Anatomy-Aware Reverse Symmetric Self-Paced Learning for Prostate Cancer Detection

Dan Zhang[1], Di Zhao[2], Yu Sun[3], Alan Wang[1], Claude Aguergaray[4], and Hayley M. Reynolds[1⋆]

[1] Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand
[2] School of Computer Science, University of Auckland, Auckland, New Zealand
[3] School of Physics, University of Sydney, Sydney, Australia
[4] Department of Physics, University of Auckland, Auckland, New Zealand
hayley.reynolds@auckland.ac.nz

Abstract. Despite the success of deep learning models to detect clinically significant prostate cancer (csPCa) using multiparametric MRI (mpMRI), it faces two levels of data imbalance: (1) voxel-level class imbalance within each patient case, where lesion voxels occupy a small fraction of the image volume, and (2) case-level difficulty imbalance, where cases exhibit heterogeneity in detection difficulty. While existing methods primarily address voxel-level imbalance via loss reweighting, they often overlook the uneven distribution of case difficulty, limiting their performance. To bridge this gap, we propose Reverse Symmetric Self-Paced Learning (RSSPL), a model-agnostic case-level learning strategy. Specifically, RSSPL builds upon curriculum learning but prioritises patient cases in a hard-to-easy manner to better emphasise challenging cases earlier during training, and can be used alongside voxel-level reweighting. To facilitate this, an anatomy-aware difficulty estimate is proposed to refine case scheduling by filtering the noise outside the prostate gland. Experiments on csPCa detection across different network backbones and both standard and imbalance-aware loss functions demonstrate that RSSPL leads to reduced cross-validation variance and improved detection performance. On the independent Prostate158 test set, RSSPL achieves a statistically significant improvement at the patient level, with an absolute AUROC increase of 3.05% ($p < 0.05$). Consistent improvement trends in a brain lesion-detection task further provide preliminary evidence of RSSPL's cross-organ transferability.

Keywords: Prostate cancer detection · Multiparametric MRI · Curriculum learning · Symmetric self-paced learning.

1 Introduction

Prostate cancer is the second most frequently diagnosed cancer among men worldwide [1]. Deep Learning (DL) methods for detecting clinically significant

⋆ Corresponding author

prostate cancer (csPCa) using multiparametric MRI (mpMRI) have achieved non-inferior or even superior performance to radiologists [3]. Typical MRI sequences include T2-weighted (T2W), Apparent Diffusion Coefficient (ADC) and high-b-value Diffusion Weighted Imaging (DWI) sequences and sometimes dynamic contrast enhanced MRI (DCE MRI). Conventionally, a patient is classified as having csPCa if at least one lesion is histopathologically confirmed as International Society of Urological Pathology (ISUP) grade group ≥ 2 [2, 3]. Consequently, patients with ISUP grade group 1 or benign findings are categorised as having non-csPCa. Despite the success of DL methods, automated csPCa detection using mpMRI still faces data imbalance both within and across cases. At the voxel level, lesions typically occupy only a small fraction of the image volume, resulting in severe foreground–background imbalance within each case. At the case level, detection difficulty varies substantially across patients, resulting in difficulty heterogeneity even when assessing mpMRI within the same csPCa or non-csPCa category.

Prior work has primarily focused on mitigating voxel-level imbalance within each case through loss reweighting strategies such as weighted cross-entropy [4], focal loss [5], and their variants [6–9] to emphasise hard voxels or rare foreground regions. However, these methods do not explicitly regulate how entire cases are prioritised during training, instead following a conventional random scheduling strategy. In large screening-oriented prostate MRI datasets such as PI-CAI [3], the predominance of non-csPCa cases, coupled with the substantial variation in lesion appearance among csPCa patients (ranging from large, conspicuous lesions to subtle, small ones), creates a highly polarised distribution of case difficulty. Under such conditions, these methods that treat all cases equally may lead to suboptimal optimisation.

This suggests that the order and priority with which cases are presented during training may substantially influence model learning. Curriculum Learning (CL) [10] enables the model to progressively transition from easy to hard cases during training. However, in csPCa detection using mpMRI, datasets are often dominated by non-csPCa cases and relatively easier csPCa cases. Prioritising such cases early in training may delay exposure to more difficult csPCa cases. This motivates a fundamental question: under highly polarised case distributions, is the conventional easy-to-hard schedule still optimal?

