

Supplementary Material

1 Results

1.1 Quantitative evaluation

Tables S1, S2 and S3 present detailed versions of the Tables 3, 4 and 5 presented in the manuscript. Model rankings for each metric are presented in brackets beside the model performance. The sum of ranks column shows how the individual metric ranks were aggregated to derive the final rank for each model.

Table S1: Detecting cancer: Lesion-level evaluation on cohorts C1-test and cohort C2. Mean and standard deviation values for each metric in each cohort is reported. Numbers in brackets indicate model rank in individual metric. Column ‘‘Sum of ranks’’ represents the sum of the individual metric ranks. Column ‘‘Final rank’’ represents rank of model based on sum of individual ranks. Note, the lower the sum rank, the higher ranked (i.e. better) the model. The first three rows above the dotted lines in both cohorts represent the MRI-only models, while the rows below the dotted lines represent the MRI+correlated features-based models. CorrSigNIA outperforms all other models in both cohorts.

Evaluation on cohort C1-test											
Patients with cancer lesion volumes $\geq 250 \text{ mm}^3$ N = 37, number of cancerous lesions = 44											
Model	ROC AUC	PR AUC	Sensitivity	Specificity	Precision	NPV	F1 Score	Dice	Accuracy	Sum of Ranks	Final Rank
SPCNet	0.82±0.30 (1)	0.76±0.36 (5)	0.62±0.47 (11)	0.75±0.30 (5)	0.44±0.41 (5)	0.84±0.22 (4)	0.48±0.40 (7)	0.27±0.26 (13)	0.72±0.24 (5)	56	5
U-Net	0.82±0.34 (2)	0.81±0.34 (1)	0.84±0.35 (4)	0.30±0.31 (11)	0.35±0.27 (11)	0.57±0.47 (11)	0.45±0.25 (11)	0.30±0.20 (2)	0.43±0.26 (12)	65	8
BrU-Net	0.80±0.32 (5)	0.78±0.34 (2)	0.82±0.35 (5)	0.39±0.38 (10)	0.37±0.29 (10)	0.57±0.46 (10)	0.47±0.28 (9)	0.27±0.17 (8)	0.52±0.27 (10)	69	9
CorrSigNIA	0.81±0.31 (3)	0.73±0.38 (8)	0.73±0.43 (6)	0.72±0.33 (7)	0.50±0.41 (1)	0.86±0.25 (3)	0.55±0.39 (1)	0.30±0.24 (4)	0.72±0.26 (6)	39	1
SPCNet + CR-3-S	0.79±0.31 (8)	0.72±0.38 (10)	0.62±0.46 (10)	0.80±0.21 (2)	0.45±0.39 (4)	0.88±0.15 (1)	0.49±0.38 (6)	0.27±0.23 (10)	0.75±0.17 (1)	52	4
SPCNet + CR-123-I	0.79±0.33 (9)	0.72±0.37 (9)	0.66±0.47 (9)	0.77±0.29 (4)	0.48±0.42 (3)	0.87±0.21 (2)	0.52±0.41 (3)	0.29±0.24 (6)	0.75±0.23 (3)	48	2
SPCNet + CR-123-S	0.73±0.35 (12)	0.70±0.38 (12)	0.55±0.48 (13)	0.82±0.24 (1)	0.41±0.41 (8)	0.83±0.21 (5)	0.44±0.40 (12)	0.27±0.24 (12)	0.75±0.20 (2)	77	11
