



OPEN

Machine learning prediction of Gleason grade group upgrade between in-bore biopsy and radical prostatectomy pathology

Kaan Ozbozduman^{1✉}, Irem Loc¹, Selahattin Durmaz², Duygu Atasoy³, Mert Kilic⁴, Hakan Yildirim⁵, Tarik Esen^{4,6}, Metin Vural⁵ & M. Burcin Unlu^{7,8}

This study aimed to enhance the accuracy of Gleason grade group (GG) upgrade prediction in prostate cancer (PCa) patients who underwent MRI-guided in-bore biopsy (MRGB) and radical prostatectomy (RP) through a combined analysis of prebiopsy and MRGB clinical data. A retrospective analysis of 95 patients with prostate cancer diagnosed by MRGB was conducted where all patients had undergone RP. Among the patients, 64.2% had consistent GG results between in-bore biopsies and RP, whereas 28.4% had upgraded and 7.4% had downgraded results. GG1 biopsy results, lower biopsy core count, and fewer positive cores were correlated with upgrades in the entire patient group. In patients with GG > 1, larger tumor sizes and fewer biopsy cores were associated with upgrades. By integrating MRGB data with prebiopsy clinical data, machine learning (ML) models achieved 85.6% accuracy in predicting upgrades, surpassing the 64.2% baseline from MRGB alone. ML analysis also highlighted the value of the minimum apparent diffusion coefficient (ADC_{min}) for GG > 1 patients. Incorporation of MRGB results with tumor size, ADC_{min} value, number of biopsy cores, positive core count, and Gleason grade can be useful to predict GG upgrade at final pathology and guide patient selection for active surveillance.

Prostate cancer is the second most frequent cancer diagnosed in men, and despite many improvements in detection and treatment, it is still one of the leading causes of cancer-related mortality¹. Histology from needle biopsy is crucial for risk stratification and proper selection of treatment options tailored to the characteristics of each tumor and patient. Historically, samples for histopathology have been obtained by standard transrectal ultrasound (TRUS)-guided biopsy using a systematic scheme. However, TRUS-guided prostate biopsy has long been known to result in undersampling, and the other diagnostic uncertainty of this technique is discordance between needle biopsy and RP histologic grading^{2,3}, which might lead to under or overtreatment.

In recent years, mp-MRI and subsequent MRI-targeted biopsy techniques have proven to be a highly accurate pathway for detecting clinically significant PCa while simultaneously decreasing the detection of clinically insignificant cancers⁴. The three common MRI-targeted biopsy techniques include visual registration of MRI images with real-time ultrasound images, software-assisted fusion of MRI images and real-time ultrasound images, and MRI-guided in-bore biopsy. MRI-guided in-bore biopsy technique by using ADC maps for fine manual adjustments during the procedure and real-time feedback for needle placement can help to obtain an adequate fraction of the tumor to reach higher concordance between biopsy and final pathology^{5,6}. Although there is increased concordance between MRI-targeted biopsy and final pathology from RP, upgrades of the Gleason grade group (GG) are still important, especially for patients with GG1, who are candidates for active surveillance. Active surveillance is a management strategy for low-risk PCa patients designed to avoid overtreatment and the potential side effects of surgery⁷. High Prostate Imaging Reporting and Data System (PI-RADS) scores and/or large tumor sizes on mp-MRI were reported to be predictive factors of upgrading GG1 lesions⁸.

¹Bogazici University Physics Department, Istanbul, Turkey. ²Department of Radiology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey. ³Department of Radiology, University of Koc School of Medicine, Istanbul, Turkey. ⁴Department of Urology, VKF American Hospital, Istanbul, Turkey. ⁵Department of Radiology, VKF American Hospital, Istanbul, Turkey. ⁶Department of Urology, University of Koc School of Medicine, Istanbul, Turkey. ⁷Faculty of Engineering, Ozyegin University, Istanbul, Turkey. ⁸Faculty of Aviation and Aeronautical Sciences Ozyegin University, Istanbul, Turkey. ✉email: kaan.ozbozduman@boun.edu.tr

In this retrospective study, an analysis of the clinical variables of patients who underwent MRI-targeted in-bore biopsy and subsequent RP was conducted, with two primary objectives: (i) to identify the clinical variables that were pertinent to the upgrade of GG in the final pathology and (ii) to investigate the possibility of predicting GG upgrade through the utilization of machine learning methods on an individual patient basis, thereby providing a foundation for personalized treatment planning.

Methods

Patients

Following the internal review board approval of Koc University, a retrospective examination was conducted on the datasets of 400 men who underwent mp-MRI and subsequent MRI-targeted in-bore biopsy at American Hospital (Istanbul, Turkiye) between 2012 and 2022, with the high likelihood target, i.e. PI-RADS 4 and 5. The research was carried out adhering to the principles outlined in the Declaration of Helsinki. Among the patients who were diagnosed with PCa, 95 of them (median age 64, range 42–78) underwent RP as a definitive treatment. Of these, 20 patients were diagnosed with GG1 PCa based on the in-bore biopsy results. Although GG1 patients normally do not require active treatment and are actively surveilled, the shared decision for definitive treatment took into account the following criteria: History of prostate cancer in the father or brother, International Prostate Symptom Score (IPSS) > 19, tumor positivity of 2 cores or more, and PI-RADS 4 or PI-RADS 5 lesions bigger than 10 mm. The time interval between in-bore biopsy and RP was less than 6 months for most of the patients. None of the patients had received either radiotherapy or hormone therapy before RP. All biopsy cores and radical prostatectomy specimens were evaluated by a dedicated uropathologist with 16 years of experience, according to the recommendation of the International Society of Urological Pathology (ISUP).

Our study focused on per index lesion level. All high-likelihood lesions (PI-RADS 4 and 5) detected on mp-MRI were targeted by in-bore biopsy. Index lesions were depicted according to PI-RADS version 2 guideline⁹. All index lesions detected on mp-MRI and sampled by in-bore biopsy were confirmed at whole-mount step-section specimens after RP. Since non-index lesions were not clinically related to patient outcomes, they were not analyzed^{10–13}.

Multiparametric MRI and measurements

All multiparametric-MRI examinations were conducted on a 3.0 Tesla MRI Scanner (Magnetom Skyra, Siemens AG, Germany) with sixteen-channel body coil. Butylscopolamine were used to suppress bowel peristalsis during the examination. The MRI protocol included T2-weighted imaging in axial, coronal and sagittal planes, diffusion-weighted imaging (DWI), and dynamic contrast-enhanced pulse sequences (Table 1).

Tumor size, tumor location and PI-RADS scores were interpreted in consensus by three radiologists. ADC values were measured by two radiologists who were blinded to clinical variables and pathology results. The ADC_{mean} values were obtained by drawing a regions of interest (ROIs) that cover the largest tumor area excluding the tumor edges. While the ADC_{min} values were obtained by drawing an ROI on the area that visually depicts the lowest ADC value within the tumor (Fig. 1). The interobserver agreement regarding the ADC measurements was assured with 79.3% and 67.7% correlations for mean and min values, respectively. The pathologic interpretation was the same as our previous publication¹⁴.

In-bore biopsy technique

In-bore biopsy was performed in an outpatient setting on the same 3 T MRI scanner. All biopsy procedures were carried out by a single radiologist [M.V.] who had more than 15 years of experience in urogenital radiology and interventions.