To this end, we propose the Reverse Symmetric Self-Paced Learning (RSSPL) strategy to present cases to the model in a hard-to-easy manner during training, ensuring that rare but informative harder cases are not overwhelmed by the predominance of easier cases. To implement the proposed scheduling strategy, a quantitative measure of case difficulty is required. Existing approaches typically derive this difficulty from model feedback, such as input gradients [11]. However, in medical imaging, if input gradients are computed over the entire image volume, these signals can be unstable or misleading when dominated by large background regions, and they rarely incorporate anatomical knowledge that is clinically relevant. To address this, we estimate case difficulty using input gradients restricted to the prostate gland, providing a more robust and clinically meaningful sig-

nal. An overview of the proposed RSSPL is shown in Fig. 1. Through sufficient experiments on csPCa detection across different network backbones and loss functions, and additional validation of RSSPL on a brain lesion detection task, we demonstrate that hard-to-easy scheduling leads to more stable optimisation under highly polarised distributions of case difficulty. The main contributions of this paper are:

- We propose RSSPL, a novel framework designed to address polarised case difficulty in csPCa detection. It is model-agnostic and compatible with voxel-level reweighting, allowing for seamless integration with different architectures and loss functions to address data imbalance across multiple levels.
- We propose an anatomy-aware difficulty estimation method that integrates anatomical priors into the learning process, offering a novel way to utilise clinical knowledge for case-level scheduling.
- We demonstrate consistent performance gains and reduced cross-validation variance in prostate cancer detection, supporting its clinical viability. Additional brain lesion experiments provide preliminary evidence of cross-organ transferability to broader tasks characterised by polarised case difficulty.

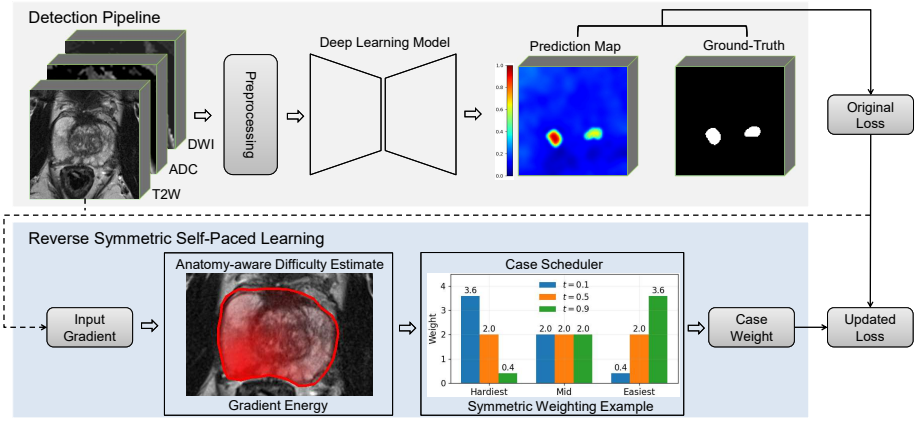

Fig. 1. Overview of the reverse symmetric self-paced learning framework for prostate cancer detection. Case difficulty is estimated in an anatomy-aware manner and used to schedule training cases via hard-to-easy strategy.

2 Method

2.1 Anatomy-aware Difficulty Estimation

We estimate case difficulty via input gradients, following recent evidence that this approach provides more informative model feedback than loss-based methods [12–14] by explicitly taking input features into account [11]. However, in

prostate MRI, standard gradients calculated over the entire image often contain noise from large background regions. To obtain a clinically relevant signal, we constrain the gradient computation to the prostate gland using a binary mask. Let f_θ be the network, x the input volume with its label y, and $\ell(\cdot)$ the training loss. We define the input gradient as $g = \nabla_x \ell(f_\theta(x), y)$. Given a binary gland mask M derived from the anatomical annotations provided in the dataset, the difficulty score $d(x)$ is computed as:

$$d(x) = \frac{\|g \odot M\|_2^2}{|M| + \epsilon} \tag{1}$$

where $\odot$ is the elementwise product, $|M|$ denotes the number of foreground voxels, and ϵ is a small constant for numerical stability. Dividing by $|M|$ ensures that the difficulty score is normalised by the gland size, preventing larger anatomical structures from biasing the estimation.

