ProTeUS: A Spatio-Temporal Enhanced Ultrasound-Based Framework for Prostate Cancer Detection

Tarek Elghareb[1,5], Mohamed Harmanani[2,5‡], Minh Nguyen Nhat To[1,5‡], Paul Wilson[2,5], Amoon Jamzad[2,5], Fahimeh Fooladgar[1], Baraa Abdelsamad[1], Obed Dzikunu[1], Samira Sojoudi[1], Gabrielle Reznik[3], Michael Leveridge[4], Robert Siemens[4], Silvia Chang[3], Peter Black[3], Parvin Mousavi[2,5⋆], and Purang Abolmaesumi[1⋆]

[1] The University of British Columbia, Vancouver, BC, Canada
purang@ece.ubc.ca
[2] Queen's University, Kingston, ON, Canada
[3] Vancouver General Hospital, Vancouver, BC, Canada
[4] Kingston General Hospital, Kingston, ON, Canada
[5] Vector Institute, Toronto, ON, Canada

Abstract. Deep learning holds significant promise for enhancing real-time ultrasound-based prostate biopsy guidance through precise and effective tissue characterization. Despite recent advancements, prostate cancer (PCa) detection using ultrasound imaging still faces two critical challenges: (i) limited sensitivity to subtle tissue variations essential for detecting clinically significant disease, and (ii) weak and noisy labeling resulting from reliance on coarse annotations in histopathological reports. To address these issues, we introduce ProTeUS, an innovative spatio-temporal framework that integrates clinical metadata with comprehensive spatial and temporal ultrasound features extracted by a foundation model. Our method includes a novel hybrid, cancer involvement-aware loss function designed to enhance resilience against label noise and effectively learn distinct PCa signatures. Furthermore, we employ a progressive training strategy that initially prioritizes high-involvement cases and gradually incorporates lower-involvement samples. These advancements significantly improve the model's robustness to noise and mitigate the limitations posed by weak labels, achieving state-of-the-art PCa detection performance with an AUROC of 86.9%. Our code is publicly accessible at github.com/DeepRCL/ProTeUS.

Keywords: Prostate Cancer · Ultrasound Imaging · Deep Learning · Foundation Models · Progressive Training · Time-Series Analysis

⋆ P. Mousavi and P. Abolmaesumi are joint senior authors.
‡ Equal contribution.

1 Introduction

Prostate cancer (PCa) is the second most commonly diagnosed malignancy among men and the fifth leading cause of cancer-related death [15]. Early and accurate detection significantly enhances patient outcomes, yet standard diagnostic techniques, such as systematic transrectal ultrasound (TRUS)-guided biopsies, suffer from limited sensitivity (40–50%), often failing to detect clinically significant cancers or prompting unnecessary biopsies due to their coarse sampling and the inherent limitations of TRUS imaging [1,13,14]. While multi-parametric MRI (mpMRI) fused with TRUS improves detection accuracy [19], it introduces additional complexity and costs, and requires specialized expertise [2], which restricts widespread clinical adoption. Therefore, developing reliable standalone ultrasound-based methods remains clinically attractive for real-time, accurate tissue characterization.

Recent advancements in deep learning (DL) have shown promising results in improving PCa detection via various ultrasound imaging modalities [7,17,22,25], including B-mode imaging [8,17,23], radio frequency (RF) data [22], contrast-enhanced ultrasound (CEUS) [21], high frequency ultrasound [24], and shear wave elastography [18]. Temporal enhanced Ultrasound (TeUS), which analyzes time-series RF signals, has emerged as particularly promising due to its capacity to capture subtle tissue dynamics often missed in static imaging methods [3,9]. In parallel to these developments, large-scale foundation models pre-trained on diverse datasets are gaining traction for their robust performance in medical imaging applications, including PCa detection [7,25].

Despite these advances, existing approaches remain limited by inadequate robustness against subtle tissue variations and weak, noisy labels provided by histopathology reports, which commonly lack precise spatial annotations. Such weak labeling introduces noise in the training set of DL models, potentially undermining model performance [9]. In particular, cores with low cancer involvement pose a greater risk of mislabeled regions, as only a small fraction of the tissue is actually malignant. Training on these uncertain samples can destabilize the learning process, especially in the early stages, causing the model to overfit to noisy signals [10,20].