During the biopsy, the patients were positioned in the prone position. The needle guide, lubricated with 2% lidocaine gel, was inserted into the rectum and attached to a commercially available biopsy device (DynaTRIM, Invivo). To adjust the needle guide placement, sagittal T2W turbo spin-echo images were acquired and transferred to a workstation (DynaCAD, Invivo) in the first place. Subsequently, the software then calculated the

| Parameter | Axial T2-WI TSE | Sagittal T2-WI TSE | Coronal T2-WI TSE | DWI | DCE | T1-WI |
|------------------------------|-----------------|--------------------|-------------------|-----------|-----------|-----------|
| TR (ms) | 5290 | 5670 | 3800 | 4800 | 9 | 445 |
| TE (ms) | 111 | 113 | 105 | 98 | 1.76 | 9.80 |
| Field of view (mm) | 200 × 200 | 200 × 200 | 200 × 200 | 260 × 260 | 260 × 260 | 300 × 400 |
| Matrix size | 512 × 297 | 320 × 224 | 384 × 230 | 192 × 154 | 192 × 155 | 384 × 297 |
| Slice thickness (mm) | 3/0.6 | 2.5/0 | 2.5/0 | 3.6/0 | 3.6/0 | 6/1.2 |
| Flip angle (°) | 180 | 180 | 180 | 180 | 15 | 120 |
| Scan time (mins) | 03:23 | 02:46 | 02:11 | 05:38 | 04:38 | 1:16 |
| Temporal res. (s) | – | – | – | – | 8 | – |
| B value (s/mm ²) | – | – | – | 1600 | – | – |

Table 1. Multiparametric MRI protocol.

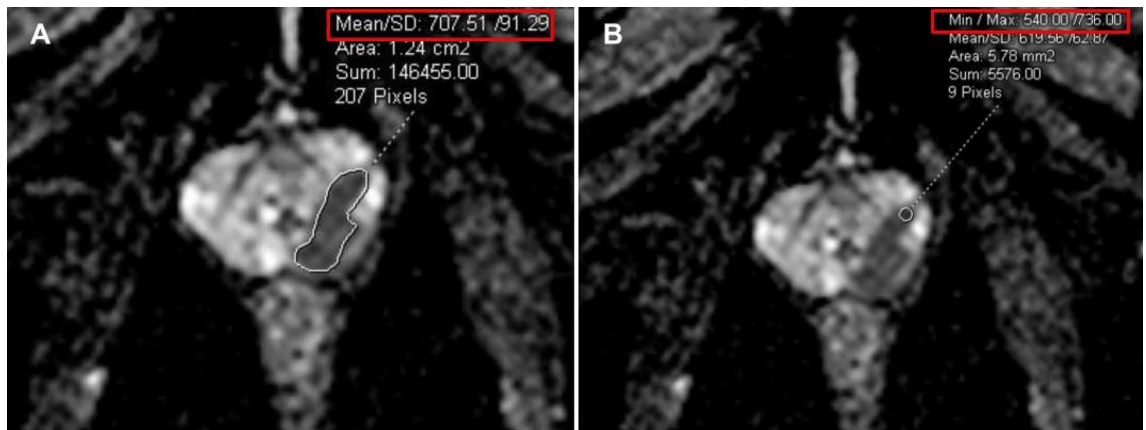


Figure 1. ADC images demonstrate a PI-RADS 5 lesion in the left peripheral zone, which was subsequently confirmed as prostate cancer (Gleason grade 3). The mean ADC and minimum ADC values were measured as shown in Figure (A) and (B), respectively.

target's rough coordinates relative to the needle guide's tip, which was manually adjusted toward the target. ADC maps were also utilized during the manual needle adjustments to guide the needle to the area with the lowest ADC values (Fig. 1).

Following the initial adjustments, repeat sagittal and multiplanar reconstructed axial and coronal T2-weighted images were obtained for further fine manual adjustments until the needle guide was accurately pointed to the designated target (Fig. 2). Biopsy cores were obtained using an MRI-compatible, 18-gauge biopsy gun with needle lengths of 150 or 175 mm (In vivo, Gainesville, FL). To ensure accurate sampling of the targeted lesion, the fired needle was left deployed in the prostate, and sagittal and reconstructed T2-weighted images were acquired. Only the suspicious target detected on pre-biopsy mpMRI was sampled without performing a complementary systematic biopsy. During the course of our study, we increased the number of biopsy cores in relation to the growing evidence that focal saturation can improve the compatibility of needle biopsy with whole-mount specimen pathology. On a case-by-case basis, the number of biopsy cores was also affected by the patient's comorbidities, the history of previous negative biopsy, the size and location of the target, and feedback from needle-in images. The number of biopsy cores that were obtained per each lesion ranged from 2 to 5.

Clinical parameters

Pre-biopsy clinical variables include patient age, prostate volume, prostate specific antigen (PSA), PSA density (PSAD), tumor size, tumor location (either in peripheral zone (PZ) or transition zone (TZ)), assigned PI-RADS score, mean and minimum ADC values acquired by diffusion weighted images. Biopsy records include number of biopsy cores, number of positive biopsy cores, the ratio of positive cores to total number of cores, total biopsy core length (CL), total biopsy tumor length (TL), TL/CL ratio, and biopsy-assigned GG. Table 2 shows the characteristics of the patients that are involved in this study.

Statistical and machine learning analysis

In order to identify clinical parameters that are predictive for GG upgrade, univariate statistics and multivariate machine learning (ML) analyses were performed. For the univariate statistical tests logistic regression was employed. Odds ratio (OR) with confidence interval (CI) that excludes 1 and $p < 0.05$ are considered significant.

The baseline prediction accuracy was calculated by comparison of in-bore biopsy and radical prostatectomy Gleason grades, which was used as the benchmark to evaluate the performance of ML models. ML studies were conducted by selecting algorithms that are robust to overfitting for relatively small datasets such as ours, namely, support vector machine (SVM) with linear and radial basis function (RBF) kernels, least absolute shrinkage and selection operator (LASSO) regression, and ridge regression. To assess the performances of the ML algorithms, we used sensitivity, specificity, the area under the receiver operator characteristic (ROC) curve (AUC)¹⁵, and the Youden index¹⁶ metrics. Our analyses employed 3 different grouping strategies for the patient cohort: (i) we included all patients and studied all patients with a GG upgrade, (ii) we included GG > 1 patients and studied all patients with a GG upgrade, and (iii) we included all patients and studied only those with clinically significant upgraded cases, from GG1 to GG > 1.

The evaluation of performance metrics was conducted through a rigorous process involving 100 randomly selected train-test splits across the dataset, ensuring a comprehensive examination of the model's robustness and consistency. We adhered to a train-test split ratio of 80% for training data and 20% for testing data. Furthermore, to assess the model's generalizability and mitigate the risk of overfitting, we employed a 3-fold cross-validation strategy.