2.2 Reverse Symmetric Self-paced Learning

Symmetric Self-Paced Learning (SSPL) [11] is a specialised Curriculum Learning (CL) framework that regulates model attention through a dynamic weighting scheme. It ensures a balanced transition by initially assigning larger weights to easier cases and lower weights to more challenging ones; as training progresses, the weights of easy cases gradually decrease while those of hard cases increase. However, this easy-to-hard schema may be sub-optimal for csPCa detection where case difficulty distributions are highly polarised. Building upon this framework, we propose *Reverse Symmetric Self-Paced Learning* (RSSPL). We extend the dynamic weighting scheme by incorporating the clinically-relevant difficulty estimation introduced in Sec. 2.1, whilst inverting the learning order to follow a hard-to-easy schema, shifting the model attention from harder cases towards easier ones as training progresses.

Rank-based Scheduling Framework To rigorously evaluate the impact of learning orders, we formulate a rank-based dynamic weighting framework. Let $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^N$ be the training set and $\ell_i(\theta)$ the loss for case i. The weighted objective is:

$$\min_\theta \sum_{i=1}^{N} w_i(t;\sigma)\, \ell_i(\theta), \tag{2}$$

where $w_i(t;\sigma)$ represents the case weight at normalised training progress $t \in [0,1]$, defined as the ratio between the current epoch and the total number of epochs. As illustrated in the case scheduler module within Fig. 1, We achieve the attention transition by defining the weight through a linear interpolation between two moving endpoints:

$$w_i(t;\sigma) = w_{\text{first}}(t) - \frac{r_i(\sigma) - 1}{N - 1}\big(w_{\text{first}}(t) - w_{\text{last}}(t)\big), \tag{3}$$

where the rank $r_i(\sigma)$ is the sorted index of case i based on its difficulty score $d(x_i)$ (Eq. (1)). And the direction parameter $\sigma \in \{+1, -1\}$ determines the sorting order, with $+1$ and -1 denoting ascending and descending ranks, respectively. The moving endpoints, $w_{\text{first}}(t)$ and $w_{\text{last}}(t)$, represent the weights assigned to the top-ranked ($r_i = 1$) and bottom-ranked ($r_i = N$) cases, respectively. These endpoints are formulated as:

$$w_{\text{first}}(t) = \texttt{min} + \texttt{span} \cdot (1 - t), \quad w_{\text{last}}(t) = \texttt{min} + \texttt{span} \cdot t. \tag{4}$$

Here, `min` and `span` modulate the weight range, facilitating a symmetric swap of attention between harder and easier cases as t increases.

The Proposed RSSPL Strategy In our RSSPL, we set $\sigma = -1$, yielding a descending rank where the most challenging cases (highest $d(x_i)$) are ranked first. This ensures that rare but difficult cases have larger initial weights and prioritised via Eq. 3 at the high-gradient phase during early training ($t \to 0$). As training progresses, the symmetric swap in Eq. 4 gradually shifts the model's attention towards easier cases, potentially facilitating overall model convergence.

Relation to SSPL This rank-based formulation allows RSSPL to be directly compared with SSPL. Specifically, when $\sigma = +1$, it corresponds to SSPL (easy-to-hard).

3 Experiments

3.1 Datasets and Evaluation Metrics

The public PI-CAI dataset [3] consists of MRI data from 425 csPCa (ISUP ≥ 2) and 1,075 non-csPCa cases, providing a combination of radiologist-derived and AI-generated annotations for both lesions and gland zones. Prostate158 [15] is another public dataset containing 158 cases (102 csPCa cases and 56 non-csPCa cases) with radiologist-provided annotations. Both datasets contain three modalities: T2W, ADC, and high b-value DWI. Evaluation follows the PI-CAI protocol [3] using patient-level AUROC, lesion-level AP, and their average (AVG) as the primary metric. For 5-fold cross-validation, results are summarised as mean $\pm$ 2*standard deviation (SD) across the five folds, consistent with established evaluation protocols [18–20].

