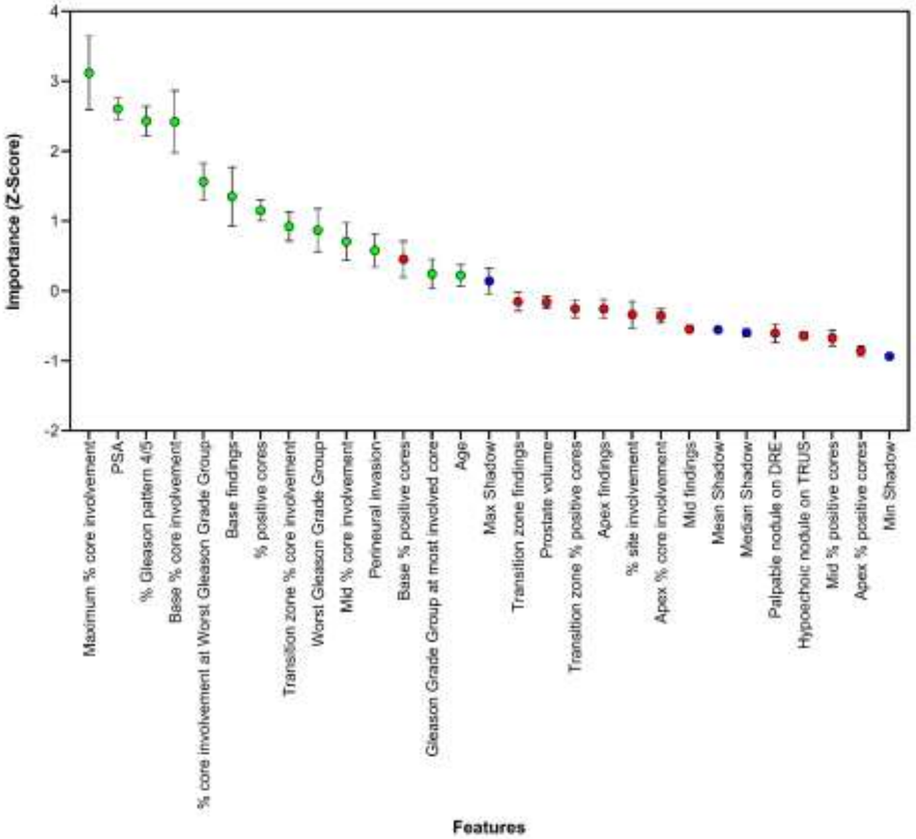
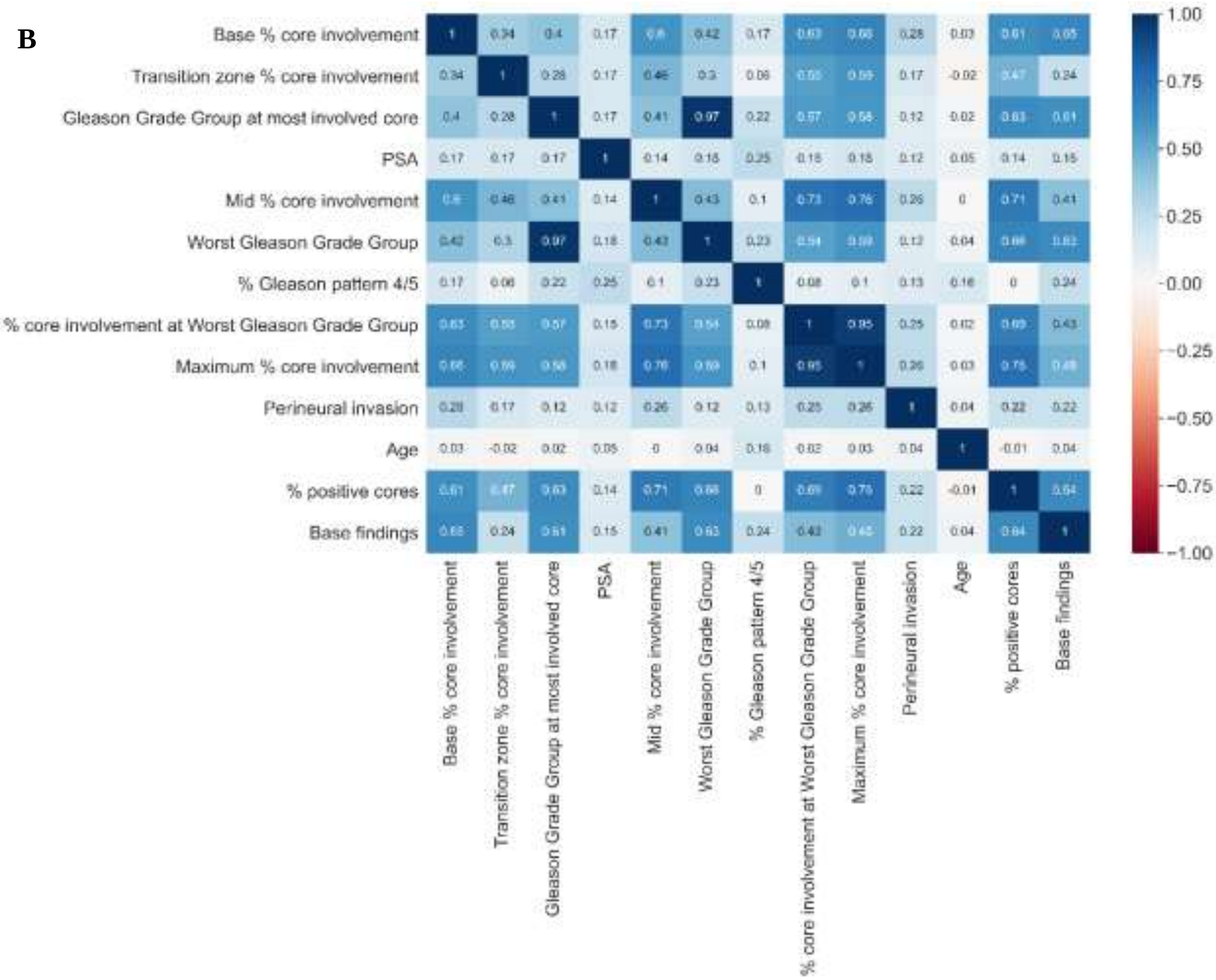
Kwong JCC, et al. Explainable artificial intelligence to predict the risk of side-specific extraprostatic extension in pre-prostatectomy patients