SPCNet + CR-4-I	0.80±0.34 (6)	0.77±0.36 (4)	0.58±0.47 (12)	0.80±0.30 (3)	0.44±0.42 (6)	0.82±0.25 (6)	0.47±0.42 (10)	0.28±0.23 (7)	0.73±0.24 (4)	58	7
SPCNet + CR-4-S	0.79±0.34 (10)	0.75±0.37 (6)	0.69±0.44 (7)	0.72±0.32 (6)	0.50±0.41 (1)	0.82±0.28 (7)	0.54±0.39 (2)	0.30±0.23 (3)	0.71±0.27 (7)	49	3
U-Net + CR-3-I	0.80±0.34 (6)	0.77±0.34 (3)	0.89±0.29 (1)	0.27±0.33 (12)	0.34±0.23 (12)	0.46±0.48 (12)	0.47±0.24 (8)	0.29±0.17 (5)	0.44±0.26 (11)	70	10
U-Net + CR-123-I	0.80±0.31 (4)	0.72±0.38 (10)	0.88±0.32 (2)	0.41±0.37 (9)	0.40±0.29 (9)	0.65±0.46 (9)	0.51±0.28 (4)	0.31±0.21 (1)	0.55±0.28 (9)	57	6
BrU-Net + CR-3-I	0.77±0.38 (11)	0.75±0.38 (7)	0.88±0.32 (2)	0.10±0.25 (13)	0.30±0.22 (13)	0.20±0.39 (13)	0.42±0.23 (13)	0.27±0.17 (8)	0.32±0.23 (13)	93	13
BrU-Net + CR-123-I	0.72±0.39 (13)	0.66±0.40 (13)	0.69±0.44 (7)	0.58±0.38 (8)	0.43±0.38 (7)	0.70±0.38 (8)	0.49±0.37 (5)	0.27±0.23 (10)	0.62±0.30 (8)	79	12
Evaluation on cohort C2											
Patients with cancer N = 147, number of cancerous lesions = 189											
Model	ROC AUC	PR AUC	Sensitivity	Specificity	Precision	NPV	F1 Score	Dice	Accuracy	Sum of Ranks	Final Rank
SPCNet	0.75±0.36 (6)	0.76±0.35 (6)	0.50±0.47 (7)	0.77±0.36 (3)	0.42±0.43(4)	0.69±0.33 (3)	0.43±0.42 (6)	0.25±0.25 (6)	0.69±0.24 (3)	44	5
U-Net	0.78±0.33 (4)	0.77±0.34 (4)	0.76±0.39 (3)	0.57±0.39 (7)	0.50±0.35 (1)	0.66±0.41 (5)	0.56±0.33 (1)	0.33±0.23 (2)	0.63±0.26 (5)	32	3
BrU-Net	0.79±0.35 (3)	0.79±0.34 (3)	0.89±0.29 (2)	0.27±0.36 (8)	0.41±0.26 (5)	0.41±0.47 (8)	0.53±0.25 (2)	0.33±0.20 (1)	0.48±0.26 (8)	40	4
CorrSigNIA	0.81±0.31 (1)	0.79±0.33 (2)	0.58±0.46 (4)	0.76±0.36 (4)	0.49±0.43 (2)	0.71±0.33 (2)	0.49±0.41 (4)	0.30±0.26 (4)	0.71±0.25 (2)	25	1
SPCNet + CR-123-I	0.77±0.33 (5)	0.76±0.34 (5)	0.55±0.47 (5)	0.81±0.33 (1)	0.47±0.44 (3)	0.75±0.29 (1)	0.48±0.42 (5)	0.30±0.27 (5)	0.74±0.22 (1)	31	2
U-Net + CR-3-I	0.49±0.43 (9)	0.54±0.39 (9)	0.40±0.46 (9)	0.59±0.40 (6)	0.30±0.38 (9)	0.56±0.36 (7)	0.31±0.37 (9)	0.19±0.22 (9)	0.53±0.29 (7)	74	9
U-Net + CR-123-I	0.67±0.38 (7)	0.66±0.38 (7)	0.52±0.47 (6)	0.67±0.39 (5)	0.41±0.41 (6)	0.65±0.35 (6)	0.42±0.39 (7)	0.24±0.24 (7)	0.62±0.28 (6)	57	7
BrU-Net + CR-3-I	0.81±0.33 (2)	0.81±0.33(1)	0.90±0.27(1)	0.12±0.26 (9)	0.36±0.20 (8)	0.22±0.40 (9)	0.49±0.20 (3)	0.32±0.19 (3)	0.39±0.21 (9)	45	6
BrU-Net + CR-123-I	0.61±0.40 (8)	0.62±0.39 (8)	0.41±0.46 (8)	0.78±0.34 (2)	0.37±0.44 (7)	0.68±0.30 (4)	0.36±0.41 (8)	0.21±0.25 (8)	0.66±0.25 (4)	57	7