3.2 Implementation Details

To ensure fair and consistent comparisons, all methods were implemented within the nnUNet framework [16] following its automated self-configuring pipeline for preprocessing (including data augmentation), optimiser settings, and model selection strategies across all experiments. Model development was conducted on the public PI-CAI dataset [3] using five-fold cross-validation. To ensure a strictly

Table 1. Prostate lesion detection results. Performance is reported as point estimates for Prostate158 test set. For each backbone–loss configuration, SSPL and RSSPL are compared against the corresponding EW baseline within the same group.

Group	Strategy	Validation AVG	Test AVG (Prostate158)
UNet-CE	EW	64.15 ± 8.02	71.21
	SSPL	64.65 ± 10.25	72.36
	RSSPL	**68.28 ± 8.50 (↑4.13%)**	**71.84 (↑0.63%)**
UNet-CEF	EW	65.86 ± 6.16	70.71
	SSPL	67.09 ± 8.77	70.43
	RSSPL	**67.74 ± 6.91 (↑1.88%)**	**71.07 (↑0.36%)**
MNet-CE	EW	68.66 ± 6.56	76.55
	SSPL	69.06 ± 7.38	76.11
	RSSPL	**70.34 ± 7.16 (↑1.68%)**	**78.72 (↑2.17%)**
MNet-CEF	EW	69.07 ± 4.54	77.24
	SSPL	68.78 ± 9.77	78.38
	RSSPL	**71.07 ± 6.76 (↑2.00%)**	**79.12 (↑1.88%)**

independent evaluation, Prostate158 [15] was held out from all training and validation phases and used only for final testing. Postprocessing (e.g., lesion candidate retrieval) and evaluation were performed using the publicly available `picai_eval` toolkit [3]. Difficulty scores for case scheduling were computed once per epoch and ranked globally. Each model training was performed on an Tesla V100 32GB GPU for 500 epoches. For the hyperparameters `min` and `span`, we fixed `min` = 0 and conducted a search over `span` $\in \{1, 2, 4, 6\}$, selecting `span` = 4 based on validation performance.

3.3 Prostate Cancer Lesion Detection Results

To demonstrate that RSSPL is model-agnostic, we evaluated two backbones: the default U-Net [4] used in nnUNet and the MNet proposed by Yuan et al [17], which achieved top-ranking position in the PI-CAI Grand Challenge [3]. To verify RSSPL's compatibility with voxel-level reweighting, we evaluated two loss configurations: standard Cross-Entropy (CE) and a combination of CE and Focal Loss (CEF), an imbalance-aware loss configuration recommended by PI-CAI.

Results in Table 1 shows that SSPL exhibits divergent trends between validation and test sets; for instance, while performance improves on the validation set for MNet-CE, it simultaneously decreases on the test set. In contrast, RSSPL provides more consistent gains over the EW baseline across all configurations. Notably, RSSPL's consistently lower 5-fold cross-validation SD compared to SSPL underscores its superior stability, ensuring the reliable performance essential for clinical tasks. Moreover, RSSPL yields larger test-set gains on MNet than on UNet (e.g., $\sim$2% vs. $<$1% AVG improvement on Prostate158), suggesting that RSSPL interacts synergistically with advanced model architectures to enhance generalisation.

Table 2. Paired bootstrap comparison on Prostate158. Reported are the mean paired difference (Δ), 95% CI, and two-sided p-values; significant results in bold.

Metric	**Δ mean (%)**	**95% CI (%)**	p
Lesion-level AP	0.35	[−1.50, 2.27]	0.705
Patient-level AUROC	**3.05**	**[0.78, 5.51]**	**0.012**

Paired Bootstrap Evaluation To evaluate the statistical significance of the gains reported in Table 1, we focus on the best-performing MNet-CEF configuration. Specifically, we conducted stratified patient-level bootstrap resampling (2,000 replicates) on the Prostate158 test set to compare the effectiveness of RSSPL against its corresponding EW baseline, preserving the original proportion of csPCa and non-csPCa patients in each replicate. As shown in Table 2, RSSPL achieves consistent improvements across the two metrics. Notably, at the patient level, RSSPL achieves a substantial AUROC increase of 3.05% ($p < 0.05$). This significant gain aligns with our design motivation, as RSSPL prioritises rare yet informative cases at the patient level, ensuring they are sufficiently learned during training.