APPENDIX

Supplementary Figure 1. Data preparation using dimensionality reduction. This is a stepwise reduction of the training cohort to select the most important and independent features.
(A) Relevant features were selected using Boruta feature selection. This method involves comparing the importance (Z-score) of each feature to random noise (i.e., shadow attributes) using a random forest model. Shadow attributes are created for each feature and features with higher Z-scores than the maximum Z-score among shadow attributes are considered important, while those with lower Z-scores are removed from the dataset. The graphical output of the Boruta feature selection process with the importance (Z-score) of each feature based on SHAP is shown. Green values are features that have been selected (higher Z-score than maximum Z-score of shadow attributes). Red values are features that have been rejected (lower Z-score than maximum Z-score of shadow attributes). Blue values represent metrics of the shadow attributes.
(B) Next, highly correlated features were removed using a cutoff of Pearson correlation >0.8. A correlogram illustrating the Pearson correlation coefficients of the 13 most important features identified using Boruta feature selection is shown. Here, one can see that Gleason grade group at most involved core and % core involvement at "Worst Gleason grade group" exhibited high collinearity and were removed to generate the final feature subset for model training and hyperparameter tuning.

A





	Base % core involvement	Transition zone % core involvement	Gleason Grade Group at most involved core	PSA	Mid % core involvement	Worst Gleason Grade Group	% Gleason pattern 4/5	% core involvement at Worst Gleason Grade Group	Maximum % core involvement	Perineural invasion	Age	% positive cores	Base findings
Base % core involvement	1	0.34	0.4	0.17	0.6	0.42	0.17	0.63	0.66	0.28	0.03	0.61	0.65
Transition zone % core involvement	0.34	1	0.28	0.17	0.46	0.3	0.06	0.55	0.59	0.17	-0.02	0.47	0.24
Gleason Grade Group at most involved core	0.4	0.28	1	0.17	0.41	0.97	0.22	0.57	0.58	0.12	0.02	0.63	0.61
PSA	0.17	0.17	0.17	1	0.14	0.18	0.25	0.15	0.18	0.12	0.05	0.14	0.15
Mid % core involvement	0.6	0.46	0.41	0.14	1	0.43	0.1	0.73	0.76	0.26	0	0.71	0.41
Worst Gleason Grade Group	0.42	0.3	0.97	0.18	0.43	1	0.23	0.54	0.59	0.12	0.04	0.66	0.63
% Gleason pattern 4/5	0.17	0.06	0.22	0.25	0.1	0.23	1	0.08	0.1	0.13	0.16	0	0.24
% core involvement at Worst Gleason Grade Group	0.63	0.55	0.57	0.15	0.73	0.54	0.08	1	0.95	0.25	0.02	0.69	0.43
Maximum % core involvement	0.66	0.59	0.58	0.18	0.76	0.59	0.1	0.95	1	0.26	0.03	0.75	0.48
Perineural invasion	0.28	0.17	0.12	0.12	0.26	0.12	0.13	0.25	0.26	1	0.04	0.22	0.22
Age	0.03	-0.02	0.02	0.05	0	0.04	0.16	0.02	0.03	0.04	1	-0.01	0.04
% positive cores	0.61	0.47	0.63	0.14	0.71	0.66	0	0.69	0.75	0.22	-0.01	1	0.64
Base findings	0.65	0.24	0.61	0.15	0.41	0.63	0.24	0.43	0.48	0.22	0.04	0.64	1

Supplementary Table 1. Data dictionary of all features and the label of interest.

	Data type	**Description of method of collection or measurement**	**Range of values for numerical features, coded values for categorical features**
Clinical features			
Age*	Numerical	Age at the time of radical prostatectomy, in years	39.8–74.9
PSA*	Numerical	Most recent PSA at time of radical prostatectomy, ng/mL	1.05–65
Global biopsy features			
% Gleason pattern 4/5*	Numerical	From prostate biopsy report, in %	0–100
Perineural invasion*	Binary	From prostate biopsy report	0=No; 1=Yes
Prostate volume*	Numerical	From radiology report at the time of transrectal ultrasound-guided prostate biopsy, in mL	3.49–115.7
Side-specific features (i.e., left or right prostatic lobe)			
Palpable nodule on DRE*	Binary	Direct from clinical notes based on DRE by the urologist	0=No, 1=Yes
Hypoechoic nodule on TRUS*	Binary	Direct from radiology report at the time of transrectal ultrasound guided prostate biopsy	0=No, 1=Yes
% site involvement	Numerical	Calculated from prostate biopsy report. Number of sites (base, mid, apex, transition zone) with positive biopsy cores divided by total number of sites biopsied on the ipsilateral side, in %	0–100
% positive cores*	Numerical	Calculated from prostate biopsy report. Number of positive biopsy cores divided by total number of cores taken on the ipsilateral side, in %	0–100
Worst Gleason grade group*	Categorical	Direct from prostate biopsy report. Highest Gleason grade group among the ipsilateral cores	0=Normal 1=HGPIN 2= ASAP 3=Grade group 1 4 =Grade group 2 5=Grade group 3 6=Grade group 4 7=Grade group 5