Table S2: Detecting clinically significant cancer: Lesion-level evaluation on cohorts C1-test and cohort C2. Mean and standard deviation values for each metric in each cohort is reported. Numbers in brackets indicate model rank in individual metric. Column “Sum of ranks” represents the sum of the individual metric ranks. Column “Final rank” represents rank of model based on sum of individual ranks. Note, the lower the sum rank, the higher ranked (i.e. better) the model. The first three rows above the dotted lines in both cohorts represent the MRI-only models, while the rows below the dotted lines represent the MRI+correlated features-based models. CorrSigNIA ranks second in cohort C1-test, closely following another CorrSigNIA variant, whereas it ranks first in cohort C2.

Evaluation on cohort C1-test												
Patients with clinically significant cancer with lesion volume > 250 mm ³ N = 34, number of clinically significant cancerous lesions = 37												
Model	ROC AUC	PR AUC	Sensitivity	Specificity	Precision	NPV	F1 Score	Dice	Accuracy	Agg cancer overlap	Sum of Ranks	Final Rank
SPCNet	0.83±0.30 (3)	0.77±0.37 (3)	0.68±0.47 (10)	0.76±0.29 (4)	0.47±0.41 (4)	0.88±0.21 (3)	0.52±0.41(4)	0.29±0.25 (8)	0.75±0.24 (5)	0.18±0.28 (9)	53	4
U-Net	0.84±0.32 (1)	0.81±0.34 (1)	0.88±0.34 (4)	0.33±0.31 (10)	0.35±0.28 (10)	0.65±0.46 (10)	0.46±0.27 (10)	0.30±0.20 (5)	0.46±0.28 (11)	0.30±0.33 (5)	67	7
Branched U-Net	0.81±0.32 (8)	0.75±0.38 (9)	0.84±0.36 (5)	0.39±0.37 (9)	0.34±0.28 (11)	0.62±0.46 (11)	0.45±0.28 (11)	0.27±0.18 (11)	0.51±0.27 (10)	0.20±0.26 (7)	92	12
CorrSigNIA	0.82±0.31 (5)	0.74±0.38 (10)	0.79±0.40 (6)	0.72±0.35 (6)	0.55±0.40 (1)	0.88±0.25 (1)	0.59±0.39 (1)	0.32±0.24 (1)	0.73±0.27 (6)	0.26±0.26 (8)	48	2
SPCNet + CR-3-S	0.79±0.31 (11)	0.70±0.39 (12)	0.66±0.47 (11)	0.79±0.22 (3)	0.45±0.39 (6)	0.91±0.13 (1)	0.51±0.40 (5)	0.28±0.23 (10)	0.77±0.18 (1)	0.14±0.26 (13)	73	8
SPCNet + CR-123-I	0.83±0.32 (4)	0.76±0.37 (5)	0.71±0.46 (9)	0.76±0.29 (4)	0.50±0.41 (2)	0.89±0.21 (2)	0.55±0.41 (3)	0.31±0.23 (3)	0.76±0.23 (3)	0.18±0.28 (9)	44	1
SPCNet + CR-123-S	0.77±0.34 (12)	0.73±0.39 (11)	0.60±0.48 (13)	0.81±0.25 (1)	0.44±0.41 (7)	0.86±0.20 (5)	0.48±0.42 (9)	0.29±0.24 (7)	0.77±0.21 (2)	0.17±0.27 (10)	77	9
SPCNet + CR-4-I	0.82±0.33 (7)	0.78±0.36 (2)	0.63±0.47 (12)	0.80±0.30 (2)	0.47±0.43 (5)	0.85±0.25 (6)	0.51±0.43 (6)	0.29±0.23 (6)	0.76±0.25 (4)	0.16±0.26 (12)	62	6