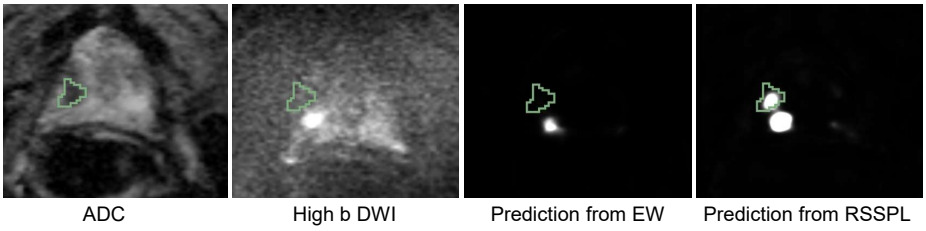

Fig. 2. Qualitative comparison on Prostate158. The baseline model (MNet-CEF-EW) fails to detect the lesion (green circle) due to ADC–DWI misalignment, while the RSSPL-trained model (MNet-CEF-RSSPL) detects it.

Qualitative Analysis We qualitatively evaluated the best-performing MNet-CEF configuration to demonstrate RSSPL's advantages over the EW baseline. As shown in Fig. 2, the baseline model is prone to missing lesions in cases with severe ADC–DWI misalignment. By prioritising challenging cases, RSSPL enhances the model's robustness in detection even in the presence of severe challenges.

Difficulty Signals Ablation Study To evaluate the impact of different difficulty estimation methods, we conducted studies using a standard UNet backbone with CE loss (Table 3). Compared with loss-based or whole-image input gradient approaches, integrating anatomical priors into the gradient calculation leads to

superior performance across all metrics. This supports the use of anatomy-aware signals for more reliable difficulty assessment.

Table 3. Difficulty signals ablation study results. Gradient-A is short for Anatomy-aware Gradient. Best results are shown in **bold**.

Difficulty Signal	AVG	Patient-level AUROC	Lesion-level AP
Loss	67.62 ± 5.70	82.99 ± 4.93	52.24 ± 6.99
Gradient	67.87 ± 7.40	82.87 ± 4.17	52.87 ± 11.24
Gradient-A (Ours)	**68.28 ± 8.50**	**83.51 ± 3.93**	**53.06 ± 13.33**

3.4 Brain Lesion Validation Study

To evaluate the potential transferability of RSSPL, we conducted validation on a polarised brain lesion dataset, which we constructed by combining 120 cases with lesions from BraTS2023 [21] and 580 healthy cases from IXI [22]. We conducted the validation using a standard UNet-CE configuration. For simplicity, the whole-image input gradient was utilised as the difficulty signal for RSSPL in this task. Results in Table 4 show that while patient-level AUROC is saturated at 100% for both strategies, RSSPL achieves a notable increase of over 2% in lesion-level AP. Although this experiment is not intended as a full benchmark comparison, RSSPL demonstrates a consistent improvement trend over the baseline training strategy. This provides preliminary evidence that the proposed RSSPL is not strictly limited to prostate imaging and may possess broader applicability to other imbalanced medical imaging tasks.

Table 4. Cross-organ validation results on the brain lesion detection task.

Configuration	AVG	Patient-level AUROC	Lesion-level AP
EW	71.61 ± 4.28	100.00 ± 0.00	43.22 ± 8.55
RSSPL	**72.97 ± 4.19**	**100.00 ± 0.00**	**45.93 ± 8.38**

4 Conclusion

This study revisits the efficacy of curriculum learning in the context of highly polarised data distributions. Our findings challenge the conventional easy-to-hard schema, demonstrating that for tasks such as csPCa detection, prioritising rare yet difficult cases early provides a more stable and effective optimisation approach. As a model-agnostic strategy, RSSPL offers a plug-and-play

solution that enhances case scheduling without adding architectural complexity. The cross-organ experiments on brain lesions provide preliminary evidence of RSSPL's transferability. For prostate imaging, future work could adapt the anatomical priors to be more tolerant of imperfect gland segmentation. More generally, formalising the theoretical foundation of hard-to-easy scheduling, expanding cross-organ validation, and evaluating it against a wider range of recent training strategies and architectures would strengthen its broader applicability.