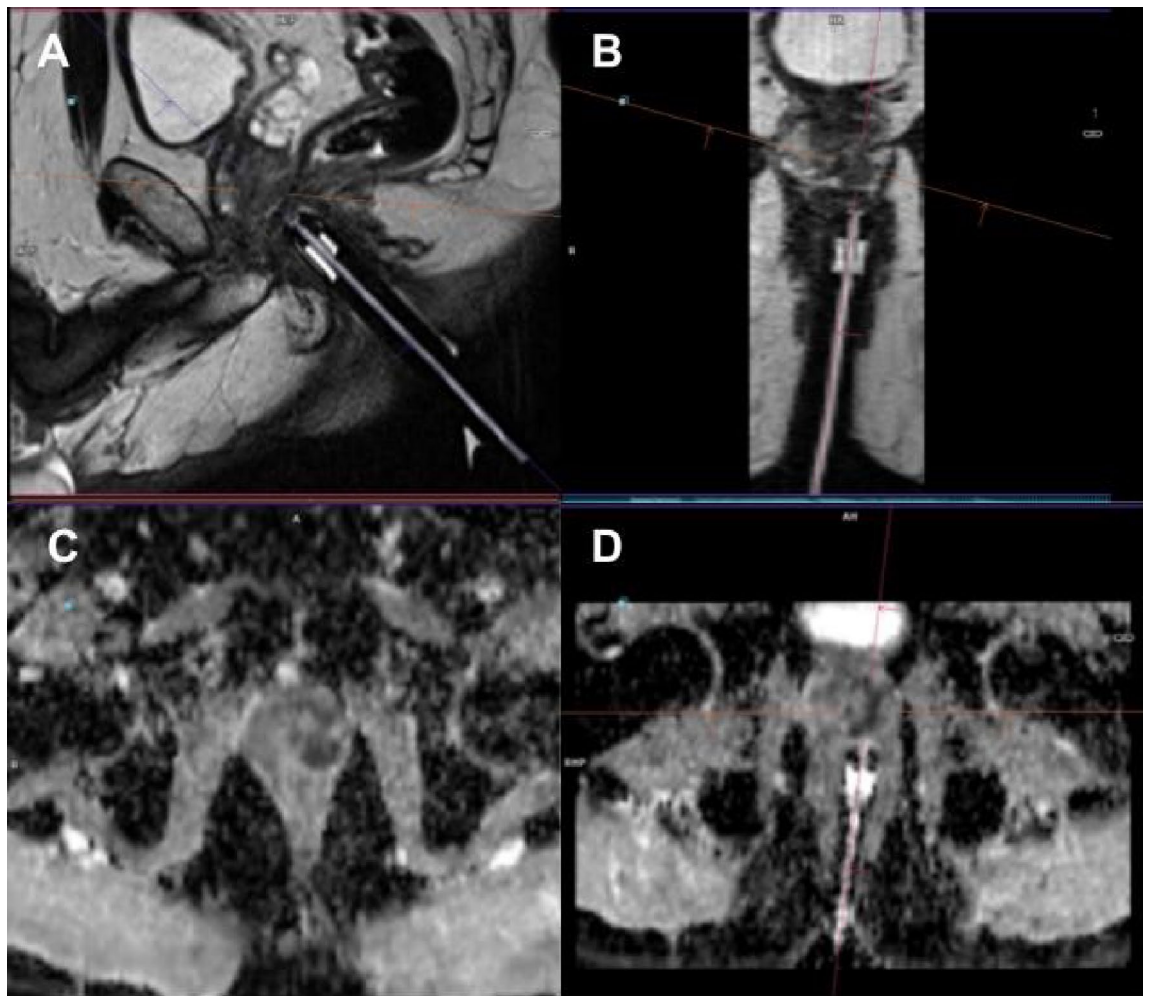


Figure 2. A 57-year-old patient presented with an elevated level of prostate-specific antigen (PSA) measuring 8.9 ng/ml, accompanied by suspicious findings on digital rectal examination. Multiparametric magnetic resonance imaging (mp-MRI) identified a PI-RADS 5 lesion in the left peripheral zone. Subsequently, an MRI-guided in-bore biopsy was performed and the diagnosis of prostate cancer (Gleason Group 4) was established. Sagittal (A) and axial (B) T2-weighted images, axial ADC image (D) showing biopsy needle positioning. Axial ADC image (C) taken during in-bore biopsy.

Informed consent

This retrospective observational study was approved by our Institutional Review Board and the requirement for informed written consent was waived by the Koc University School of Medicine ethics committee. All experiments including the study protocol study followed approved institutional guidelines.

Results

In our study cohort, concordance between biopsy and final pathology GG was recorded in 61 (64.2%) patients. Overall upgrading was recorded in 27 (28.4%) patients, whereas 7 (7.4%) patients were downgraded. Six downgrading men were lowered to the preceding GG, whereas a single case was downgraded by 2 Gleason grade groups (from GG4 to GG2). Among 27 upgraded men, 21 (77.8%) patients' Gleason grade group were increased by 1 grade. Upgrades by 2 (n=3) and 3 (n=3) grades were also observed equally in 6 cases in total. Among 75 men with biopsy GG > 1, 10 (13.3%) upgraded cases were observed whereas 58 (77.3%) cases were concordant. Table 3 shows GG distribution obtained by in-bore biopsy versus RP where diagonal elements represent concordance. The upper and lower diagonal elements represent the cases with GG upgrades and downgrades, respectively. All of the upgraded cases from clinically insignificant to clinically significant PCa (17.9% in our study cohort) consisted of upgrades from GS 3+3 to 3+4, whereas downgrading to clinically insignificant PCa did not occur. We focused on the statistics of GG upgrade only due to the lack of downgraded cases. Table 2 gives a comparative account of clinical variable characteristics in men whose GG upgraded after RP in comparison to the men whose GG did not upgrade.