% core involvement at worst Gleason grade group	Numerical	Direct from prostate biopsy report. Percentage of tumour involvement in the core with the highest Gleason grade group among the ipsilateral cores, in %	0–100
Maximum % core involvement*	Numerical	Direct from prostate biopsy report. Highest percentage of tumor involvement in a single core among the ipsilateral cores, in %	0–100
Gleason grade group at most involved core	Categorical	Direct from prostate biopsy report. Gleason grade group in the single core with the highest percentage of tumour involvement among the ipsilateral cores	0=Normal 1=HGPIN 2=ASAP 3=Grade group 1 4=Grade group 2 5=Grade group 3 6=Grade group 4 7=Grade group 5
Base findings	Categorical	Direct from prostate biopsy report. Highest Gleason grade group among the ipsilateral basal cores	0=Normal 1=HGPIN 2=ASAP 3=Grade group 1 4=Grade group 2 5=Grade group 3 6=Grade group 4 7=Grade group 5
Base % positive cores	Numerical	Calculated from prostate biopsy report. Number of positive basal cores divided by total number of ipsilateral basal cores taken, in %	0–100
Base % core involvement	Numerical	Direct from prostate biopsy report. Percentage of tumor involvement among the ipsilateral basal cores, in %	0–100
Mid findings	Categorical	Direct from prostate biopsy report. Highest Gleason grade group among the ipsilateral mid cores	0=Normal 1=HGPIN 2=ASAP 3=Grade group 1 4=Grade group 2 5=Grade group 3 6=Grade group 4 7=Grade group 5
Mid % positive cores	Numerical	Calculated from prostate biopsy report. Number of positive mid cores divided by total number of ipsilateral mid cores taken, in %	0–100

Mid % core involvement	Numerical	Direct from prostate biopsy report. Percentage of tumor involvement among the ipsilateral mid cores, in %	0–100
Apex findings	Categorical	Direct from prostate biopsy report. Highest Gleason grade group among the ipsilateral apical cores	0=Normal 1=HGPIN 2=ASAP 3=Grade group 1 4=Grade group 2 5=Grade group 3 6=Grade group 4 7=Grade group 5
Apex % positive cores	Numerical	Calculated from prostate biopsy report. Number of positive apical cores divided by total number of ipsilateral apical cores taken, in %	0–100
Apex % core involvement	Numerical	Direct from prostate biopsy report. Percentage of tumour involvement among the ipsilateral apical cores, in %	0–100
Transition zone findings	Categorical	Direct from prostate biopsy report. Highest Gleason grade group among the ipsilateral transition zone cores	0=Normal 1=HGPIN 2=ASAP 3=Grade group 1 4=Grade group 2 5=Grade group 3 6=Grade group 4 7=Grade group 5
Transition zone % positive cores	Numerical	Calculated from prostate biopsy report. Number of positive transition zone cores divided by total number of ipsilateral transition zone cores taken, in %	0–100
Transition zone % core involvement	Numerical	Direct from prostate biopsy report. Percentage of tumor involvement among the ipsilateral transition zone cores, in %	0–100
Label			
ssEPE	Binary	Direct from pathology report of prostatectomy specimen. Presence of tumor that has extended beyond the prostatic capsule on the ipsilateral lobe	0 = No; 1 = Yes

Features marked with an asterisk were selected a priori based on literature review.
ASAP: atypical small acinar proliferation; DRE: digital rectal exam; HGPIN: high-grade prostatic intraepithelial neoplasia; IQR: interquartile range; PSA: prostate-specific antigen; SD: standard deviation; ssEPE: side-specific extraprostatic extension; TRUS: transrectal ultrasound.

Supplementary Table 2. Final ML model and hyperparameter specifications	
Model version	XGBoost version 1.3.3
Model class	XGB Classifier
Hyperparameter search space	n_estimators: 600–1200 max_depth: 7–11 subsample: 0.8, 0.9 base_score: 0.3 (based on reported incidence of ssEPE) learning_rate: 0.01, 0.05, 0.1 colsample_bylevel=0.5 colsample_bynode=0.5 colsample_bytree=0.5
Final hyperparameters	n_estimators=831 max_depth=10 base_score=0.3 scale_pos_weight=1 subsample=0.8 learning_rate=0.05 gamma=5 booster='gbtree' colsample_bylevel=0.5 colsample_bynode=0.5 colsample_bytree=0.5 eval_metric='auc' importance_type='gain' min_child_weight=1 num_parallel_tree=1 tree_method='exact' reg_alpha=0 reg_lambda=1 random_state=42 objective='binary:logistic'
Final features	Age PSA % Gleason pattern 4/5 Perineural invasion % positive cores Worst Gleason Grade Group Maximum % core involvement Base finding Base % core involvement Mid % core involvement Transition zone % core involvement

Hyperparameter tuning involves adjusting model parameters to optimize performance. A grid search with stratified tenfold cross-validation and mean AUROC as the scoring metric was used to select the final hyperparameters. In stratified tenfold cross-validation, the training cohort is randomly partitioned into ten equal folds, with each fold containing the same proportion of ssEPE cases. Nine folds are used for model training and hyperparameter tuning while the remaining fold makes up the validation cohort. This process was repeated ten times such that each fold served as the validation cohort once. The final model and combination of hyperparameters were determined based on the highest mean AUROC across all validation cohorts to improve generalizability of the model. AUROC: area under receiver operating characteristic; ML: machine learning; PSA: prostate-specific antigen; ssEPE: side-specific extraprostatic extension.

