Integrating clinical data with AI to optimise decision-making in prostate MRI

Abstract submitted for presentation at RSNA 2023

Author Block: Antony Rix; Nadia Moreira Da Silva; Jobie Budd; Michael Yeung; Francesco Giganti; Lucy Davies; Paul Richard Burn; Richard Hindley; Nikhil Vasdev; Alison Jane Bradley; Giles Maskell; Adrian Andreou; Sidath Liyanage; Raj Persad; Jonathan Aning; Tristan Barrett; Mark Hinton; Anwar Roshanali Padhani; Evis Sala; Aarti Shah

Purpose:

To determine whether combining prostate MRI AI-based decision support outputs, clinical data and PI-RADS scores in a multi-modal predictive model enhances detection of clinically significant prostate cancer.

Methods and Materials:

MRI, clinical history, histopathology, and PI-RADS scores were obtained retrospectively from five sites in a multi-vendor, multiple field strength study. After exclusions for AI contraindications including prior treatment and quality issues, model training used data from 352 patients and a held-out test set comprised data from 235 patients (Gleason grade group (GGG)≥2, prevalence 34%).

Our automated multi-stage AI-based software segments and calculates the volume of prostate whole gland and transition zone (TZ) on MRI, and segments and scores lesions/patients for GGG≥2 disease likelihood.

Biopsy-verified GGG≥2 was used as ground truth, with MRI-negative patients not undergoing biopsy assumed negative. Sensitivity, specificity, and AUC were evaluated at patient level on the held-out test set, with 95% confidence intervals obtained through bootstrapping. Combinations of AI, clinical and PI-RADS data were tested for significant improvement to the AI score and PI-RADS assessment, at pre-determined thresholds equivalent to PI-RADS 3.

Results:

mpMRI PI-RADS scores alone detected GGG≥2 with sensitivity 1.00 (95% CI 1.00-1.00), specificity 0.67 (0.61-0.75) and AUC 0.94 (0.91-0.97).

GGG≥2 was detected by bpMRI AI with sensitivity 0.97 (0.93-1.00), specificity 0.55 (0.47-0.62) and AUC 0.88 (0.84-0.92). Combining AI score and TZ-PSA density (PSAD) improved specificity (sensitivity 0.95 (0.90-0.99), specificity 0.70 (0.63-0.77) and AUC 0.90 (0.85-0.93)).

The addition of AI and TZ-PSAD to PI-RADS scores maintained high sensitivity of 0.99 (0.96-1.00), while significantly improving specificity to 0.83 (0.77-0.89, KS p-value<0.001) and AUC to 0.96 (0.93-0.98, DeLong p-value 0.003).

TZ volume based PSAD had modest additional benefit compared to whole-prostate PSAD. Other variables offered <5% specificity improvements or non-significant benefits. Findings with bpMRI and mpMRI AI models were similar.

Limitations:

Most MRI-negative cases did not receive biopsy in this retrospective study.

Conclusion:

The use of PSAD improves the predictive accuracy of prostate MRI AI decision support, with significant improvement in specificity at similar sensitivity. Combining PI-RADS, PSAD and AI offers substantial improvement compared to AI or PI-RADS assessments alone.

Clinical Relevance:

The improved specificity achieved through integrating patient PSAD and radiologists' PI-RADS scores with AI software can potentially reduce false positive cases, further aiding patient selection for biopsy using MRI.