| Continuous features (unit) | All (N=95) | | Upgrade (N=27) | No upgrade (N=68) |
|---------------------------------|-------------------------------|--|------------------------------|-----------------------------|
| | Mean (Std) | | | |
| | Median [range (IQR)] | | | |
| Age (years) | 63.0 (6.6) | | 61.4 (7.9) | 63.6 (5.9) |
| | 64 [42-78 (58-67)] | | 61 [42-72 (56-67)] | 64 [53-78 (59-67)] |
| PSA (ng/ml) | 6.5 (4.3) | | 5.4 (3.1) | 7.0 (4.7) |
| | 5.4 [1.5-26.0 (4.0-7.0)] | | 5.0 [1.5-16.0 (3.9-6.0)] | 5.7 [2.1-26.0 (4.1-7.2)] |
| Prostate volume (ml) | 46.8 (22.6) | | 46.0 (20.9) | 47 (23.5) |
| | 42.0 [16.0-151.0 (30.0-58.0)] | | 48 [19-110 (28.2-56.9)] | 42 [16-151 (30.7-58)] |
| PSAD (ng/ml/ml) | 0.15 (0.11) | | 0.14 (0.07) | 0.15 (0.12) |
| | 0.12 [0.00-0.71 (0.09-0.17)] | | 0.14 [0.04-0.27 (0.09-0.18)] | 0.12 [0.0-0.71 (0.09-0.16)] |
| ADC _{mean} | 774 (184) | | 769 (190) | 776 (182) |
| | 754 [322-1203 (646-882)] | | 759 [322-1129 (691-877)] | 742 [401-1203 (631-882)] |
| ADC _{min} | 634 (163) | | 639 (199) | 632 (148) |
| | 639 [185-955 (531-772)] | | 664 [185-934 (550-782)] | 628 [350-955 (524-712)] |
| Tumor size (mm) | 12.3 (6.7) | | 14.7 (8.8) | 11.4 (5.4) |
| | 11.0 [4.0-40.0 (8.0-15.0)] | | 12.0 [5-40 (9-19.5)] | 11 [4-33 (7-13)] |
| Total biopsy core length (mm) | 41.1 (15.0) | | 38.8 (15.8) | 42.0 (14.7) |
| | 40 [10-79 (31-52)] | | 36 [12-73 (27-48.5)] | 40 [10-79 (32.5-50.5)] |
| Total biopsy tumor length (mm) | 20.6 (12.1) | | 17.7 (10.9) | 21.8 (12.4) |
| | 18 [0.3-66 (12.2-26)] | | 16 [0.3-40 (11.5-23)] | 19 [3-66 (13-27.5)] |
| TL/CL | 0.50 (0.23) | | 0.46 (0.26) | 0.52 (0.21) |
| | 0.5 [0.03-1.0 (0.34-0.64)] | | 0.44 [0.03-0.95 (0.34-0.59)] | 0.50 [0.07-1.0 (0.35-0.66)] |
| Positive biopsy core ratio | 0.91 (0.19) | | 0.90 (0.23) | 0.91 (0.18) |
| | 1.0 [0.25-1.0 (1.0-1.0)] | | 1.0 [0.25-1.0 (1.0-1.0)] | 1.0 [0.25-1.0 (0.94-1.0)] |
| Categorical features | Value(%) | | | |
| PI-RADS | 4 (62.1%), 5 (37.9%) | | 4 (48.1%), 5 (51.9%) | 4 (67.6%), 5 (32.4%) |
| Prostate zone | PZ (85.3%), TZ (14.7%) | | PZ (85.2%), TZ (14.8%) | PZ (85.3%), TZ (14.7%) |
| Number of biopsy cores | 2 (11.6%), 3 (41.1%) | | 2 (25.9%), 3 (51.9%) | 2 (5.8%), 3 (36.8%) |
| | 4 (42.1%), 5 (5.2%) | | 4 (22.2%) | 4 (50.0%), 5 (7.4%) |
| Number of positive biopsy cores | 1 (6.3%), 2 (15.8%) | | 1 (11.2%), 2 (29.6%) | 1 (4.4%), 2 (10.3%) |
| | 3 (47.4%), 4 (25.3%) | | 3 (44.4%), 4 (14.8%) | 3 (48.5%), 4 (29.4%) |
| | 5 (5.2%) | | | 5 (7.4%) |
| Biopsy Gleason grade | 1 (21.1%), 2 (42.1%) | | 1 (63.0%), 2 (25.9%) | 1 (4.4%), 2 (48.5%) |
| | 3 (20.0%), 4 (12.6%) | | 3 (7.4%), 4 (3.7%) | 3 (25.0%), 4 (16.2%) |
| | 5 (4.2%) | | | 5 (5.9%) |

Table 2. Patient characteristics comparison for GG upgraded vs non-upgraded cohorts.

| In-bore biopsy GG | RP Specimen GG | | | | | Total |
|-------------------|----------------|----|----|---|----|-------|
| | 1 | 2 | 3 | 4 | 5 | |
| 1 | 3 | 17 | 0 | 0 | 0 | 20 |
| 2 | 0 | 33 | 3 | 1 | 3 | 40 |
| 3 | 0 | 3 | 14 | 0 | 2 | 19 |
| 4 | 0 | 1 | 3 | 7 | 1 | 12 |
| 5 | 0 | 0 | 0 | 0 | 4 | 4 |
| Total | 3 | 54 | 20 | 8 | 10 | 95 |

Table 3. Confusion table of Gleason grades by MRI-guided in-bore biopsy versus RP pathology. Diagonal elements indicate GG concordance. Upper- and lower-diagonal elements indicate GG upgrade and downgrade cases, respectively.

Statistical analysis

Univariate analyses were conducted using logistic regression, and the results are shown in Table 4. Biopsy GG1 stands out as the most significant predictive factor for a GG upgrade at RP (95% CI 0.06–0.32, $p < 0.0001$), such that 17 of the 20 patients with GG1 were upgraded. A smaller number of biopsy cores (95% CI 0.3–0.76, $p = 0.002$) and fewer positive biopsy cores (95% CI 0.35–0.87, $p = 0.011$) were found to be independent predictive clinical factors by univariate analysis.

The fact that the majority of upgraded patients (17 out of 27) in our study had biopsy GG1 poses the risk that this bulk would saturate our statistics and prevent us from identifying other important upgrade risk factors. Therefore, we repeated the statistical analysis for biopsy GG > 1 patients only, where increasing tumor size (95% CI 1.07–3.54, $p = 0.028$) and decreasing number of biopsy cores (95% CI 0.36–0.97, $p = 0.038$) were the statistically significant predictive factors. Furthermore, we studied the clinically significant upgraded cases (from GG1 to GG > 1), yet none of the clinical parameters turned out to be significant indicators.

The most significant cutoff thresholds for the statistically significant parameters were found by binarizing the parameters using various thresholds and minimizing the p -value. The results indicate that the number of biopsy cores and positive biopsy cores should at least be equal to or larger than 3 and 2, respectively, to decrease the likelihood of GG upgrade. In addition, for the GG > 1 subgroup, a tumor size equal to 20 mm stands out as the best diagnostic criterion.

Machine learning

The baseline prediction accuracy set by the in-bore biopsy GG was 64.2% for all patients and 77.3% for the patients with biopsy GG > 1. We aimed to improve this model by introducing clinical variables. To select the optimum clinical features that maximize the performance of the ML models, we first scaled all clinical variables to the [0, 1] range and then ordered all clinical variables according to their chi-square statistics to GG upgrade. First, machine learning models were trained using only the most correlated feature. Then, at each step, we added the next feature in order and observed its effect on the model performance, measured by the Youden index. At a certain point, the ML models reached a maximum Youden index, and we kept the feature set at that point as our predictive variables. Figure 3a shows the performance of SVM with linear and RBF kernels and LASSO and ridge regressions as a function of the clinical feature set, including the overall patient cohort, after 100 random train-test split iterations. The most favorable results were obtained using an SVM with an RBF kernel (Youden index: 0.575 ± 0.013 , accuracy: 0.856 ± 0.004 , sensitivity: 0.621 ± 0.013 , specificity: 0.953 ± 0.003 , and AUC: 0.865 ± 0.007) with two predictive clinical features: total number of cores and in-bore biopsy GG.

The same procedure was repeated for biopsy GG > 1 patients (see Fig. 3b), where ridge regression yielded optimum results (Youden index: 0.590 ± 0.024 , accuracy: 0.904 ± 0.005 , sensitivity: 0.652 ± 0.023 , specificity: 0.938 ± 0.004 , and AUC: 0.944 ± 0.005) with 10 predictive clinical features, namely, the total number of cores, PI-RADS score, tumor size, ADC_{min}, PSAD, in-bore biopsy GG, number of positive cores, prostate volume, core length, and PSA. The steepest improvement in ML model performances was caused by ADC_{min} to feature set for