To address these significant gaps, we introduce ProTeUS, a novel learning framework to integrate spatial image features and fine-grained temporal ultrasound signatures through a foundation model, that also encodes clinical metadata to enhance PCa detection. Our key contributions include:

- A spatio-temporal learning strategy integrating both global spatial and fine-grained temporal features to enhance diagnostic robustness.
- Incorporation of clinical metadata to contextualize image features and reduce the impact of noisy labeling.
- An innovative hybrid loss function explicitly designed to enhance model resilience to noisy, weak labels.
- A progressive training methodology prioritizing clearly defined cases, incrementally adapting to challenging, ambiguous samples.

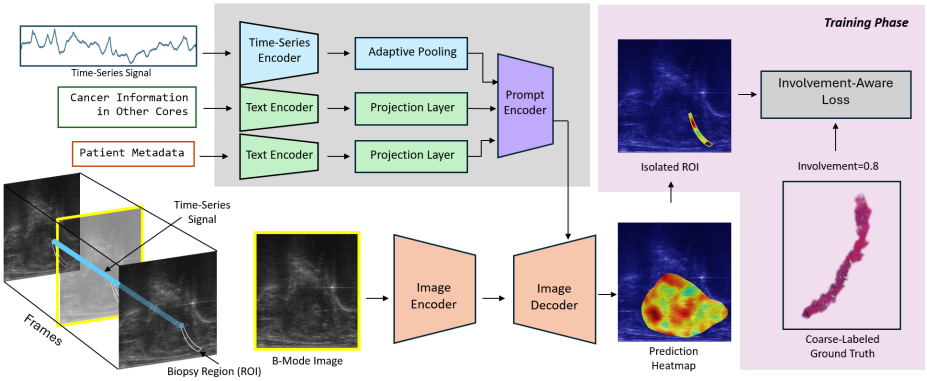

Fig. 1: Overview of the proposed pipeline, integrating time-series RF signals, patient metadata, cancer information in other cores (during training only), and B-mode features for robust prostate cancer detection.

These advances lead to significant improvements in prostate cancer detection performance, offering a robust foundation for more precise biopsy guidance and improved patient outcomes.

2 Materials and Methods

2.1 Dataset

Acquisition: A private dataset was collected as part of a clinical PCa study approved by the institutional health research ethics board. All participants provided both verbal and written informed consent. The dataset comprises 883 biopsy cores obtained from 131 patients. Raw RF ultrasound data were captured using a BK3500 ultrasound system equipped with an E10C4 endocavity transducer. Each biopsy core, approximately 18 mm in length, is associated with 200 consecutive TRUS RF frames recorded over 5 seconds while the transducer was held stationary prior to tissue sampling. Histopathological analysis served as the ground truth for each biopsy core, identifying cancer presence and quantifying its involvement as a percentage.
Preprocessing: To generate image inputs, the RF data were converted into B-mode images. For training, a random frame was selected from each sequence, while for validation and testing, the last frame was used for consistency. In the case of time-series data, temporal signals were extracted from consecutive frames within the biopsy needle region.

2.2 Methods

Figure 1 illustrates our proposed framework. A temporal signal is extracted from all pixels within the biopsy region (a single-pixel signal is shown for illustration)

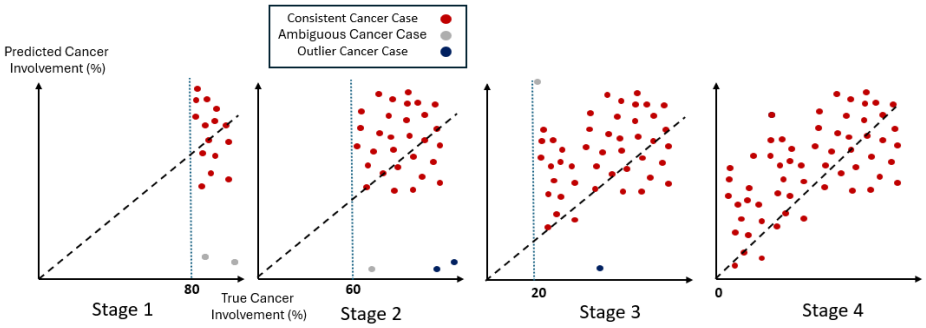

Fig. 2: Progressive training in four stages, gradually adding low-involvement cases. High-loss samples are flagged as ambiguous, re-evaluated in the next stage, and removed if losses persist.

and fed into a time-series encoder. The resulting embeddings are aggregated via adaptive pooling and passed to a prompt encoder. Patient metadata and other-core information (used during the training phase only) share the same text encoder but use separate projection layers. These prompts and image features extracted by the image encoder guide the image decoder in generating a cancer segmentation mask for the entire prostate. During training, the biopsy region is isolated, and an involvement-aware loss function is applied.