SPCNet + CR-4-S	0.81±0.32 (8)	0.76±0.37 (5)	0.72±0.44 (7)	0.72±0.32 (5)	0.50±0.41 (2)	0.85±0.29 (7)	0.55±0.39 (2)	0.31±0.23 (3)	0.72±0.27 (7)	0.21±0.29 (6)	52	3
U-Net + CR-3-I	0.82±0.32 (6)	0.75±0.36 (7)	0.93±0.25 (1)	0.24±0.28 (11)	0.31±0.19 (12)	0.48±0.49 (12)	0.44±0.20 (12)	0.28±0.18 (9)	0.41±0.23 (12)	0.36±0.29 (4)	86	10
U-Net + CR-123-I	0.80±0.29 (2)	0.75±0.37 (8)	0.90±0.29 (3)	0.41±0.37 (8)	0.39±0.29 (9)	0.66±0.46 (9)	0.50±0.27 (7)	0.31±0.21 (2)	0.53±0.29 (9)	0.46±0.34 (2)	59	5
BrU-Net + CR-3-I	0.80±0.36 (10)	0.77±0.37 (3)	0.91±0.28 (2)	0.09±0.20 (12)	0.26±0.17 (13)	0.22±0.41 (13)	0.39±0.19 (13)	0.26±0.16 (13)	0.29±0.18 (13)	0.53±0.30 (1)	93	13
BrU-Net + CR-123-I	0.73±0.41 (13)	0.70±0.40 (13)	0.72±0.44 (7)	0.56±0.38 (7)	0.41±0.37 (8)	0.20±0.39 (8)	0.49±0.37 (8)	0.27±0.23 (12)	0.60±0.31 (8)	0.42±0.36 (3)	87	11
Evaluation on cohort C2												
Patients with clinically significant cancer N = 90, number of clinically significant cancerous lesions = 110												
Model	ROC AUC	PR AUC	Sensitivity	Specificity	Precision	NPV	F1 Score	Dice	Accuracy	Agg cancer overlap	Sum of Ranks	Final Rank
SPCNet	0.79±0.34 (6)	0.79±0.34 (6)	0.58±0.47 (7)	0.74±0.37 (3)	0.48±0.42 (4)	0.71±0.34 (4)	0.49±0.41(7)	0.29±0.25 (6)	0.70±0.25 (4)	0.14±0.24 (9)	56	7
U-Net	0.83±0.30 (4)	0.81±0.32 (4)	0.81±0.37 (3)	0.58±0.39 (7)	0.53±0.34 (3)	0.68±0.41 (6)	0.60±0.32 (1)	0.38±0.23 (1)	0.67±0.26 (5)	0.26±0.28 (2)	36	3
Branched U-Net	0.84±0.31 (3)	0.82±0.33 (3)	0.94±0.22 (1)	0.26±0.35 (8)	0.42±0.25 (7)	0.41±0.48 (8)	0.55±0.23 (4)	0.35±0.19 (3)	0.48±0.25 (8)	0.24±0.24 (4)	49	4
CorrSigNIA	0.86±0.26 (1)	0.83±0.31 (2)	0.67±0.43 (4)	0.74±0.38 (4)	0.55±0.41 (1)	0.72±0.36 (3)	0.57±0.39 (2)	0.35±0.25 (5)	0.71±0.25 (2)	0.24±0.31 (5)	29	1
SPCNet + CR-123-I	0.81±0.30 (5)	0.79±0.33 (5)	0.65±0.45 (5)	0.80±0.34 (2)	0.54±0.42 (2)	0.77±0.31 (1)	0.56±0.41 (3)	0.36±0.21 (2)	0.76±0.21 (1)	0.21±0.29 (7)	33	2
U-Net + CR-3-I	0.55±0.44 (9)	0.59±0.40 (9)	0.46±0.47 (9)	0.62±0.40 (6)	0.34±0.39 (9)	0.61±0.36 (7)	0.36±0.39 (9)	0.22±0.24 (9)	0.58±0.29 (7)	0.15±0.34 (8)	82	9
U-Net + CR-123-I	0.72±0.37 (7)	0.70±0.37 (7)	0.61±0.46 (6)	0.70±0.37 (5)	0.48±0.42 (4)	0.70±0.35 (5)	0.50±0.41 (5)	0.28±0.25 (7)	0.67±0.28 (6)	0.26±0.31 (3)	55	6
BrU-Net + CR-3-I	0.86±0.31 (2)	0.85±0.30 (1)	0.93±0.24 (2)	0.12±0.23 (9)	0.35±0.17 (8)	0.25±0.42 (9)	0.49±0.18 (6)	0.35±0.19 (3)	0.39±0.19 (9)	0.62±0.30 (1)	50	5
BrU-Net + CR-123-I	0.67±0.39 (8)	0.67±0.38 (8)	0.49±0.47 (8)	0.80±0.33 (1)	0.44±0.45 (6)	0.73±0.30 (2)	0.44±0.43 (8)	0.25±0.26 (8)	0.71±0.25 (3)	0.23±0.30 (6)	58	8