Per-patient sensitivity vs. false positive rate, identifying patients with any Gleason>=3+4 cancer

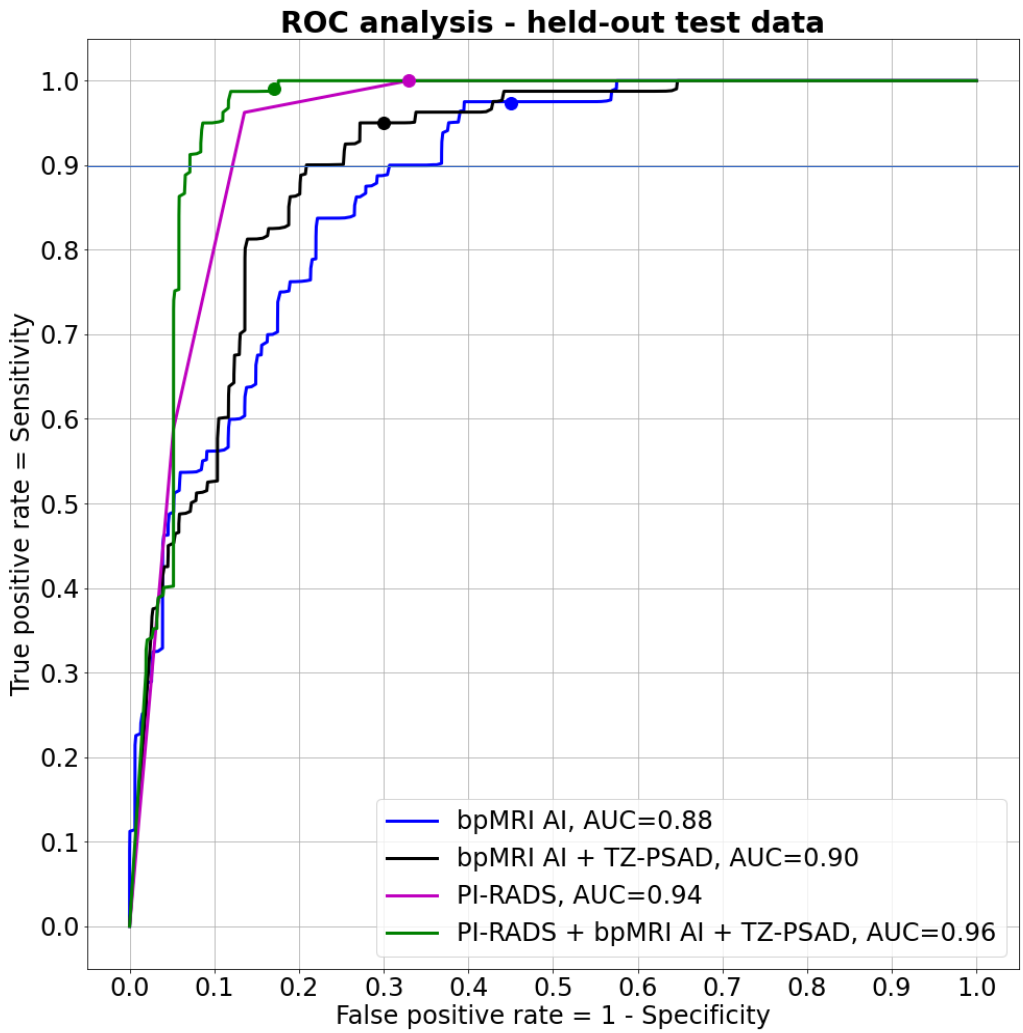

Figure illustrates the patient-level ROC curves obtained on the held-out test data (n=235) for each predictor compared against ground truth:

- bpMRI AI alone
- Combined model using bpMRI AI score together with TZ-PSAD computed using PSA data and AI TZ volume
- PI-RADS category
- Combined model using PI-RADS, bpMRI AI score and TZ-PSAD.

The combined models were trained on a separate training data set (n=352). The sensitivity and specificity measures reported in the text and plotted in the figure were obtained at thresholds that were pre-determined from the training data to correspond to PI-RADS 3.

Ground truth for this evaluation is biopsy-verified Gleason grade group ≥2 or higher cancer. Patients in this study received standard-of-care biopsy according to local practice. MRI-negative (PI-RADS 1-2) patients who did not receive a biopsy were assumed negative. Note that this implies near-100% sensitivity for the clinical PI-RADS assessment. It is therefore appropriate to consider the improvements to specificity (false-positive reductions) when assessing the potential added value of data integration in this analysis.