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

UW Medicine
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

UW Medicine
DEPARTMENT OF
REHABILITATION MEDICINE

DISCLOSURES

- 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

UW Medicine
DEPARTMENT OF
REHABILITATION MEDICINE

OBJECTIVE

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

METHODS

STUDY SAMPLE

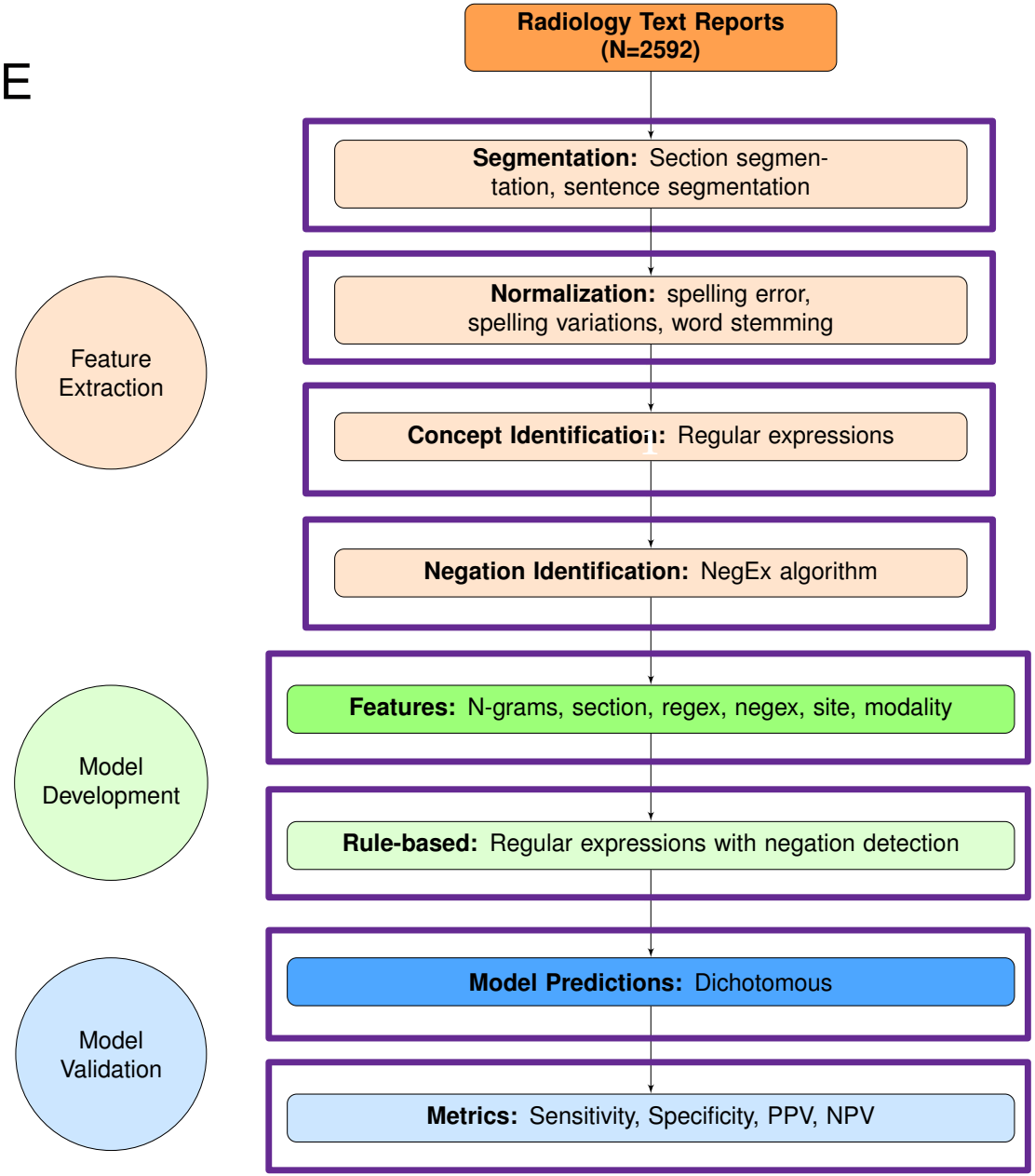
- An “imaging set” was created of lumbar spine imaging reports
- Set was enriched for spondyloarthropathy ICD 9/10 codes

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



UW Medicine
DEPARTMENT OF REHABILITATION MEDICINE

METHODS

STATISICAL 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

	Imaging reports without Spondyloarthropathy	Imaging reports with Spondyloarthropathy
Number of studies	2292	302
Mean age (sd)	66.22 (15.49)	60.49 (17.64)
Image type (%)		
XR	817 (35.6)	251 (83.1)
CT	54 (2.4)	2 (0.7)
MRI	1421 (62.0)	49 (16.2)
Gender (%)		
Female	862 (40.1)	124 (42.8)
Male	1289 (59.9)	166 (57.2)
Unknown	1 (0.0)	0 (0.0)

UW Medicine
DEPARTMENT OF REHABILITATION MEDICINE

RESULTS

TABLE 2: NLP Performance Characteristics

	Percentage (95% confidence interval)
Sensitivity	**95%** *(92-97%)*
Specificity	**98%** *(98-99%)*
Positive predictive value	**91%** *(88-94%)*
Negative predictive value	**99%** *(99-99.9%)*

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

- These performance characteristics reflect the case-enriched sample
  - may not be generalizable to the general population
- Incidentally, “spondyloarthropathy” was noted to be used inappropriately
  - reference to spondylosis rather than true inflammatory arthritis

ACKNOWLEDGEMENTS

This work was funded by the following grants from the National Institutes of Health (NIH):

- UH2 AT007766
- UH3 AR066795

UW Medicine
DEPARTMENT OF REHABILITATION MEDICINE

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
2. Tan, K. et al. Comparison of Natural Language Processing Rules-based and Machine-learning Systems to Identify Lumbar Spine Imaging Findings Related to Low Back Pain. Academic Radiology. 2018; LOE: Level 2