INTRODUCTION – WHY RadQA?

• Machine reading comprehension (MRC) is widely explored to better comprehend unstructured text, by enabling machines to answer specific questions given a textual passage.

• Much attention in MRC drawn toward biomedical scientific articles.

• Limited work toward building a challenging MRC dataset for electronic health record (EHR) data.

• Most existing datasets are small and/or publicly unavailable to build advanced models.

• Question collection in most datasets includes bias and does not reflect real-world user needs.

• Almost all datasets use discharge summaries.

• Thus, we propose RadQA, a new EHR MRC dataset.

FINAL REPORT

INDICATION: 64 year old male with status post recent STEMI. Now with increasing edema and shortness of breath.

FINDINGS: The heart is (enlarged in size) but stable in the interval. Mediastinal contour is unaltered. There is upper zone redistribution of the pulmonary artery vasculature. Perihilar haziness as well as diffuse bilateral pulmonary opacities. These findings are consistent with acute CHF.

There are also bilateral pleural effusions. There is hawking in a single view from previous study.

 IMPRESSION: 1. Findings consistent with pulmonary edema due to CHF. 2. Bilateral pleural effusions.

Q: Are there any infiltrates in the lung? A: Bilateral pleural effusions.

Q: Did the cardiac silhouette enlarge? A: No.


Tab 1. RadQA example. Fndg – Findings. Lmp – Impression.

RadQA DATASET

• Questions reflect true information needs of clinicians (inspired from the clinical referral section of radiology reports).

• Contains 3074 unique question-report pairs for 1009 radiology reports

• Each question has at least two answers for a report (in its Findings and Impressions sections), resulting in 6148 distinct question-answer evidence pairs (including unanswerable questions)

• Answers are often times phrases or span multiple lines

• Questions require wide variety of reasoning & domain knowledge to answer

Tab 2. Existing MRC datasets. UN-Q – Unanswerable questions.

Table 3. Reasoning categories in RadQA.

Tab 3. Reasoning categories in RadQA.

<table>
<thead>
<tr>
<th>Reasoning</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Variation (Syntactic)</td>
<td>Key links between words are synonyms</td>
<td>Q: Was the PCC line placed correctly?</td>
</tr>
<tr>
<td>Lexical Variation (world/medical knowledge)</td>
<td>Key links between words are domain or medical knowledge</td>
<td>Q: Is there any obstruction in the lungs?</td>
</tr>
<tr>
<td>Synoptic Variation</td>
<td>Declarative form of options does not syntactically match the anat sentences</td>
<td>Q: Are there any fractures in the pelvis?</td>
</tr>
<tr>
<td>Confusion</td>
<td>Anaphora or intra-sentence framing</td>
<td>Q: 1. Was the PCC placed?</td>
</tr>
<tr>
<td>Incomplete Context</td>
<td>Missing contextualization in anat sentences</td>
<td>Q: We find any atrophic in the cardiac arteries that require getting bi-atrial CABG?</td>
</tr>
<tr>
<td>Change Information</td>
<td>Questions related to internal changes</td>
<td>Q: Has thyoid gland progressed?</td>
</tr>
<tr>
<td>Diagnosis Knowledge</td>
<td>Questions require diagnostic understanding to anat</td>
<td>Q: Are there signs of pericarditis?</td>
</tr>
<tr>
<td>Anatomy Knowledge</td>
<td>Questions require anatomy understanding to anat</td>
<td>Q: Did the gastric cancer metastasize to chest?</td>
</tr>
<tr>
<td>Require Specification</td>
<td>Questions require specific information in anat</td>
<td>Q: What is the status of the sclera through iridescence?</td>
</tr>
<tr>
<td>No Answer</td>
<td>Ans is present but negated</td>
<td>Q: Is there any medial shift due to pneumothorax?</td>
</tr>
</tbody>
</table>

CONCLUSION

• The performance of the best transformer language model, MIMIC-BERT, is 63.55 (F1), which falls significantly short of the best human performance of 90.31.

• This demonstrates the challenging nature of RadQA that leaves ample scope for future research.

ACKNOWLEDGEMENTS

U.S. National Library of Medicine, National Institutes of Health, (R01LM012104), National Institute of Biomedical Imaging and Bioengineering (R21EB029575); and UTHHealth Innovation for Cancer Prevention Research Training Program Predoctoral Fellowship (CRIPTR RP210042).

REFERENCES