| | All patients | | Biopsy GG > 1 patients | | Clinically significant GG upgrade cases | |
|---------------------------------|--------------------------|-------------------|--------------------------|---------------|---|--------|
| Total population | N = 95 | | N = 75 | | N = 95 | |
| Upgraded cases | N = 27 | | N = 10 | | N = 17 | |
| Clinical features | OR (CI) | P val. | OR (CI) | P val. | OR (CI) | P val. |
| Age | 0.75 (0.50–1.14) | 0.181 | 1.04 (0.66–1.64) | 0.855 | 0.69 (0.45–1.05) | 0.085 |
| PSA | 0.75 (0.48–1.17) | 0.210 | 0.90 (0.57–1.42) | 0.653 | 0.80 (0.52–1.23) | 0.306 |
| Prostate volume | 0.97 (0.65–1.45) | 0.866 | 0.95 (0.61–1.50) | 0.841 | 1.00 (0.67–1.49) | 0.994 |
| PSAD | 0.92 (0.61–1.38) | 0.677 | 0.99 (0.63–1.56) | 0.974 | 0.91 (0.60–1.36) | 0.633 |
| ADC _{mean} | 0.98 (0.66–1.46) | 0.920 | 0.80 (0.51–1.28) | 0.354 | 1.22 (0.81–1.83) | 0.344 |
| ADC _{min} | 1.03 (0.69–1.54) | 0.879 | 0.79 (0.50–1.26) | 0.322 | 1.29 (0.86–1.95) | 0.220 |
| Tumor size | 1.50 (0.95–2.37) | 0.079 | 1.95 (1.07–3.54)* | 0.028* | 0.93 (0.62–1.40) | 0.734 |
| Total biopsy core length | 0.84 (0.56–1.26) | 0.407 | 0.80 (0.51–1.27) | 0.351 | 0.98 (0.65–1.46) | 0.907 |
| Total biopsy tumor length | 0.76 (0.50–1.15) | 0.191 | 1.00 (0.63–1.57) | 0.989 | 0.71 (0.46–1.09) | 0.116 |
| TL/CL | 0.82 (0.54–1.23) | 0.332 | 1.11 (0.71–1.75) | 0.651 | 1.02 (0.47–1.09) | 0.115 |
| PI-RADS | 1.39 (0.92–2.10) | 0.112 | 1.44 (0.91–2.29) | 0.121 | 1.05 (0.70–1.57) | 0.813 |
| Tumor zone | 1.00 (0.67–1.49) | 0.990 | 1.05 (0.67–1.66) | 0.820 | 0.94 (0.63–1.41) | 0.774 |
| Number of biopsy cores | 0.48 (0.30–0.76)* | 0.002* | 0.59 (0.36–0.97)* | 0.038* | 0.71 (0.47–1.08) | 0.111 |
| Number of positive biopsy cores | 0.56 (0.35–0.87)* | 0.011* | 0.69 (0.42–1.11) | 0.121 | 0.71 (0.47–1.09) | 0.115 |
| Positive biopsy core ratio | 0.96 (0.64–1.44) | 0.858 | 1.03 (0.66–1.63) | 0.881 | 0.91 (0.61–1.37) | 0.649 |
| Biopsy GG 1 | 0.14 (0.06–0.32)* | <0.001* | | | | |

Table 4. Results of the univariate statistical analysis, given for three different patient groupings: (i) entire cohort, (ii) biopsy GG > 1 subgroup patients, (iii) the entire cohort where clinically significant upgraded cases (from GG1 to GG > 1) considered only. Statistically significant results are marked by an asterisks. Significant values are in bold.

| | AUC | Accuracy | Sensitivity | Specificity | Youden index |
|---------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| (a) All patients | | | | | |
| Linear SVM | 0.859 ± 0.005 | 0.800 ± 0.005 | 0.416 ± 0.013 | 0.958 ± 0.006 | 0.374 ± 0.013 |
| RBF SVM | 0.865 ± 0.007 | 0.856 ± 0.004 | 0.621 ± 0.013 | 0.953 ± 0.003 | 0.575 ± 0.013 |
| LASSO | 0.859 ± 0.005 | 0.802 ± 0.004 | 0.452 ± 0.014 | 0.946 ± 0.006 | 0.399 ± 0.012 |
| Ridge | 0.858 ± 0.005 | 0.796 ± 0.005 | 0.449 ± 0.014 | 0.939 ± 0.007 | 0.388 ± 0.012 |
| (b) Biopsy GG > 1 patients only | | | | | |
| Linear SVM | 0.944 ± 0.004 | 0.901 ± 0.005 | 0.637 ± 0.023 | 0.936 ± 0.004 | 0.573 ± 0.024 |
| RBF SVM | 0.894 ± 0.007 | 0.878 ± 0.004 | 0.445 ± 0.024 | 0.936 ± 0.004 | 0.381 ± 0.023 |
| LASSO | 0.930 ± 0.005 | 0.895 ± 0.005 | 0.613 ± 0.024 | 0.933 ± 0.005 | 0.545 ± 0.023 |
| Ridge | 0.944 ± 0.005 | 0.904 ± 0.005 | 0.652 ± 0.023 | 0.938 ± 0.004 | 0.590 ± 0.024 |

Table 5. Metrics of the best performing ML models in feature selection. The mean scores and their standard deviations of randomly selected 100 train-test splits of (a) all patients and (b) biopsy GG > 1 patients only. Significant values are in bold.

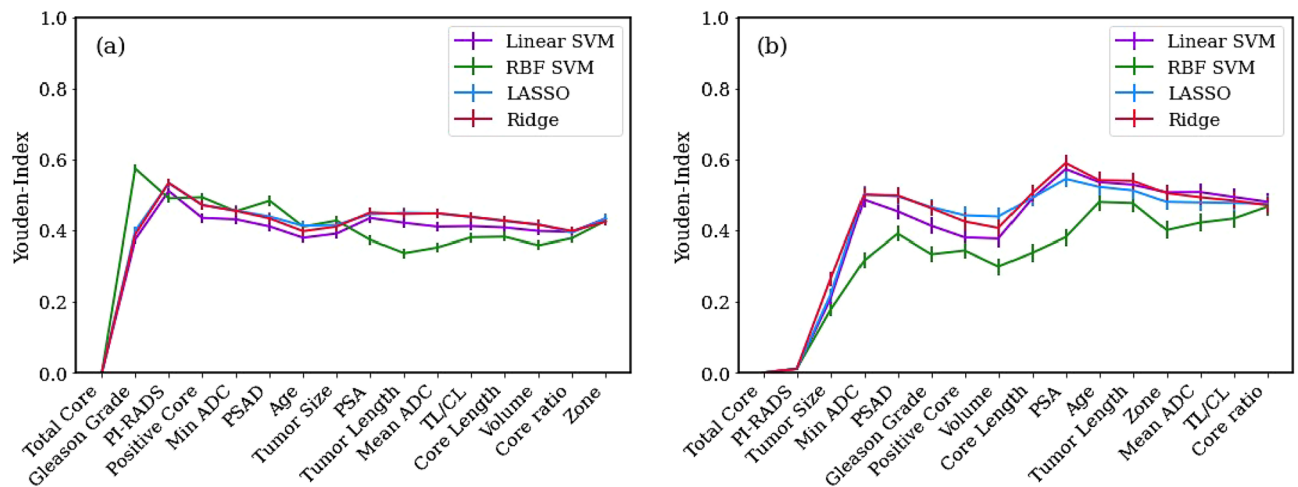


Figure 3. Feature selection by four machine learning models using Youden index as performance metric. Results are shown for (a) overall patient cohort and (b) only biopsy GG > 1 group. Error bars indicate standard error of the mean Youden index.

GG > 1 patients. The number of biopsy cores and tumor size also significantly improved the model performance for the entire cohort and biopsy GG > 1 patients, respectively. Table 5 shows the overall results of the feature selection study with the means and standard errors of the model performance metrics.