Table S3: Patient-level evaluation on cohorts C1-test and C3-test (Total number of subjects N = 70, including 55 men with cancer from C1-test and 15 men with normal MRI from C3-test. Numbers in brackets represent model rank in particular metric. Column “Sum of ranks” represents the sum of the individual metric ranks. Column “Final rank” represents rank of model based on sum of individual ranks. Note, the lower the sum rank, the higher ranked (i.e. better) the model. The first three rows above the dotted lines in both cohorts represent the MRI-only models, while the rows below the dotted lines represent the MRI+correlated features-based models. CorrSigNIA outperforms all other models in patient-level evaluation.

Model	Sensitivity	Specificity	Precision	NPV	F1 score	Accuracy	Sum of Ranks	Final Rank
SPCNet	0.71 (10)	0.80 (3)	0.93 (3)	0.43 (3)	0.80 (9)	0.73 (4)	32	3
U-Net	0.87 (4)	0.00 (12)	0.76 (13)	0.00 (12)	0.81 (8)	0.69 (11)	60	13
Branched U-Net	0.87 (4)	0.13 (9)	0.79 (9)	0.22 (9)	0.83 (5)	0.71 (8)	44	9
CorrSigNIA	0.78 (6)	0.87 (1)	0.96 (1)	0.52 (1)	0.86 (2)	0.80 (1)	12	1
SPCNet + CR-3-S	0.73 (9)	0.53 (7)	0.85 (8)	0.35 (8)	0.78 (12)	0.69 (11)	55	12
SPCNet + CR-123-I	0.71 (10)	0.80 (3)	0.93 (3)	0.43 (3)	0.80 (9)	0.73 (4)	32	3
SPCNet + CR-123-S	0.67 (11)	0.87 (1)	0.95 (2)	0.42 (5)	0.79 (11)	0.71 (8)	38	7
SPCNet + CR-4-I	0.67 (11)	0.80 (3)	0.93 (3)	0.40 (6)	0.78 (12)	0.70 (13)	48	10
SPCNet + CR-4-S	0.78 (6)	0.73 (6)	0.91 (6)	0.48 (2)	0.84 (3)	0.77 (3)	26	2
U-Net + CR-3-I	0.91 (1)	0.07 (10)	0.78 (12)	0.17 (10)	0.84 (3)	0.73 (4)	40	8
U-Net + CR-123-I	0.89 (3)	0.07 (10)	0.78 (11)	0.14 (11)	0.83 (5)	0.71 (8)	48	10
BrU-Net + CR-3-I	0.91 (1)	0.00 (12)	0.77 (9)	0.00 (12)	0.83 (1)	0.71 (2)	37	5
BrU-Net + CR-123-I	0.78 (6)	0.53 (7)	0.86 (7)	0.40 (6)	0.82 (7)	0.73 (4)	37	5

