

Image Based Phenotyping for PIRADS in VHA Clinical Notes and Radiology Reports

Cecelia J Madison^{*1}, Hiten D. Patel^{*2,3}, Peter N Danilov¹, Yutai Li^{2,3}, Clayton Neill^{2,3}, Mitchell M. Huang^{2,3}, Joseph Nicolas^{2,3}, Johnathan Stanley¹, Patrick R Alba^{1,4}, Elizabeth Hanchrow¹, Tiffany Quilter¹, Lacy D Castleton¹, Julie A Lynch^{1,4}

*contributed equally, presenting author Hiten D. Patel

¹Salt Lake City VA, VA Informatics and Computing Infrastructure (VINCI), Salt Lake City, UT

²Surgery Service, Jesse Brown VA Medical Center, Chicago, IL

³Department of Urology, Feinberg School of Medicine, Northwestern University, Chicago, IL

⁴University of Utah School of Medicine, Division of Epidemiology, Salt Lake City, UT

Background

The Prostate Imaging Reporting and Data System (PIRADS) standardizes prostate cancer risk reporting from multiparametric MRI. Electronic health records lack structured PIRADS fields with data stored within free text notes, limiting retrospective research and clinical cohort identification for operational initiatives. We developed and validated an automated approach to identify and extract PIRADS scores from Veterans Health Administration (VHA) data.

Methods

We searched VHA procedure ontologies (CPT, HCPCS, ICD 9/10 PCS), identifying 136 codes for prostate imaging across multiple modalities and anatomic regions. We retrieved associated radiology reports and further cross-referenced with independent keyword searches of the radiology domain. A total of 500 notes were manually reviewed and annotated. A natural language processing (NLP) system using the UIMA framework LEO was developed to extract PIRADS scores. We compared clinical characteristics between patients with and without PIRADS scores.

Results

Procedure codes identified 9,843,427 imaging events in 3,721,577 patients (>98% with reports). Keyword search found 108,572 notes (84,827 patients), with 94.2% captured by procedure codes. NLP achieved F1-score >0.9, extracting PIRADS from 94,434 procedures (74,438 patients) across 50 unique codes (5 most common: 72197, 72195, 72196, 76872, 76498). Patients with PIRADS scores had significantly higher urology encounter rates (96.7% vs 64.2%), shorter time to imaging (112 vs 509 days), and greater proportion with elevated PSA by diagnosis code as well as lab test value (99.8% vs 88.3%; all $p < 0.001$).

Conclusion

We demonstrate a validated method to extract PIRADS scores from free-text reports, enabling population-level prostate cancer research and quality improvement in the VHA with potential adaptations to other health systems.

Funding Acknowledgements: VA Data for Veterans Grant (VINCI), Prostate Cancer Foundation (24YOUN22), Department of Defense (HT9425-25-1-0498)

Conflicts of Interest Disclosure Statement: The authors have no relevant disclosures.

Variables	NLP did extract PIRADS n = 91,234	NLP did <i>not</i> extract PIRADS n = 211,136	p-value
Urology clinic encounter, n(%)	88,180 (96.7%)	135,600 (64.2%)	p < 0.001
Time between last Urology clinic visit and imaging procedure (mean, SD)	111.8 ± 426.5 days	508.5 ± 1099.4 days	p < 0.001
BPH diagnosis, n(%)	39,207 (43.0%)	63,339 (30.0%)	p < 0.001
Urethral stricture diagnosis, n(%)	991 (1.1%)	4,463 (2.1%)	p < 0.001
Elevated PSA by ICD, n(%)	51,092 (56.0%)	49,792 (23.6%)	p < 0.001
Elevated PSA by lab test value, n(%)	91,069 (99.8%)	186,454 (88.3%)	p < 0.001
Distribution of PSA values, n(%)	See Figure		p < 0.001