The performance of the machine learning models was also evaluated by receiver operating characteristic (ROC) curves, where the area under the curve (AUC) was used for performance assessment. Figure 4a shows the ROC curves for the four classifier models used. The mean AUC was obtained using 100 random train-test splits on the overall patient group. The RBF SVM model outperformed by achieving AUC: 0.865 ± 0.007 . Figure 4b shows the ROC curves computed from biopsy GG > 1 patients only. Compared to those of the previous case, the model performances were enhanced. Ridge regression and linear SVM were favored, with an AUC of 0.944 ± 0.004 .

The use of ML algorithms significantly increased the predictability of GGs at RPs. The final pathological GG estimation accuracy of the ML models reached 0.856 ± 0.004 and 0.904 ± 0.005 for the entire cohort and biopsy GG > 1 patient groups, respectively. Compared to the baseline accuracy established by in-bore biopsy alone, these values indicate 21.4% and 13.1% accuracy enhancement for the two cohorts.

Discussion

Adverse pathology after RP can have serious management consequences, and men with clinically significant disease may be undertreated. Conversely, an overestimated GG would result in overtreatment and hence a reduction in the quality of life of the patient. Therefore, it is of utmost importance to determine the relevant clinical variables that affect GS concordance.

Gleason grade concordance in the literature ranges from 38 to 63%^{17–20}. Upgraded cases occur at a rate of 25% to 56%^{17–20}, significantly outweighing downgrading cases in the majority of the studies, which range from 8% to 16%^{18,19,21}. Although the upgrades in our GG1 group (17 to 20) are remarkable, Costa et al. reported a 66.7% upgrade in the GG1 group²². Liu et al. also reported significant upgrade potential in the GG1 subgroup¹⁸.

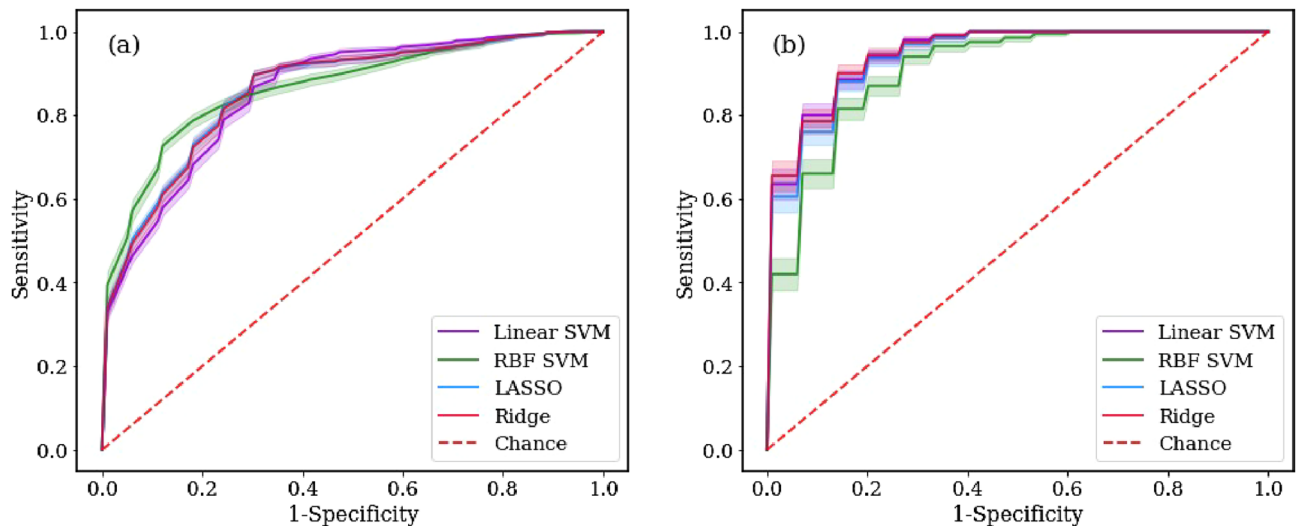


Figure 4. ROC curves for the machine learning models used. Model performances, assessed by AUC metric, using **(a)** overall patient cohort and using **(b)** only biopsy GG > 1 cases are compared. Shaded regions denote standard error of the mean ROC curves obtained by 100 iterations.

In our study, the GG upgrade rate was comparable to that in recent studies executed with MRI-targeted biopsy techniques and significantly lower than that in studies with TRUS-guided systematic biopsy^{23–25}.

A discordant GG between needle biopsy and final pathology is associated with interobserver variability among different pathologists, borderline grades, and more significantly sampling errors²⁶. In support of these arguments, Maruyama et al. reported that GG concordance improved by 6.2% after second-opinion pathology²⁷.

In the literature, multiple clinical variables are reported to be predictive for GG discordance, where GG upgrade indicators include older age²⁰, higher PSA²⁰, lower prostate volume²⁸, higher PSAD²⁸, higher PI-RADS score²⁷, and higher tumor percentage in biopsy cores²⁹.

In our study, univariate analysis revealed that age, prostate volume, PSA, PSAD, PI-RADS score, total biopsy core length, total biopsy tumor length, and tumor percentage in biopsy cores were nonsignificant variables for GG upgrade, whereas the number of biopsy cores, number of positive biopsy cores, Gleason grade, and tumor size were found to be significant predictors of GG upgrade. Although the ADC_{mean} and the ADC_{min} values were found to be irrelevant variables for GG upgrades in univariate analysis, in the GG > 1 patient group, multivariate machine learning analysis found the ADC_{min} value as a useful variable for predicting GG upgrades.

Although the diagnosis of PCa is shifting to targeted biopsy, no agreement has been reached on optimum number of cores. Recent studies showed that more than two biopsy cores had no incremental value in determining the GG^{30,31}, however there are some contrary publications suggesting that additional cores from sextants adjacent to designated target (so called focal saturation) can increase biopsy yield and the concordance between needle and final pathology by excluding the effect of GS heterogeneity³². According to Tracy et al. the likelihood of GG upgrade decreases with an increase in the number of targeted cores³³. Our study results also revealed inverse correlation between number of total and positive cores and GG upgrade likelihood at final pathology. Compared to our study with corresponding 28.4% and 17.9% rates, and 3.4 ± 0.8 cores taken, Ahdoot et al. reported 30.9% and 8.7% rates and Costa et al. reported 13% and 4.4% rates for any GG upgrading and clinically significant upgrading at final pathology with 5.8 ± 2.7 and average 3.2 MRI-targeted cores, respectively^{22,23}.

Intratumoral heterogeneity of tumors is a well known concept and increase in fraction of heterogeneous genetic fusion parallel to tumor size is reported in prostate cancer³⁴. Langer et al. showed that peripheral zone prostate cancer is heterogeneous in nature and 36% percent of tumors consists of scattered few malignant glands intermixed with healthy tissue and classified as sparse tumors³⁵. To our knowledge, the effect of tumor size on GG upgrade was not studied in literature. Our study showed that tumors with larger sizes were upgraded more than tumors with smaller size, which is statistically significant for GG > 1 subgroup ($p = 0.028$). Our analysis revealed 20 mm as a threshold for GG > 1 group, and showed that tumors over 20 mm have a higher possibility to upgrade after biopsy. Due to size criteria of PI-RADS 2.1³⁶, our threshold with 20 mm falls into PI-RADS 5 category. The correlation between PI-RADS scores and Gleason grades is well known^{37–39}, besides that in accordance with our results, the correlation with upgrades of GG and PI-RADS score was demonstrated by Alqahtani et al.⁴⁰. Meta-analysis about active surveillance stated precarious results with active surveillance of PI-RADS 4 and 5, which can be related to our finding with high upgrade ratios in GG1 group⁴¹. In addition to that, our model stated that tumor size has a value for predicting upgrade after in-bore prostate biopsy, which is a novel finding. This finding, if supported by future research with larger series, may have important implications for clinical practice, including considering focal saturation in tumors with large dimensions.