1.2 Qualitative analysis of false negative and false positive predictions

To assess the clinical utility of CorrSigNIA, false negative and false positive predictions were qualitatively analysed in cohorts C1-test and C3-test.

Fig. S1 shows slices from apex to base of a patient in cohort C1-test, where CorrSigNIA missed the clinically significant lesion of volume 548 mm^3 on the left of the image (right of the patient). The false negative prediction could be due to the small size of the lesion and the subtle appearance of the lesion on MRI. It may however be noted that CorrSigNIA correctly identifies the false positive annotation of the radiologist (red) on the right of the image (left of patient) as benign tissue.

Fig. S2 shows slices from apex to base of a man with normal prostate MRI in cohort C3-test. CorrSigNIA generated a false positive prediction (predominantly indolent cancer) for this patient on the right anterior side of the prostate (left side of image).

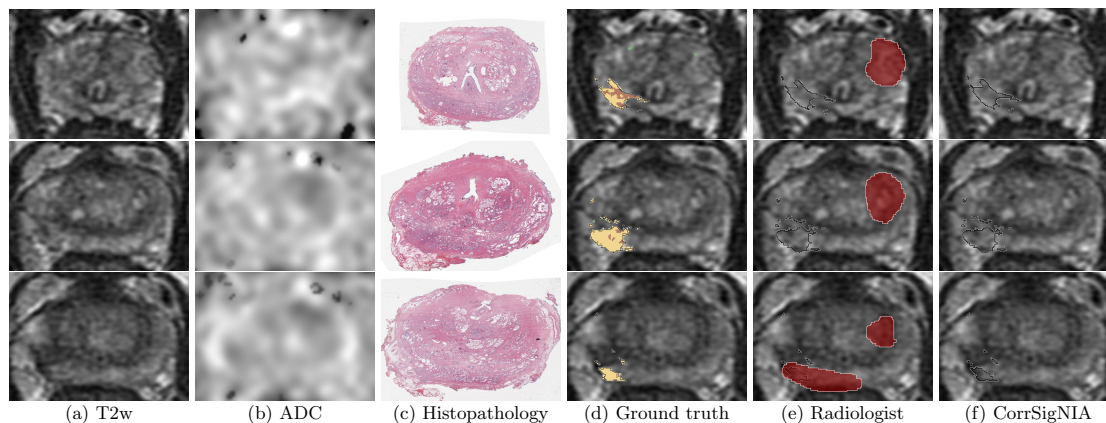


Fig. S1: CorrSigNIA false negative predictions in a patient from cohort C1-test, shown from apex to base. Registered (a) T2w image, (b) ADC image, and (c) histopathology image. T2w image overlaid with (d) ground truth labels: cancer from expert pathologist (black outline), aggressive cancer (yellow) and indolent cancer (green) histologic grade labels, pixels within pathologist outline without automated histologic grade labels (brown), (e) radiologist annotations (red), together with pathologist cancer annotations (black outline), (f) CorrSigNIA predictions, together with pathologist cancer annotations (black outline). CorrSigNIA misses the cancerous lesion on the left of the image (lesion volume 548 mm^3), but it also correctly identifies the false positive annotation of the radiologist as benign tissue.