Supplementary Table 3. Comparison of baseline characteristics between the training and testing cohorts for this study and Sayyid et al

	Kwong et al		**Sayyid et al**	
	Training cohort	**Testing cohort**	**Training cohort**	**Testing cohort**
No. of lobes	900	122	1506	622
Age, median (IQR)	62 (57–66)	62 (57–65)	62 (57–67)	61 (56–65)
PSA (ng/mL), median (IQR)	7.06 (5.50–9.30)	8.20 (6.00–12.20)	5.70 (4.34–8.15)	5.00 (4.00–8.00)
Prostate volume (mL), median (IQR)	34 (25–44)	35 (27–42)	34 (27–44)	35 (28–46)
Palpable nodule on DRE, n (%)	192 (21.3)	27 (22.1)	298 (20.1)	145 (28.9)
Hypoechoic nodule on TRUS, n (%)	106 (11.8)	13 (10.7)	494 (33.2)	98 (19.5)
% positive cores, median (IQR)	33.3 (14.3–66.7)	42.9 (16.7–83.3)	33.0 (13.0–60.0)	33.0 (20.0–60.0)
Worst Gleason grade group, n (%)				
Normal	183 (20.3)	23 (18.9)	225 (14.9)	40 (8.0)
HGPIN	11 (1.2)	0 (0)	128 (8.5)	36 (7.2)
ASAP	6 (0.7)	0 (0)	–	–
Grade group 1	208 (23.1)	22 (18.0)	433 (28.8)	232 (46.2)
Grade group 2	320 (35.6)	42 (34.4)	470 (31.2)	104 (20.7)
Grade group 3	108 (12.0)	21 (17.2)	146 (9.7)	52 (10.4)
Grade group 4	32 (3.6)	9 (7.4)	68 (4.5)	25 (5.0)
Grade group 5	32 (3.6)	5 (4.1)	36 (2.4)	13 (2.6)
Maximum % core involvement, median (IQR)	20.0 (5.0–50.0)	40.0 (5.0–75.0)	15.0 (1.0–50.0)	5.0 (0.0–8.0)
ssEPE, n (%)	276 (30.7)	51 (41.8)	298 (19.8)	145 (28.9)

DRE: digital rectal exam; HGPIN: high-grade prostatic intraepithelial neoplasia; IQR: interquartile range; PSA: prostate-specific antigen; ssEPE: side-specific extraprostatic extension; TRUS: transrectal ultrasound.

Supplementary Table 4. Bias assessment of the ML, LR, and baseline models based on patient demographics and disease characteristics

		Baseline		**LR**		**ML**	
		AUROC (95% CI)	**AUPRC (95% CI)**	**AUROC (95% CI)**	**AUPRC (95% CI)**	**AUROC (95% CI)**	**AUPRC (95% CI)**
Age	**≤60 years**	0.79 (0.74–0.83)	0.64 (0.56–0.72)	0.81 (0.77–0.85)	0.68 (0.61–0.75)	0.84 (0.80–0.88)	0.73 (0.65–0.79)
	>60 years	0.74 (0.70–0.77)	0.63 (0.56–0.68)	0.77 (0.73–0.80)	0.66 (0.60–0.72)	0.78 (0.74–0.81)	0.68 (0.63–0.73)
Institution	**CVH**	0.74 (0.70–0.77)	0.59 (0.53–0.66)	0.78 (0.74–0.81)	0.64 (0.58–0.70)	0.81 (0.77–0.83)	0.69 (0.63–0.74)
	MH	0.75 (0.66–0.83)	0.70 (0.59–0.81)	0.76 (0.67–0.84)	0.75 (0.65–0.85)	0.81 (0.73–0.88)	0.78 (0.67–0.87)
D'Amico risk classification	**Intermediate**	0.69 (0.65–0.73)	0.57 (0.51–0.63)	0.72 (0.68–0.76)	0.62 (0.56–0.68)	0.76 (0.72–0.80)	0.64 (0.58–0.70)
	High	0.71 (0.63–0.78)	0.80 (0.71–0.86)	0.76 (0.68–0.83)	0.85 (0.78–0.90)	0.81 (0.74–0.87)	0.89 (0.83–0.92)