Diffusion weighted imaging is a key component of mp-MRI that contributes to tumor detection, as well as to the assessment of tumor aggressiveness. Tissue microstructure such as dense cellularity or atrophic glands can result in distinct imaging findings. Hambrook et al. showed a high discriminatory performance can be achieved in the differentiation of low, intermediate, and high-grade PCa by ADC value⁴². In active surveillance patient group, ADC value was identified as an independent predictor of both upgrading on repeat biopsy and

time to radical therapy^{26,43,44}. Park et al. reported a significant inverse correlation between the ADC_{mean} and the ADC_{min} values and the possibility of GG upgrade in a patient group of GG1⁴⁵. In our study, statistical analysis revealed no significant correlation between the ADC_{mean} and ADC_{min} values and GG upgrade in both groups of patients whereas in ML studies ADC_{min} value was found to be useful in the prediction of GG upgrade in the GG > 1 patient group. This discrepancy between ADC_{min} and ADC_{mean} can be explained by the heterogeneous nature of PCa.

Various ML algorithms previously used for GG upgrade prediction are logistic regression¹⁸, LASSO regression^{18,46}, SVM^{18,47}, k-Nearest Neighbours (kNN)⁴⁶, decision trees⁴⁶, and random forests^{18,46}. Due to the lack of large datasets, medical problems pose a particular challenge for ML models. Many machine learning algorithms require a considerable amount of data. Otherwise, the ML model may overfit the training data and generate poor results on the tests. For this reason, ML models such as decision tree and random forest that require massive datasets are not suitable candidates for our problem, agreed by the former studies in literature^{18,46}. SVM, Ridge and LASSO models were used in this study as they are less prone to overfitting for relatively small datasets.

Our results show ML-assisted GG estimation accuracy was increased by 21.4% for the overall patient group, surpassing the 13.1% enhancement for upgrade estimations among GG > 1 cases, in line with the literature where Liu et al.¹⁸ showed ML application improved the prediction accuracy from 39.2% to 71.2%. These accuracy enhancements indicate ML models are useful tools to utilize clinical records for personalized treatment planning. Moreover, ML models unraveled the significance of more clinical features than revealed by statistics alone such as ADC_{min} (see Fig. 3b), outlining the power of ML concept, where features considered statistically insignificant can be utilized for predictive models.

The potential limitations of this study are retrospective design, small sample size that affects both statistical and ML studies, and possible increase in selection bias due to recruitment of patients over 8 years of time. Biopsy GG1 patients upgraded at RP pathology more often compared to other biopsy Gleason grade groups. The reason for this may be due to bias in data collection, as most low-risk GG1 patients are assigned to active surveillance rather than RP. Additionally, even though a 3-fold cross-validation strategy was employed in our study, an external validation is crucial for confirming the model's effectiveness and applicability in different clinical settings. Future studies should aim to incorporate such validation to ensure the model's reliability and utility in the clinical management of prostate cancer, enhancing its potential contribution to personalized patient care.

Our study pioneers the application of machine learning methodologies to predict upgrades in MRI-guided in-bore biopsy patients, boasting the second-largest study population, which compares MRI-guided in-bore biopsy and radical prostatectomy results⁴⁸. Overall, our study suggests that a combination of clinical factors (the number of biopsy cores, the number of positive biopsy cores, Gleason grade, tumor size and ADC_{min} value) and machine learning models may be valuable in predicting the likelihood of GG upgrade following RP and could potentially improve patient outcomes.

Conclusion

Determining the relevant clinical variables that affect GS concordance in MRI-targeted biopsy is of utmost importance in the era of MRI pathway. Univariate statistics revealed the number of biopsy cores, number of positive biopsy cores, and Gleason grade were statistically significant GG upgrade indicators and inversely correlated to GG upgrade possibility. Machine learning analysis found the ADC_{min} value as a useful variable in the prediction of GG upgrade. As a novel finding, tumor size measured by mpMRI is shown to be positively correlated with GG upgrade likelihood for GG > 1 subgroup. Tumor size and ADC_{min} can be useful markers to assess risk of upgrade prior to biopsy, so biopsy number and patient selection for active surveillance can be decided in terms of these markers. The findings of our study contribute to identifying patients predisposed to GG upgrade during RP. By comparing patient characteristics with our documented outcomes, we can pinpoint high-risk cases for GG upgrade and potentially adjust the threshold for performing RP in such cases.

Data availability

The data that support the findings of this study are available upon reasonable request from the corresponding author.

Received: 30 December 2023; Accepted: 6 March 2024

Published online: 11 March 2024

References

1. Rawla, P. Epidemiology of prostate cancer. *World J. Oncol.* **10**, 63 (2019).
2. Siu, W., Dunn, R. L., Shah, R. B. & Wei, J. T. Use of extended pattern technique for initial prostate biopsy. *J. Urol.* **174**, 505–509 (2005).
3. Djavan, B. et al. Prospective evaluation of prostate cancer detected on biopsies 1, 2, 3, and 4: When should we stop?. *J. Urol.* **166**, 1679–1683 (2001).
4. Venderink, W. et al. Multiparametric magnetic resonance imaging for the detection of clinically significant prostate cancer: What urologists need to know. Part 3: Targeted biopsy. *Eur. Urol.* **77**, 481–490 (2020).
5. Prince, M. et al. In-bore versus fusion MRI-targeted biopsy of PI-RADS category 4 and 5 lesions: A retrospective comparative analysis using propensity score weighting. *Am. J. Roentgenol.* **217**, 1123–1130 (2021).
6. Kilic, M. et al. Pathological accuracy in prostate cancer: Single-center outcomes of 3 different magnetic resonance imaging-targeted biopsy techniques and random systematic biopsy. *Türk Üroloji Dergisi/Turk. J. Urol.* **48**, 346–353 (2022).
7. Sanmugalingam, N. et al. The PRECISE recommendations for prostate MRI in patients on active surveillance for prostate cancer: A critical review. *Am. J. Roentgenol.* **221**, 649–660 (2023).
8. Kilic, M. et al. Accuracy of sampling PI-RADS 4–5 index lesions alone by MRI-guided in-bore biopsy in biopsy-naive patients undergoing radical prostatectomy. *Eur. Urol. Focus* **6**, 249–254 (2020).

