Appendix 1 The detailed description of the model architecture

The network architecture, including the encoder path and the decoder path, was automatically determined by the nnU-Net framework via the data properties. In our case, for the T2-weighted images (T2WI), U-Net architecture consisted of 5 downsampling blocks and 5 upsampling blocks. In contrast, for the diffusion-weighted images (DWI) with a lower resolution, the model consisted of 3 downsampling blocks and 3 upsampling blocks. The detailed information is elaborated upon in *Table S2*.

Appendix 2 The detailed fine-tuning processes

The original training and fine-tuning cohorts were combined to train another 50 epochs, with a smaller initial learning rate of 0.0001. The parameters of the entire network were all updated. Similar to the original training procedure, the batch size of 2 was calculated with the nnU-Net framework based on the graphic processing unit (GPU) memory and the number of parameters of the model. The loss function was the combination of the Dice loss and the binary cross entropy (BCE) loss.

Appendix 3 Statistical analysis of the influencing factors for the Dice similarity coefficient on DWI images in the PXtest cohort without fine-tuning

A multivariate logistic regression analysis was performed to analyze the potential influencing factors in the decline of the DWI model's performance in the external testing cohort. The candidate influencing factors were echo time, repetition time, MR scanner, echo train length, and slice thickness.
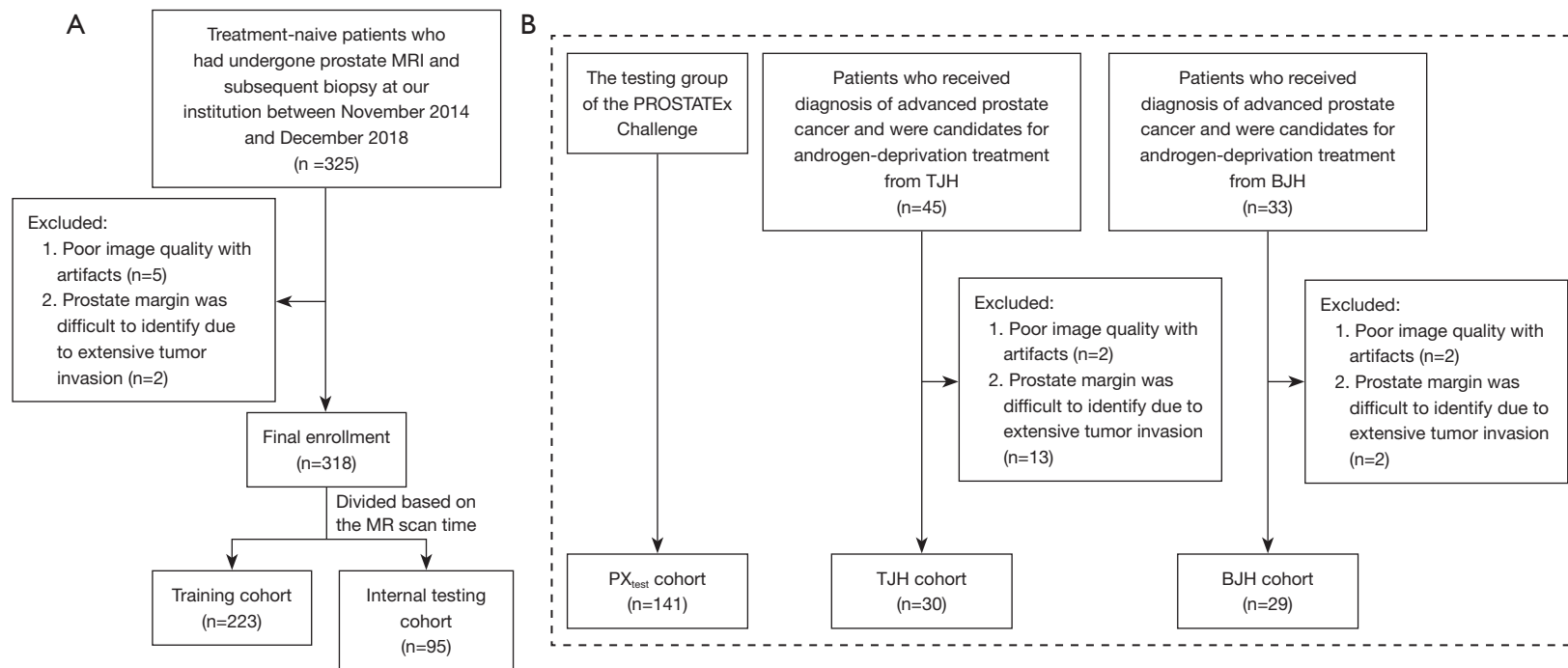
Figure S1 Flowchart of the selection process of the datasets in this study. PX$_{test}$ cohort, testing group of the PROSTATEx Challenge dataset; TJH cohort, Tongji Hospital cohort; BJH cohort, Beijing Hospital cohort.