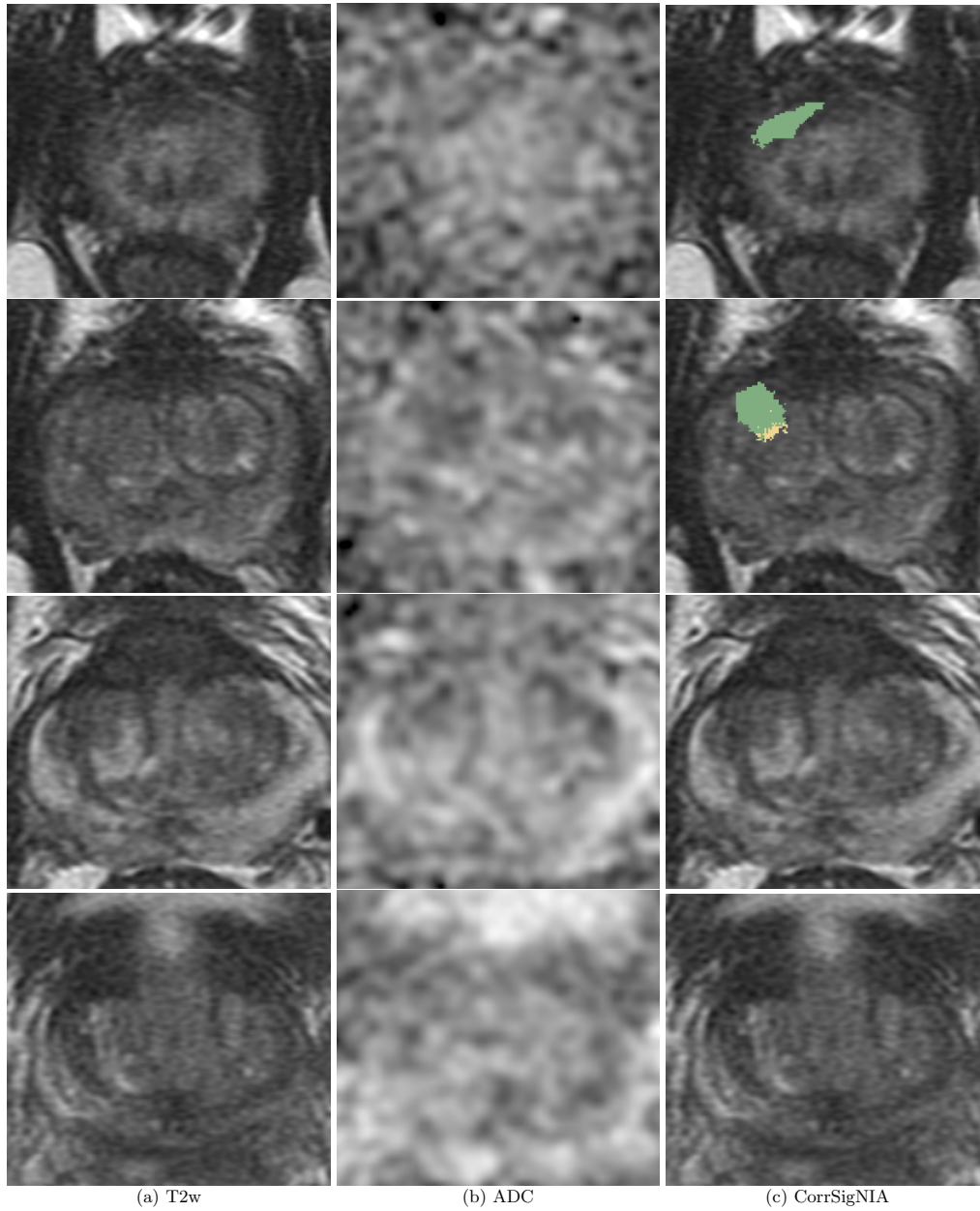


Fig. S2: CorrSigNIA false positive predictions in a man with normal prostate MRI from cohort C3-test, shown from apex to base. Registered (a) T2w image, (b) ADC image, (c) T2w image overlaid with CorrSigNIA predictions. CorrSigNIA generated a false positive prediction (predominantly indolent cancer) in the right anterior prostate (left side of image).