Multiple centre external validation of an AI solution for prostate cancer diagnostic imaging

Abstract submitted for presentation at RSNA 2023

Author Block: Aarti Shah; Nadia Sofia Moreira Da Silva; Michael Yeung; Francesco Giganti; Lucy Davies; Paul Richard Burn; Richard Hindley; Nikhil Vasdev; John Hayes; Sophie Squire; Alison Jane Bradley; Giles Maskell; Adrian Andreou; Sidath Liyanage; Mark De Bono; Raj Persad; Nimalan Sanmugalingam; Tristan Barrett; Jonathan Aning; Mark Hinton; Antony Rix; Evis Sala

Purpose:

Clinical translation of AI solutions for detection of clinically significant prostate cancer (csPCa) has been limited by the lack of validation on multi-centre datasets including multiple MRI scanners, vendors, field strengths and imaging protocols. Here, we evaluate the ability of an AI solution to generalise to real-world external validation data including blinded validation on an unseen site.

Materials and Methods:

AI-based software was developed using PROSTATEx and retrospective data from five sites (794 patients, 34% csPCa). The software was evaluated on a blinded external validation set (252 patients – 42 per site, 31% csPCa, 9% with prior negative biopsy) of multiparametric (mpMRI) data obtained from six sites; one site was unseen during development, and data from other sites was from later time periods than the development set. This external data included six scanner models from two vendors, with different field strengths (1.5T/3.0T) and acquisition protocols. The software automatically outputs scores intended to identify Gleason score (GS)≥3+4 csPCa per-patient. csPCa was confirmed by biopsy (GS≥3+4 / PI-RADS ≥3), with PI-RADS 1/2 patients that did not receive a biopsy assumed negative. Exclusion criteria included quality issues such as severe motion and metal prostheses, active surveillance, prior prostate or bladder surgery or treatment including brachytherapy, TURP, prostatectomy, ablation, HIFU/focal therapy, or water vapour therapy. Performance was evaluated using ROC analysis, with 95% confidence intervals estimated by bootstrapping.

Results:

For selecting patients for biopsy, the AI identified patients with csPCa with sensitivity 94% (95% CI 88-99%), specificity 57% (49-64%), NPV 95% (90-99%), and AUC 0.85 (0.80-0.90) using mpMRI data from the blinded external validation set. Comparing between sites, the AUC ranged from 0.70-0.98, with a pooled AUC of 0.86±0.11. On the unseen site, the AUC was 0.95 (0.87-1.00).

Reporting radiologists had per-patient sensitivity 99% (95% CI 96-100%) due to the assumed ground truth, specificity 73% (67-80%), NPV 99% (98-100%), and AUC 0.95 (0.92-0.97). In a 2019 Cochrane meta-analysis of 12 major studies (37% csPCa), radiologists identified patients with GS≥3+4 csPCa with sensitivity 86% and specificity 42%.

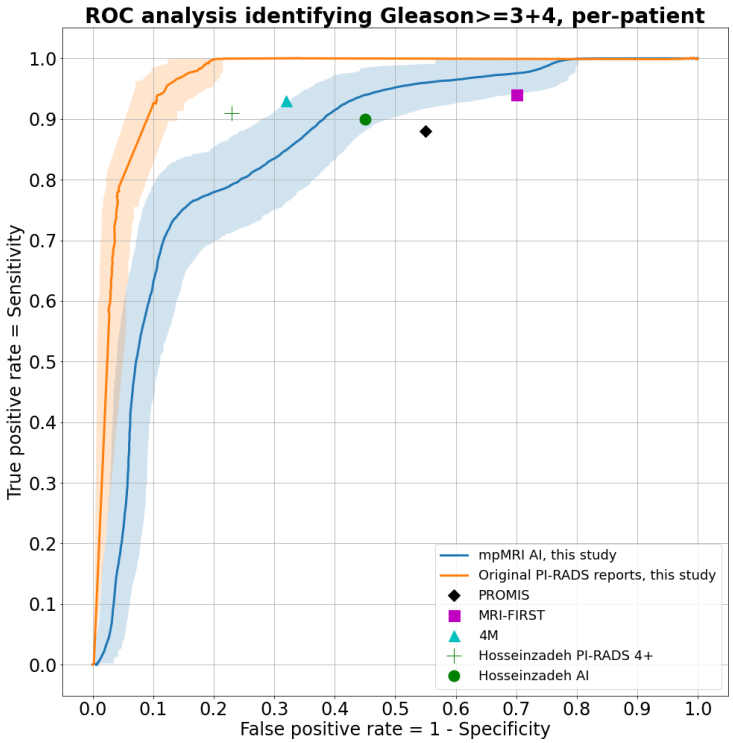
Conclusion:

The proposed AI solution shows comparable performance to radiologists in major expert studies, on a large real-world, multi-centre, external validation dataset with different scanners, vendors, field strengths and imaging protocols.

Clinical Relevance:

AI could support prostate cancer detection in clinical practice, generalises to multiple sites, scanners and imaging protocols, and is robust to novel data.

Per-patient sensitivity vs. false positive rate, with comparators



Per-patient results, AI (mpMRI external validation data, 252 patients, 34% csPCA)

Threshold	Sensitivity	Specificity	NPV	AUC
4.0	0.82 (0.74-0.90)	0.71 (0.65-0.78)	0.90 (0.84-0.94)	0.85 (0.80-0.90)
3.5	0.94 (0.88-0.99)	0.57 (0.49-0.64)	0.95 (0.90-0.99)	
3.0	0.98 (0.93-1.00)	0.34 (0.27-0.41)	0.97 (0.91-1.00)	
2.5	1.00 (1.00-1.00)	0.21 (0.16-0.28)	1.00 (1.00-1.00)	

Comparator studies

Study	Sensitivity	Specificity	NPV	AUC
PROMIS (1)	0.88 (0.84-0.91)	0.45 (0.39-0.51)	0.76 (0.69-0.82)	0.67
MRI-FIRST (2)	0.94	0.30	0.89	0.81
4M (3)	0.93	0.68	0.96	0.85
This study – Radiologists* (4)	0.99	0.73 (0.67-0.80)	0.99	0.95
Hosseinzadeh PI-RADS 4+ (5)	0.91	0.77	0.95	0.89
Hosseinzadeh DL-CAD AI (5)	0.90	0.55	0.93	0.85

* MRI-negative cases did not receive biopsy which may positively bias performance.

95% confidence intervals estimated through bootstrapping are shown where available.

Comparator studies with extensive biopsy ground truth (all cited at PI-RADS/Likert 3 threshold unless stated):

1. PROMIS: Ahmed HU, et al. Lancet 2017;389:815-822
2. MRI-FIRST: Rouvière O, et al. Lancet Oncol. 2019;20(1):100-109
3. 4M: van der Leest M, et al. Eur. Urol. 2019;75(4):570–578
4. 252 patients from blinded external data; note that most MRI-negative cases did not receive biopsy, resulting in the assumption that sensitivity and NPV are near 100%
5. Hosseinzadeh M, et al. Eur Radiol. 2022; 32(4): 2224–2234 (AI study)