production dataset containing over a million exams, we demonstrated how a generic radiology report processing pipeline can be developed to extract a range of critical findings. A key strength of this work is the extensibility of the framework to other findings, and the control granularity provided to create specific extraction rules. To the best of the authors’ knowledge, this is the first scalable attempt to report on the prevalence of multiple critical findings observed in radiology reports.

Our results show that most of the critical findings are diagnosed on CT and MR exams and that intracranial hemorrhage and fluid collection are the most prevalent. This work can

enable radiologists and administrators to better understand the prevalence of critical results by modality and subspecialty; enable a range of quality assurance initiatives, such as automatically detecting if there is documentation of a critical result communication to a provider per departmental guidelines; and identify where improvements to critical findings communication-related procedures are needed.

Despite using a production dataset containing data from multiple network hospitals, the current study has several limitations. *First*, all reports were created using common dictation macros that are shared across the network hospitals, and therefore, the rule-based methods used to parse radiology reports and identify findings may serve as a model but may not be immediately generalizable to other institutions, even

Table 2 Term list for intracranial hemorrhage

Type	Term(s)
Term (required)	Intracranial Hemorrhage
Synonyms (optional)	Hemorrhage, hemmorage (common misspelling), bleeding, hematoma, ruptured aneurysm, impending rupture
Anatomies (optional)	Intracranial, intracerebral, spinal, intraparenchymal, subarachnoid, intraventricular, intramedullary, subdural, epidural, retroperitoneal, extra-axial
Subspecialty (optional)	Neuro, Emergency

**Table 3** Algorithm performance on training set by finding (FN false-negative, FP false-positive, TN true-negative, TP true-positive)

Finding	FN	FP	TN	TP	Total
AcuteAppendicitis	1	4	33	37	75
AcuteCholecystitis	3		47	29	79
AcuteStroke	5	8	362	43	418
BrainHerniation	3		30	20	53
FluidCollection	7	5	48	37	97
IntracranialHemorrhage	11	16	159	69	255
Laceration	2	1	32	21	56
Pneumoperitoneum	2	5	72	37	116
Pneumothorax	2	6	64	39	111
PulmonaryEmbolism	2	1	19	11	33
Total	38	46	866	343	1293

**Table 4** Algorithm performance on test set by finding (FN false-negative, FP false-positive, TN true-negative, TP true-positive)

Finding	FN	FP	TN	TP	Total
AcuteAppendicitis	–	–	9	9	18
AcuteCholecystitis	–	–	12	8	20
AcuteStroke	2	6	88	10	106
BrainHerniation	1	–	8	5	14
FluidCollection	2	1	12	9	24
IntracranialHemorrhage	6	2	42	14	64
Laceration	1	1	9	5	16
Pneumoperitoneum	–	1	18	10	29
Pneumothorax	1	2	15	9	27
PulmonaryEmbolism	2	–	5	2	9
Total	15	13	218	81	327

though our dataset contained over 1.2 million reports created by over 150 different radiologists across different network hospital settings. Given that we are using key terms with synonyms from a relatively large academic center, we expect to have covered all terms and variations reasonably well, although this requires further validation to ensure scalability. *Second*, the algorithm has limitations when evaluating the clinical context of an exam, as observed by the false-positives and false-negatives we encountered. For example, the algorithm failed with several complex sentences, for instance, “The left cerebellar infarct is resulting in mass effect on the inferior aspect of the fourth ventricle which is increased since the prior exam, however, without development of hydrocephalus.” refers to an acute stroke; however, the algorithm failed due to the reference to prior exam (i.e., a false-negative). Another example is “There is a small eccentric filling defect in a right lower lobe segmental pulmonary artery (4/93), which may represent pulmonary embolus, possibly chronic, especially given patient’s known history of DVT.” where the ground truth was marked as critical due to clinical context, although the algorithm determined that this is non-critical due to the reference to “chronic.” Similarly, “A paraspinous and mediastinal hematoma is noted posterior and lateral to the trachea.” was flagged as having a critical finding (intracranial hemorrhage), although the context suggested that this is not a critical finding, resulting in a false-positive. Therefore, it is possible that our results may not represent the true prevalence rates of these conditions. As such, improvements to the algorithm are needed to take more clinical context into consideration. *Third*, we considered any increase/enlargement in blood, gas, and so on as a clinically significant finding. This is a conservative and reasonable approach to flagging cases, although this may not always be the case. Therefore, further refinement on clinical scenarios is needed. *Fourth*, we have not accounted for any dictation, transcription,

spelling, or other human factors that could affect the accuracy of the report content, although we do not expect such factors to have any large systemic impact based on our review of cases. *Fifth*, the primary use case of interest for our current work was to use retrospective reporting data to build ground truth annotations at scale so that imaging exams with critical findings can be reliably identified for various downstream tasks such as worklist prioritization. Therefore, detecting “criticals” was more important than missing some cases with a positive finding for this use case. The observed sensitivity of 84.4% (95% CI 77.9–89.1%) and specificity of 94.4% (95% CI 91.7–96.4%) was a good tradeoff for our purposes, although for applications more oriented toward clinical diagnosis, sensitivity may be more important. *Lastly*, our algorithm validation was performed on a relatively small dataset, and further validation of the algorithm is necessary in order for the results to be generalizable. One practical approach to do this is to run the algorithm on a larger training set and then

**Table 5** Exams with a critical finding by scanned modality

Modality	No. of exams	No. of exams with a critical finding
CT	229,058	14,318 (6.3%)
MR	89,298	3202 (3.6%)
PT	6116	32 (0.5%)
US	136,177	432 (0.3%)
CR	622,367	1895 (0.3%)
XA	18,118	41 (0.2%)
NM	22,344	31 (0.1%)
FL	21,051	14 (0.1%)
MG	66,329	1 (0.0%)
Total	1,210,858	19,966 (1.6%)

**Table 6** Distribution of exams by finding

Finding	No. of exams	No. of exams with a critical finding ( <i>n</i> = 19,966)
IntracranialHemorrhage	5927	29.7%
FluidCollection	4702	23.6%
AcuteStroke	4496	22.5%
Pneumothorax	2736	13.7%
PulmonaryEmbolism	1465	7.3%
Pneumoperitoneum	673	3.4%
AcuteCholecystitis	574	2.9%
AcuteAppendicitis	556	2.8%
Laceration	412	2.1%
BrainHemiation	398	2.0%
Total	21,939	

focus on false-negatives and false-positives to refine the algorithm, while ensuring that accuracy is not adversely affected.