Table S1 MRI acquisition parameters for the axial T2-weighted imaging and diffusion-weighted imaging sequences [median (range)]

Sequences	Datasets		Cases	Scanner	MR field strength (T)	TR (msec)	TE (msec)	Matrix	Slice thickness (mm)	Pixel spacing
T2WI	Training cohort	Our institution	223	GE Discovery MR 750	3.0	4,422 (2,672–5,367)	108 (97–116)	512×512	3 (3–6)	0.51 (0.43–0.78)
	Internal testing cohort	Our institution	95	GE Discovery MR 750	3.0	4,424 (2,672–5,534)	108 (86–116)	512×512	3 (3–4)	0.51 (0.51–0.53)
	External testing cohort 1	PX$_{test}$ cohort	90	Siemens Magnetom Skyra	3.0	5,660 (5,660–8,624)	104 (101–104)	384×384–640×640	3 (3–3.5)	0.5 (0.3–0.5)
			51	Siemens Magnetom TrioTim	3.0	4,494 (4,480–5,870)	103 (102–104)	256×256–320×320	3 (3–5)	0.56 (0.56–0.70)
	External testing cohort 2	TJH cohort	29	Siemens Magnetom Skyra	3.0	6,750 (6,130–7,970)	104	384×384	3 (3–3.5)	0.47
			1	Siemens Magnetom Aera	1.5	4,920	90	640×640	4	0.38
	External testing cohort 3	BJH cohort	10	Siemens Magnetom Espree	1.5	4,650 (3,800–5,040)	115 (102–118)	256×256–320×320	4 (3–4)	0.75 (0.75–0.78)
			3	GE Optima MR360	1.5	3,682 (3,660–3,682)	1101 (110–111)	512×512	4	0.47 (0.47–0.59)
			1	GE Signa EXCITE	1.5	4,120	122	512×512	5	0.625
			9	Philips Achieva	3.0	4,265 (2,500–5,183)	100 (80–100)	480×480–672×672	4	0.42 (0.30–0.5)
			5	GE Discovery MR 750	3.0	4,525 (4,103–4,775)	87 (87–89)	512×512	4	0.47
			1	GE SIGNA Pioneer	3.0	4,952	86	512×512	4	0.47
DWI	Training cohort	Our institution	223	GE Discovery MR 750	3.0	2,800 (2,000–4,000)	65 (62–70)	256×256	3 (3–6)	1.41 (0.94–1.56)
	Internal testing cohort	Our institution	95	GE Discovery MR 750	3.0	2,800	70 (69–70)	256×256	3 (3–4)	1.41
	Fine-tuning cohort	PX$_{train}$ cohort	203	Siemens Magnetom Skyra, and TrioTim	3.0	2,700 (2,500–3,300)	63 (63–81)	128×84–128×120	3 (3–4.5)	2
	External testing cohort	PX$_{test}$ cohort	90	Siemens Magnetom Skyra	3.0	2,700 (2,700–3,200)	63	128×84–128×120	3 (3–5)	2
			51	Siemens Magnetom TrioTim	3.0	2,800 (2,500–3,224)	70 (64–81)	106×128–128×88	3 (3–4)	2

T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; TR, repetition time; TE, echo time; PX$_{test}$ cohort, testing group of the PROSTATEx Challenge dataset; TJH cohort, Tongji Hospital cohort; BJH cohort, Beijing Hospital cohort; PXtrain cohort, training group of the PROSTATEx Challenge dataset.

Table S2 The architecture of the U-Net models for T2-weighted imaging and diffusion-weighted imaging

Block type	Model for T2WI		Model for DWI	
	Conv kernel	Pooling	Conv kernel	Pooling
Downsample 1	[1, 3, 3] ×2	[1, 2, 2]	[1, 3, 3] ×2	[2, 1, 1]
Downsample 2	[1, 3, 3] ×2	[1, 2, 2]	[1, 3, 3] ×2	[2, 1, 1]
Downsample 3	[3, 3, 3] ×2	[2, 2, 2]	[3, 3, 3] ×2	[1, 2, 2]
Downsample 4	[3, 3, 3] ×2	[1, 2, 2]	-	-
Downsample 5	[3, 3, 3] ×2	[1, 2, 2]	-	-
Bridge	[3, 3, 3] ×2	-	[3, 3, 3] ×2	-
Upsample 1	[3, 3, 3] ×2	[1, 2, 2]	-	-
Upsample 2	[3, 3, 3] ×2	[1, 2, 2]	-	-
Upsample 3	[3, 3, 3] ×2	[2, 2, 2]	[3, 3, 3] ×2	[1, 2, 2]
Upsample 4	[1, 3, 3] ×2	[1, 2, 2]	[1, 3, 3] ×2	[2, 1, 1]
Upsample 5	[1, 3, 3] ×2	[1, 2, 2]	[1, 3, 3] ×2	[2, 1, 1]
Output	[1, 1, 1]	-	[1, 1, 1]	-

T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging.

Table S3 Multivariate regression analyses of factors affecting the Dice similarity coefficient on DWI images in the PX_{test} cohort without fine-tuning

Factors	Coefficients	Lower (2.5%)	Upper (97.5%)	P value
Echo time	0.004	−0.021	0.029	0.7633
Repetition time	0.001	0.001	0.002	<0.001
MR scanner	−0.729	−1.323	−0.136	0.016
Echo train length	−0.028	−0.041	−0.015	<0.001
Slice thickness	−0.273	−0.405	−0.140	<0.001

DWI, diffusion-weighted imaging; PX_{test} cohort, testing group of the PROSTATEx Challenge dataset.