Automated Report Checker

Vineet Khanna, MD
Joshua Polster, MD
Authors and Disclosures

• Vineet Khanna, MD
  – No Disclosures

• Joshua Polster, MD
  – No Disclosures
Purpose- Background

• Radiology report represents the primary mode of communication between radiologists and referring clinicians.
  – Report accuracy important for patient care

• Prior studies have shown that up to 22% of final reports with automatic voice recognition software contain substantive errors

• Types of Errors
  – Substitution errors (“acute cirrhosis” instead of “arteriosclerosis”)
  – Nonsense phrases/errors (“interest and it is” instead of “infraspinatus”)
  – Omission errors (dictated word is not transcribed)
Purpose

• We introduce and test a user-customizable software program designed to detect and highlight context-specific errors in radiology reports, facilitating the radiologist's efficiency and effectiveness in proofreading
Materials/Methods

• Software was created to identify the context of an imaging report then evaluate the report language for context appropriateness

• First phase- pilot
  – Assess 20 finalized MRI reports of 1 MSK radiologist for errors
    – 2\textsuperscript{nd} radiologist review reports as reference standard
    – Run software on reports using standard context rules (not user specific) to see accuracy of detecting errors
  – Assess 50 finalized radiographic reports of 1 MSK radiologist for editing time
    – Attending radiologist standard manual review versus review with software aid
Materials/Methods

• Second phase- broader scope
  – 47 finalized Shoulder MRI reports
  – 51 finalized Hip Radiographs
  – Reports for each were gathered from 3 separate musculoskeletal radiologists

• Two musculoskeletal radiologists in consensus reviewed the reports manually proofreading for errors as reference standard

• Report checking software then run on reports to highlight errors using standard context rules (not user specific)

• Comparison of highlighted errors to reference standard performed to determine sensitivity and specificity of software
Materials/Methods

• Third phase- prospective use

• 14 MRI reports generated by a single musculoskeletal radiologist with user-specific language learning
  
  • 3 Pelvis/Hip, 3 Foot, 1 Elbow, 4 Shoulder, and 3 Wrist MRIs
    – Prior to editing, preliminary report was saved
    – Software used to highlight and this highlighted report was saved
    – Using the aid of the highlighted report, radiologist self-edited reports
    – Comparison of highlighted preliminary report to final report to generate accuracy data (sensitivity and specificity)
HISTORY: Ankle pain

TECHNIQUE: MRI right ankle

COMPARISON: None

RESULT:

There is a tubular left increased T2 signal and decreased T1 signal paralleling the posterior margin of the posterior calcaneal tuberosity consistent with an insufficiency fracture. There is moderate degree of surrounding ill-defined bone marrow edema. No other fractures are seen. The anterior talofibular ligament demonstrates moderate thickening and mildly increased signal compatible with a low-grade sprain. The posterior talofibular, tibiofibular, calcaneofibular and deltoid ligaments are intact. Knee extensor tendons of the ankle are intact and normal appearing. The flexor digitorum longus, posterior tibialis and flexor hallucis longus tendons are intact. Peroneal tendons are mildly thickened compatible with tendinosis. Mild peroneal tenosynovitis. The Achilles tendon demonstrates mild tendinosis. The plantar aponeurosis is within normal limits. There is a physiologic quantity of joint fluid in the ankle. No other significant abnormality.

IMPRESSION: Calcaneal insufficiency fracture

Low-grade sprain of the anterior tilting of the ligament.
Sample report - software highlighting

True positive w/ automatic correction by software (initial report: “flexor houses longus”)

False positive

True positive

True positive
Sample report- color highlighting to distinguish errors from substitutions

RESULT:

There is a tubular left increased T2 signal and decreased T1 signal paralleling the posterior margin of the posterior calcaneal tuberosity consistent with an insufficiency fracture. There is moderate degree of surrounding ill-defined bone marrow edema. No other fractures are seen. The anterior talofibular ligament demonstrates moderate thickening and mildly increased signal compatible with a low-grade sprain. The posterior talofibular, tibiofibular, calcaneofibular and deltoid ligaments are intact. Knee extensor tendons of the ankle are intact and normal appearing. The flexor digitorum longus, posterior tibialis and flexor hallucis longus tendons are intact. Peroneal tendons are mildly thickened compatible with tendinosis. Mild peroneal tenosynovitis. The Achilles tendon demonstrates mild tendinosis. The plantar aponeurosis is within normal limits. There is a
Results

• Initial pilot
  – Effectiveness:
    – 35 substantive errors (16 of the 20 MRI reports)
    – Software detected 33 of 35 substantive errors (94%)
  – Efficiency:
    – 50 radiograph cases with manual proofreading
      – 56 minutes and 53 seconds (54.4 words/report).
    – 50 radiograph cases proofread with the aid of software
      – 44 minutes and 32 seconds (50.1 words/report).
## Results

- **Second phase**—Retrospective review of finalized reports for 47 shoulder MRIs and 51 Hip radiographs (excluding Omission errors)
  - Sensitivity: 0.56
  - Specificity: 0.98

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Substitution Errors</th>
<th>Nonsense Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positives</td>
<td>28</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>False Negatives</td>
<td>19</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>True Negatives</td>
<td>12325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Positives</td>
<td>296</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Omission Errors**: 10
Results

- Third phase -- Prospective review of 14 MRIs (Pelvis/Hip, elbow, shoulder, and wrist MRIs) prior to manual proofreading evaluated with report checker (excluding omission errors)
  - Sensitivity: 0.88
  - Specificity: 0.99

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Substitution Errors</th>
<th>Nonsense Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positives</td>
<td>16</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>False Negatives</td>
<td>2</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>True Negatives</td>
<td>2078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Positives</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

• Errors missed in manual editing can be detected with the aid of software with variable sensitivity (0.56-0.94)

• User specific language learning is more accurate than fixed standard language model

• Substitution errors most common error type found by software
  — (86% of true positive errors found)

• Commonly Missed Errors
  — Nonsense Phrases
  — Omission errors (not part of software design)

• Editing speed can be improved with software aided error detection without sacrificing accuracy
Conclusions

- Future direction
  - Applying software across a greater breadth of modalities and subspeciality radiologists