The generic processing pipeline we have developed can be used to identify a range of critical results routinely reported in radiology reports. Pattern and rule-based approaches have been reported to have high accuracy for classifying critical results in unstructured radiology reports for specific findings [8], and our rule-based approach is comparable to what several other researchers have proposed [8, 9]. Machine learning-based approaches have been used as well [10] with comparable results, but the fast, scalable approach we have taken while providing the flexibility to control granularity through rule customization is novel. Compared to machine learning techniques, the methodology is also readily understandable and much easier to optimize for an intended goal. There are several potential applications of having the means to automatically detect critical results from radiology reports. For instance, it provides valuable insights for radiology quality improvement initiatives such as compliance with timely critical results reporting policies. Also, with the rapid growth of deep learning-based techniques to automatically detect various radiological findings [11, 12], having the means to extract findings can provide a mechanism to establish ground truth data to train and improve accuracy of these learning algorithms.

This data also provides valuable data for a critical result management QI/QA process. Many institutions already have a critical result reporting and/or tracking policy in place allowing radiologists to manually identify cases with critical results. However, it is often unclear what fraction of cases is not getting entered into this system, and therefore, incidence rates are often under-reported. The data generated from this tool will have a variety of uses, including (1) characterize the distribution of critical results over types of exams and stratify exams by frequency of critical results helping to target areas with greatest incidence of these findings; (2) provide baseline data for compliance with departmental reporting guidelines and allow for remediation to ensure uniformity of usage of

the critical result reporting process that is currently in place; (3) follow-up on identified cases to ensure that the complete cycle from critical result reporting, to notification, to management, and subsequent follow-up is achieved; and (4) provide comparison data to hospital administration between the patient groups that had a critical result finding and were reported through the existing departmental critical results workflow and those that did not.

## Conclusions

The generic radiology report processing pipeline we have developed to extract a range of critical findings has the potential to enable detailed analysis of critical results reporting for research, workflow management, compliance, and quality assurance.

**Table 7** Top two critical findings by modality

Modality	Finding type	No. of exams
CR	Pneumothorax	1436 (52.5%)
CR	Pneumoperitoneum	296 (44.0%)
CT	IntracranialHemorrhage	5188 (87.5%)
CT	FluidCollection	3728 (79.3%)
MR	AcuteStroke	1984 (44.1%)
MR	IntracranialHemorrhage	725 (12.2%)
NM	AcuteCholecystitis	25 (4.4%)
PT	FluidCollection	25 (0.5%)
US	AcuteCholecystitis	262 (45.6%)
US	FluidCollection	123 (2.6%)
XA	FluidCollection	33 (0.7%)

**Acknowledgements** Authors would like to acknowledge the contributions from Dr. Martin Gunn (Professor, and Vice Chair of Informatics at UW Radiology) for all his support and guidance on this work.

## References

- Gordon, J.R., et al., Failure to recognize newly identified aortic dilations in a health care system with an advanced electronic medical record. *Ann Intern Med*, 2009. **151**(1): p. 21-7, **W5**.
- The Joint Commission, National Patient Safety Goals Effective January 2018. 2018.
- Khorasani, R., Optimizing communication of critical test results. *J Am Coll Radiol*, 2009. **6**(10): p. 721-3.
- American College of Radiology, ACR Practice Parameter for Communication of Diagnostic Imaging Findings. 2014.
- Larson, P.A., et al., Actionable findings and the role of IT support: report of the ACR Actionable Reporting Work Group. *J Am Coll Radiol*, 2014. **11**(6): p. 552-8.
- Rajpurkar P. I. J., Zhu K, Yang B, Mehta H, Duan T, et al. *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*. 2017 [cited 2020 Feb 2]; Available from: <https://arxiv.org/abs/1711.05225>.
- Critical Results Reporting. 2011 [cited 2020 Jan 25]; Available from: [http://depts.washington.edu/pacshelp/docs/Crit\\_Res\\_rpt.html](http://depts.washington.edu/pacshelp/docs/Crit_Res_rpt.html).
- Lakhani, P., W. Kim, and C.P. Langlotz, Automated detection of critical results in radiology reports. *J Digit Imaging*, 2012. **25**(1): p. 30-6.
- Lacson, R., et al., Retrieval of radiology reports citing critical findings with disease-specific customization. *Open Med Inform J*, 2012. **6**: p. 28-35.
- Tiwari, A., et al., Automatic classification of critical findings in radiology reports, in Proceedings of The First Workshop Medical Informatics and Healthcare held with the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining, F. Samah and R. Daniela Stan, Editors. 2017, PMLR: Proceedings of Machine Learning Research. p. 35–39.
- Chang, P., et al., Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *AJNR Am J Neuroradiol*, 2018. **39**(7): p. 1201-1207.
- Lakhani, P. and B. Sundaram, Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 2017. **284**(2): p. 574-582.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.