ACCURACY OF NATURAL LANGUAGE PROCESSING FOR IDENTIFYING SPONDYLOARTHROPATHY IN PRIMARY CARE PATIENTS RECEIVING LUMBAR SPINE IMAGING

MYCHAEL LAGBAS, MD
UNIVERSITY OF WASHINGTON
DEPARTMENT OF REHABILITATION MEDICINE
AUTHORS

Pradeep Suri MD, MS
Jeffrey G. Jarvik MD, MPH
Sean D. Rundell DPT, PhD
W. Katherine Tan, BS
Rini Desai, MD
Patrick J. Heagerty, PhD
Hannu T. Huhdanpaa MD, MS
Kathryn James PA, MPH
Nancy L. Organ, BA
Andrew Avins MD, MPH
Karen J. Sherman PhD, MPH
David F. Kallmes, MD
Patrick H. Luetmer, MD
Brent Griffith, MD
David R. Nerenz, PhD
Jeremy Paige, MD
I have no financial disclosures
INTRODUCTION

• Natural language processing (NLP) is the computational processing of common language

• The accuracy of NLP in identifying spondyloarthropathy in radiology imaging reports is unknown
OBJECTIVE

To determine the accuracy of an NLP algorithm for identifying spondyloarthropathy in lumbar spine imaging reports.
METHODS

STUDY SAMPLE
• An “imaging set” was created of lumbar spine imaging reports
• Set was enriched for spondyloarthropathy ICD 9/10 codes

REFERENCE STANDARD
• Two clinicians evaluated each report for the presence or absence of spondyloarthropathy
• Discrepancies were resolved by discussion
  • Adjudication by a neuroradiologist when necessary
METHODS

NLP PERFORMANCE

Radiology Text Reports (N=2592)

Segmentation: Section segmentation, sentence segmentation

Normalization: spelling error, spelling variations, word stemming

Concept Identification: Regular expressions

Negation Identification: NegEx algorithm

Features: N-grams, section, regex, negex, site, modality

Rule-based: Regular expressions with negation detection

Model Predictions: Dichotomous

Metrics: Sensitivity, Specificity, PPV, NPV
STATISTICAL ANALYSIS

- Performance characteristics were estimated, with inverse probability weighting to account for sample enrichment
  - Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value
**RESULTS**

- Prevalence of spondyloarthropathy was 12%

**TABLE 1: Study sample with and without spondyloarthropathy**

<table>
<thead>
<tr>
<th></th>
<th>Imaging reports without Spondyloarthropathy</th>
<th>Imaging reports with Spondyloarthropathy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of studies</strong></td>
<td>2292</td>
<td>302</td>
</tr>
<tr>
<td><strong>Mean age (sd)</strong></td>
<td>66.22 (15.49)</td>
<td>60.49 (17.64)</td>
</tr>
<tr>
<td><strong>Image type (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>XR</em></td>
<td>817 (35.6)</td>
<td>251 (83.1)</td>
</tr>
<tr>
<td><em>CT</em></td>
<td>54 (2.4)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td><em>MRI</em></td>
<td>1421 (62.0)</td>
<td>49 (16.2)</td>
</tr>
<tr>
<td><strong>Gender (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Female</em></td>
<td>862 (40.1)</td>
<td>124 (42.8)</td>
</tr>
<tr>
<td><em>Male</em></td>
<td>1289 (59.9)</td>
<td>166 (57.2)</td>
</tr>
<tr>
<td><em>Unknown</em></td>
<td>1 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>
### TABLE 2: NLP Performance Characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Percentage (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>95% (92-97%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>98% (98-99%)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>91% (88-94%)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>99% (99-99.9%)</td>
</tr>
</tbody>
</table>
CONCLUSIONS

• NLP has high diagnostic accuracy
  • all performance characteristics >90%

• NLP may be a useful tool for identifying specific imaging findings in large datasets

• Potential applications for future research or clinical care
LIMITATIONS

1. These performance characteristics reflect the case-enriched sample
   - may not be generalizable to the general population

2. Incidentally, “spondyloarthropathy” was noted to be used inappropriately
   - reference to spondylosis rather than true inflammatory arthritis

UW Medicine
DEPARTMENT OF REHABILITATION MEDICINE
This work was funded by the following grants from the National Institutes of Health (NIH):

- UH2 AT007766
- UH3 AR066795
REFERENCES

1. Jarvik, J. et al. Lumbar Imaging with Reporting of Epidemiology (LIRE) Protocol for a Pragmatic Cluster Randomized Trial. Contemporary Clinical Trials. 2015 45(0 0):157-163; LOE: N/A