Acknowledgments. This work was supported by the New Zealand Ministry of Business, Innovation and Employment (MBIE) Smart Ideas programme (Grant UOAX2306). We also acknowledge the computational resources provided by the New Zealand eScience Infrastructure (NeSI).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bray, F., Laversanne, M., Sung, H., et al.: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **74**(3), 229–263 (2024)
2. Molière, S., Hamzaoui, D., Ploussard, G., et al.: A systematic review of the diagnostic accuracy of deep learning models for the automatic detection, localization, and characterization of clinically significant prostate cancer on magnetic resonance imaging. European Urology Oncology (2024)
3. Saha, A., Bosma, J.S., Twilt, J.J., et al.: Artificial intelligence and radiologists in prostate cancer detection on MRI (PI-CAI): an international, paired, non-inferiority, confirmatory study. The Lancet Oncology **25**(7), 879–887 (2024)
4. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241. Springer, Munich (2015)
5. Lin, T.-Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988. IEEE, Venice (2017)
6. Yeung, M., Sala, E., Schönlieb, C.-B., et al.: Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics **95**, 102026 (2022)
7. Ghosh, A., Schaaf, T., Gormley, M.: Adafocal: Calibration-aware adaptive focal loss. Advances in Neural Information Processing Systems **35**, 1583–1595 (2022)
8. Xu, F., Yang, F., Li, X., et al.: A unified loss for handling inter-class and intra-class imbalance in medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 8842–8850. AAAI Press, Philadelphia (2025)
9. Jiang, W., Li, J., Li, Y., et al.: Boundary-aware dynamic re-weighting for semi-supervised medial image segmentation. Expert Systems with Applications, 128175 (2025)
10. Bengio, Y., Louradour, J., Collobert, R., et al.: Curriculum learning. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML), pp. 41–48. ACM, Montreal (2009)

11. Zhao, D., Koh, Y.S., Dobbie, G., et al.: Symmetric self-paced learning for domain generalization. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 16961–16969. AAAI Press, Vancouver (2024)
12. Han, B., Yao, Q., Yu, X., et al.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. Advances in Neural Information Processing Systems **31** (2018)
13. Xue, C., Yu, L., Chen, P., et al.: Robust medical image classification from noisy labeled data with global and local representation guided co-training. IEEE Transactions on Medical Imaging **41**(6), 1371–1382 (2022)
14. Shu, J., Yuan, X., Meng, D., et al.: CMW-Net: Learning a class-aware sample weighting mapping for robust deep learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(10), 11521–11539 (2023)
15. Adams, L.C., Makowski, M.R., Engel, G., et al.: Prostate158—An expert-annotated 3T MRI dataset and algorithm for prostate cancer detection. Computers in Biology and Medicine **148**, 105817 (2022)
16. Isensee, F., Jaeger, P.F., Kohl, S.A.A., et al.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
17. Yuan, Y., Ahn, E., Feng, D., et al.: Z-SSMNet: Zonal-aware self-supervised mesh network for prostate cancer detection and diagnosis with bi-parametric MRI. Computerized Medical Imaging and Graphics **122**, 102510 (2025)
18. Cao, R., Bajgiran, A.M., Mirak, S.A., et al.: Joint prostate cancer detection and Gleason score prediction in mp-MRI via FocalNet. IEEE Trans. Med. Imaging **38**(11), 2496–2506 (2019)
19. Duran, A., Dussert, G., Rouvière, O., et al.: ProstAttention-Net: A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. Med. Image Anal. **77**, 102347 (2022)
20. Saha, A., Hosseinzadeh, M., Huisman, H.: End-to-end prostate cancer detection in bpMRI via 3D CNNs: effects of attention mechanisms, clinical priori and decoupled false positive reduction. Med. Image Anal. **73**, 102155 (2021)
21. Kazerooni, A.F., Khalili, N., Liu, X., et al.: The brain tumor segmentation (BraTS) challenge 2023: Focus on pediatrics (CBTN-Connect-DIPG-ASNR-MICCAI BraTS-Peds). arXiv preprint arXiv:2305 (2024)
22. IXI Dataset, http://brain-development.org/ixi-dataset/, last accessed 2025/11/17