HIDDEN-RAD:
Hidden Causality Inclusion
in Radiology Report Generation

Key-Sun Choi†, Young-Gyun Hahm‡
J.H. Hahn †, S.Y. Lee○, Y.J. Kang‡
†Konyang Univ., ‡Teddysum Inc., ○Konyang Univ. Hospital

NTCIR-18 Pilot Task
Hidden-RAD: Purpose with example

To Recover the Hidden Causality why such Impression and Findings are drawn in the Radiology Report

Radiography (option)

Radiology report

Impression:
Pneumothorax
Finding:
The pneumothorax in this case may be attributed to a combination of factors, including trauma and anatomical location. The right pneumothorax observed at the T8-11 thoracic spine level in the right pleural space indicates a localized issue in the upper to middle region of the right lung.

Hidden causality:
The lack of symmetry in the apical, upper, middle, and lower zones suggests an asymmetric distribution of air in the pleural space, further confirming the presence of pneumothorax.
Hidden-RAD: dataset overview

- Data collection: Decision of Steps in Reading Process of real doctors.
- Hidden-RAD dataset: generated report from the collected data

The reading process was annotated with questions and answers to reduce costs, and GPT-4 was used to generate the dataset into a report format.
Task Definition

Recovering report by discovering hidden (missing) causality in radiology reports.

Task Input
- Image (option)
- Radiology report (including the finding and impression)
- Additional data available (KB, etc.) (option)

How to discover and recover hidden causality of radiology reports.

Method (task participants)

Task Output
- Radiology report to include/add Hidden Causality

Ground Truth (Hidden-RAD dataset)

Evaluation
- BLEU
- ROUGE
- METEOR
- BERT Score
- And GPT-based evaluation (TBD)

A report making from a checklist obtained from a doctor's qualification exam.

A report expressing the causality of an impression that the doctor could not express.
Confirmation checklist: ABCDE approach

• 32 checklists,
• **Chest x-ray review** is a key competency for medical students, junior doctors and other allied health professionals.
• Using A, B, C, D, E is a helpful and systematic method for chest x-ray review:

- **A: airways** – 5 checklist
- **B: breathing** (the lungs and pleural spaces) - 11 checklist
- **C: circulation** (cardiomedialstiinal contour) - 5 checklist
- **D: disability** (bones - especially fractures) - 6 checklist
- **E: everything else**, e.g. pneumoperitoneum - 5 checklist
Terminology and Definition 1/2

• Radiology Report
  – A detailed report that describes the results of an imaging test.
  – A radiology report includes information about the type of imaging test that was done and how it was done. (A brief medical history of the person having the test, including any symptoms or known diseases and why the test was needed)

• Interpretation (or “reading”)
  – Radiologic interpretation is based upon the understanding of disease processes and the behaviour of diseases in a specific anatomic region.
  – Combined with knowledge of the specific radiological characteristics of various lesions, the main features of the lesion can be identified and contribute substantially to the diagnosis.

Terminology and Definition 2/2

- **CAD (Computer-aided diagnosis)**
  - CAD is a system that assist doctors in interpreting medical images to produce a lot of information that needs to be analyzed and evaluated in a short time.

- **“Hidden Causality”**
  - In this task, Required conditions for the final impression as determined by the doctor
  - In this study, hidden causality was limited to 28 checks in the chest radiography interpretation procedure and does not imply all causal relationships in the interpretation.
  - The hidden causality was manually annotated by doctors.
Task summary

Task background

- If the hidden cause is not sufficiently included in the report, it may cause errors that would worsen the patient’s illness.
- Errors according to:
  - Limitations of the perceptual capacity of the human eye and brain to perform interpretation
  - Differences in the scope and quality of background knowledge

Task Definition

- Recovering report by discovering hidden (missing) causality in radiology reports.
- Extraction of mechanical features from images to resolve perceptual errors and learning of qualification test lists to resolve knowledge gaps.

Dataset (Hidden-RAD dataset)

- “Hidden-RAD dataset” is a set of radiology reports including Findings, Impressions, and Hidden causality.
- The current data was augmented based on MIMIC-CXR and automatically generated in report format through GPT-4.

Comparison

- No dataset containing hidden causality for lesions.
- This dataset was augmented based on MIMIC-CXR.

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Release year</th>
<th>Size</th>
<th>Resource</th>
<th>Hidden causality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden-RAD dataset (Our)</td>
<td>2024</td>
<td>300</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>PadChest</td>
<td>2020</td>
<td>206,222</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>MIMIC-CXR</td>
<td>2019</td>
<td>371,920</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>CheXpert</td>
<td>2019</td>
<td>224,316</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>National Institutes of Health Repository</td>
<td>2017</td>
<td>112,120</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PLCO</td>
<td>2012</td>
<td>185,421</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Expected result (Contribution)

- Explainable CAD technology.

- Reducing diagnostic error
  - Hidden causality is included in the report through systematic interpretation of the images. This is a complete scan of the radiograph and contributes to reducing the number of diagnostic errors.

Data release plan (2024-2026)

Data for User-Adaptable Explanation Report Generation of Medical Interpretation and Diagnosis

- **Hidden Interpretation Discovery Data**
  - 2024.5
    - HIDDEN-RAD-Chest-1
      - 300 MIMIC CXR base
    - 2024.11
      - HIDDEN-RAD-Chest-2
        - 1000 MIMIC CXR base+sync
    - 2025.11
      - 2000
    - 2026.11
      - >3000

- **Image source license**
  - PhysioNet 1.5.0

- **Hidden-RAD Data license**
  - CC BY-NC-SA

- **Knowledge Graph**
  - KG enriched

- **Medical expert Crowdsourcing**
  - crowdwork log with report ref.
  - crowdwork log w/o report ref.

- **Report Generation Task**
  - KG+ generative
  - classific
cation

- **QA Task**
  - knowledge +LLM

- **Integrated interpretation task**

- **CC BY-NC-SA**

- **CC BY-NC-SA**

- **Knowledge Graph + generative classific**
(The initial) Early experimental results

- Use the same input for tasks
- Perform analysis through GPT-V

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU-1</th>
<th>BLEU-2</th>
<th>BLEU-3</th>
<th>ROUGE-E-1</th>
<th>ROUGE-E-2</th>
<th>ROUGE-E-I</th>
<th>METEOR</th>
<th>BERT Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPT-V</td>
<td>25.41</td>
<td>11.32</td>
<td>7.11</td>
<td>21.71</td>
<td>7.14</td>
<td>1.69</td>
<td>0.19</td>
<td>0.34</td>
</tr>
</tbody>
</table>
Important dates

- Kick-off Event: March 29, 2024
- 1st-Training Set Release: May 31, 2024 / Quantity: 300 cases
- 2nd-Training Set Release: Jun 30, 2024 / Quantity: 300 cases
- Test Set Release: Dec 8, 2024 / Quantity: 400 cases
- System Submission Deadline: Jan 8, 2025
- Evaluation Results Release: Feb 1, 2025
- Submission Due of Participant Papers (draft): March 1, 2025
- Camera-Ready Participant Paper Due: May 1, 2025
- NTCIR-18 Conference: Jun 10-13, 2025