AUROC: area under the receiver-operating-characteristic curve; AUPRC: area under the precision-recall curve; CI: confidence interval; CVH: Credit Valley Hospital; LR: logistic regression; MH: Mississauga Hospital; ML: machine learning.

Supplementary Table 5. Comparison of contemporary predictive models for side-specific extraprostatic extension that have been externally validated

	Kwong et al	**Sayyid et al**[1]	**Martini et al**[2]	**Soeterik et al**[3]
Sample size of training cohort (incidence of ssEPE in %)	900 (30.7)	1506 (19.8)	829 (17.1)	1774 (25.8)
AUROC on training cohort	0.81	0.88	0.82	0.80
AUROC of external validation cohort (incidence of ssEPE in %)	0.81 (41.8)	0.74 (28.9)	0.68 (29.2)[4] 0.78 (32.0)[5]	0.83 (21.9) 0.77 (15.8)
MRI-specific findings included	No	No	Yes	Yes
Variables	1. Age 2. PSA 3. % Gleason pattern 4/5 4. Perineural invasion 5. % positive cores 6. Worst Gleason grade group 7. Maximum % core involvement 8. Base finding 9. Base % core involvement 10. Mid % core involvement 11. Transition zone % core involvement	1. Age 2. PSA 3. Prostate volume 4. DRE positivity 5. Hypoechoic nodule present 6. Side-specific percent positive cores 7. Side-specific highest core involvement 8. Side-specific Gleason grade group	1. PSA 2. Side-specific Gleason grade 3. Side-specific max percent tumor involvement in highest Gleason core 4. Side-specific documented EPE on mpMRI	1. PSA density 2. MRI findings (no lesion, lesion but no EPE, EPE) 3. Worst Gleason grade group

AUROC: area under the receiver-operating-characteristic curve; DRE: digital rectal exam; MRI: magnetic resonance imaging; PSA: prostate-specific antigen; ssEPE: side-specific extraprostatic extension.

References

1. Sayyid R, Perlis N, Ahmad A, et al. Development and external validation of a biopsy-derived nomogram to predict risk of ipsilateral extraprostatic extension. *BJU Int* 2017;120:76-82. https://doi.org/10.1111/bju.13733
2. Martini A, Gupta A, Lewis SC, et al. Development and internal validation of a side-specific, multiparametric magnetic resonance imaging-based nomogram for the prediction of extracapsular extension of prostate cancer. *BJU Int* 2018;122:1025-33. https://doi.org/10.1111/bju.14353
3. Soeterik TFW, van Melick HHE, Dijksman LM, et al. Development and external validation of a novel nomogram to predict side-specific extraprostatic extension in patients with prostate cancer undergoing radical prostatectomy. *Eur Urol Oncol* 2020; Epub ahead of print. https://doi.org/10.1016/j.euo.2020.08.008
4. Sighinolfi MC, Sandri M, Torricelli P, et al. External validation of a novel side-specific, multiparametric magnetic resonance imaging-based nomogram for the prediction of extracapsular extension of prostate cancer: Preliminary outcomes on a series diagnosed with multiparametric magnetic resonance imaging-targeted plus systematic saturation biopsy. *BJU Int* 2019;124:192-4. https://doi.org/10.1111/bju.14665
5. Soeterik TFW, van Melick HHE, Dijksman LM, et al. External validation of the Martini nomogram for prediction of side-specific extraprostatic extension of prostate cancer in patients undergoing robot-assisted radical prostatectomy. *Urol Oncol Semin Orig Investig* 2020;38:372-8. https://doi.org/10.1016/j.urolonc.2019.12.028