9. Weinreb, J. C. *et al.* PI-RADS prostate imaging - reporting and data system: 2015, version 2. *Eur. Urol.* **69**, 16–40 (2016).
10. Eggener, S. E. *et al.* Focal therapy for localized prostate cancer: A critical appraisal of rationale and modalities. *J. Urol.* **178**, 2260–2267 (2007).
11. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
12. Klotz, L. Prostate cancer overdiagnosis and overtreatment. *Curr. Opin. Endocrinol. Diabetes Obes.* **20**, 204–209 (2013) (ISSN: 1752-296X).
13. Bedi, N., Reddy, D. & Ahmed, H. U. Targeting the cancer lesion, not the whole prostate. *Transl. Androl. Urol.* **9**, 1518 (2019).
14. Vural, M. *et al.* In-bore MRI-guided prostate biopsy in a patient group with PI-RADS 4 and 5 targets: A single center experience. *Eur. J. Radiol.* **141**, 109785 (2021).
15. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006).
16. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
17. Erdem, S. *et al.* The clinical predictive factors and postoperative histopathological parameters associated with upgrading after radical prostatectomy: A contemporary analysis with grade groups. *Prostate* **80**, 225–234 (2020).
18. Liu, H. *et al.* Predicting prostate cancer upgrading of biopsy Gleason grade group at radical prostatectomy using machine learning-assisted decision-support models. *Cancer Manag. Res.* **12**, 13099–13110 (2020).
19. Bullock, N. *et al.* Pathological upgrading in prostate cancer treated with surgery in the United Kingdom: trends and risk factors from the British Association of Urological Surgeons Radical Prostatectomy Registry. *BJU Int.* **19**, 1–9 (2019).
20. Soenen, C. *et al.* Concordance between biopsy and radical prostatectomy Gleason scores: Evaluation of determinants in a large-scale study of patients undergoing RARP in Belgium. *Pathol. Oncol. Res.* **26**, 2605–2612 (2020).
21. Yu, A. *et al.* Combination MRI-targeted and systematic prostate biopsy may overestimate gleason grade on final surgical pathology and impact risk stratification. *Urologic Oncology: Seminars and Original Investigations* **40**, 591 (2022).
22. Costa, D. N. *et al.* Gleason grade group concordance between preoperative targeted biopsy and radical prostatectomy histopathologic analysis: A comparison between in-bore MRI-guided and MRI-transrectal US fusion prostate biopsies. *Radiol. Imaging Cancer* **3**, e200123 (2021).
23. Ahdoot, M. *et al.* MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. *N. Engl. J. Med.* **382**, 917–928 (2020).
24. Nawfal, G. *et al.* Multiparametric MRI with in-bore targeted biopsy in the diagnostic pathway of prostate cancer: Data from a single institution experience. *Urol. Oncol. Semin. Orig. Investig.* **39**, 781–e1 (2021).
25. Coogan, C. L. *et al.* Increasing the number of biopsy cores improves the concordance of biopsy Gleason score to prostatectomy Gleason score. *BJU Int.* **96**, 324–327 (2005).
26. Kaufmann, S. *et al.* Prostate cancer detection in patients with prior negative biopsy undergoing cognitive-, robotic- or in-bore MRI target biopsy. *World J. Urol.* **36**, 761–768 (2018).
27. Maruyama, Y. *et al.* Factors predicting pathological upgrading after prostatectomy in patients with Gleason grade group 1 prostate cancer based on opinion-matched biopsy specimens. *Mol. Clin. Oncol.* **12**, 384–389 (2020).
28. Alchin, D. R. *et al.* What are the predictive factors for Gleason score upgrade following RP?. *Urol. Int.* **96**, 1–4 (2016).
29. Epstein, J. I. *et al.* Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: Incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. *Eur. Urol.* **61**, 1019–1024 (2012).
30. Borkowetz, A. *et al.* Direct comparison of multiparametric magnetic resonance imaging (MRI) results with final histopathology in patients with proven prostate cancer in MRI/ultrasonography-fusion biopsy. *BJU Int.* **118**, 213–220 (2016).
31. Hollenbeck, B. K. *et al.* Whole mounted radical prostatectomy specimens do not increase detection of adverse pathological features. *J. Urol.* **164**, 1583–1586 (2000).
32. Padhani, A. R. *et al.* PI-RADS steering committee: The PI-RADS multiparametric MRI and MRI-directed biopsy pathway. *Radiology* **292**, 464–474 (2019).
33. Tracy, C. R. *et al.* Optimizing MRI-targeted prostate biopsy: The diagnostic benefit of additional targeted biopsy cores. *Urol. Oncol. Semin. Orig. Investig.* **39**(193), e1–193.e6 (2021).
34. Tsourlakis, M.-C. *et al.* Heterogeneity of ERG expression in prostate cancer: A large section mapping study of entire prostatectomy specimens from 125 patients. *BMC Cancer* **16**, 641 (2016).
35. Langer, D. L. *et al.* Intermixed normal tissue within prostate cancer: Effect on MR imaging measurements of apparent diffusion coefficient and T2-sparse versus dense cancers. *Radiology* **249**, 900–908 (2008).
36. Turkbey, B. *et al.* Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur. Urol.* **76**, 340–351 (2019).
37. Walker, S. M. *et al.* Prospective evaluation of PI-RADS version 2.1 for prostate cancer detection. *Am. J. Roentgenol.* **215**, 1098–1103 (2020).
38. Cash, H. *et al.* The detection of significant prostate cancer is correlated with the prostate imaging reporting and data system (PI-RADS) in MRI/transrectal ultrasound fusion biopsy. *World J. Urol.* **34**, 525–532 (2016).
39. Lim, C. S. *et al.* Prognostic value of prostate imaging and data reporting system (PI-RADS) v. 2 assessment categories 4 and 5 compared to histopathological outcomes after radical prostatectomy. *J. Magn. Reson. Imaging* **46**, 257–266 (2017).
40. Alqahtani, S. *et al.* Prediction of prostate cancer Gleason score upgrading from biopsy to radical prostatectomy using pre-biopsy multiparametric MRI PIRADS scoring system. *Sci. Rep.* **10**, 7722 (2020).
41. Zhai, L. *et al.* The role of prostate imaging reporting and data system score in Gleason 3+3 active surveillance candidates enrollment: A diagnostic meta-analysis. *Prostate Cancer Prostat. Dis.* **22**, 235–243 (2019).
42. Hambrock, T. *et al.* Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology* **259**, 453–461 (2011).
43. Dickinson, L. *et al.* Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: Recommendations from a European Consensus Meeting. *Eur. Urol.* **59**, 477–494 (2011).
44. Grignon, D. Prostate cancer reporting and staging: Needle biopsy and radical prostatectomy specimens. *Mod. Pathol.* **31**, 96–109 (2018).
45. Park, S. Y. *et al.* Diffusion-weighted imaging predicts upgrading of Gleason score in biopsy-proven low grade prostate cancers. *BJU Int.* **119**, 57–66 (2017).
46. Xie, J. *et al.* Prediction of pathological upgrading at radical prostatectomy in prostate cancer eligible for active surveillance: A texture features and machine learning-based analysis of apparent diffusion coefficient maps. *Front. Oncol.* **10**, 604266 (2021).
47. Citak-Er, F. *et al.* Final Gleason score prediction using discriminant analysis and support vector machine based on preoperative multiparametric MR imaging of prostate cancer at 3T. *BioMed. Res. Int.* **2014**, 690787 (2014).
48. van der Leest, M. *et al.* Head-to-head comparison of transrectal ultrasound-guided prostate biopsy versus multiparametric prostate resonance imaging with subsequent magnetic resonance-guided biopsy in biopsy-naïve men with elevated prostate-specific antigen: A large prospective multicenter clinical study. *Eur. Urol.* **75**, 570–578 (2019).

Author contributions

M.V. provided study concept and design. S.D., D.C., M.K., and M.V. collected and organized data. K.O. and I.L. analyzed the data. K.O., I.L., S.D. and D.C. interpreted the results. K.O., I.L., S.D., D.C., and M.V. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-56415-5>.

Correspondence and requests for materials should be addressed to K.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024