Segmentation Backbone. We use the pre-trained MedSAM foundation model [12]. The Vision Transformer (ViT) image encoder, with approximately 90M parameters, is used to extract visual features from a 1024×1024 ultrasound B-mode image. The ViT has a patch size of 16, resulting in a 64×64 feature grid, each embedding having 256 dimensions. These embeddings are then passed to a 6M-parameter mask decoder, which produces a 256×256 segmentation map guided by 256-dimensional sparse prompt embeddings. Both the ViT encoder and mask decoder are trained end-to-end on our task.

Time-Series Encoder. To incorporate the RF ultrasound signals, we employ an InceptionTime encoder [4] (6M parameters) with a depth of 9. It processes sequences of length 200, with 600 input channels and 256 output channels. We set the bottleneck channels and kernel sizes to 12 and 15, respectively. This fully trained encoder captures vital temporal features for accurate characterization of cancer.

Text Encoder. We also integrate non-imaging information through a text encoder based on PubMedBERT (110M parameters) [6]. The weights of PubMedBERT are kept frozen, yielding 768-dimensional embeddings for each textual input. These embeddings are then down-sampled to 256 dimensions via two linear projection layers (with GELU activation), ensuring compatibility with the prompt encoder. Separate projection heads handle patient metadata and additional core descriptions independently.

Prompt Encoder and Fusion Strategy. Each modality is encoded via a modality-specific projection layer into a 256-dimensional vector. These modality-

specific embeddings are concatenated and then passed through a shared linear layer to form a unified prompt representation, which guides the mask decoder.

Loss Function. We formulate PCa detection as a weakly supervised task by leveraging coarse involvement data. Let $\mathcal{R}$ be the set of ROI pixels, defined by intersecting the needle region N with the prostate mask P:

$$\mathcal{R} = \{(i,j) \mid i,j \in [1,1024],\ N[i,j] \neq 0 \wedge P[i,j] \neq 0\}. \tag{1}$$

We propose an *involvement-aware hybrid loss*, which combines two specialized domain-specific loss functions explored in our prior work [7], iMSE (involvement-aware mean squared error) and iMAE (involvement-aware mean absoluter error). Given the ground-truth involvement score $\text{inv} \in [0,1]$, representing the fraction of cancerous tissue within a core, our loss function is defined as:

$$\mathcal{L}_{\text{iMSE}}(\hat{Y}, \text{inv}) = \frac{1}{|\mathcal{R}|} \sum_{(i,j)\in\mathcal{R}} \big(\hat{Y}[i,j] - \text{inv}\big)^2, \tag{2}$$

$$\mathcal{L}_{\text{iMAE}}(\hat{Y}, \text{inv}) = \frac{1}{|\mathcal{R}|} \sum_{(i,j)\in\mathcal{R}} \big|\hat{Y}[i,j] - \text{inv}\big|, \tag{3}$$

$$\mathcal{L}_{\text{iHybrid}}(\hat{Y}, \text{inv}) = \frac{1}{2}\Big(\mathcal{L}_{\text{iMSE}} + \mathcal{L}_{\text{iMAE}}\Big). \tag{4}$$

This design combines the strong penalty for large errors (iMSE) with a linear penalty for consistent reductions (iMAE), stabilizing training under label noise and promoting more accurate modeling of spatial cancer spread.

Progressive Training Strategy. Our training scheme, depicted in Figure 2, progressively introduces more challenging cancer-core examples while filtering outliers across four stages (each stage spans five epochs). In Stage 1, all benign cores and only cancer cores with greater than 80% involvement are used. After the first stage, we identify ambiguous cases using a 0.95 error quantile criterion; these suspicious outliers are flagged for re-evaluation. Stage 2 lowers the cancer involvement threshold to 60%, again flagging ambiguous cores at the end of the stage. Outliers marked in previous stages are either permanently removed if their errors remain high, or reinstated if improved. Stages 3 and 4 repeat this procedure with thresholds of 40% and 20%, respectively, allowing the model to adapt incrementally to increasingly subtle cancer involvement levels while discarding persistent outliers.

2.3 Experiments

Benchmarking. We conduct a series of experiments on a wide range of baseline methods from the PCa detection literature, including UNet [16], SAM [11], and MedSAM [12] (without prompt embeddings), all using B-mode images exclusively. We also explore InceptionTime [4] and TimesNet [26], time-series models

Table 1: Performance comparison with prior methods for PCa detection. AUROC, Sensitivity at 60% specificity (SEN@60SPE), and Balanced Accuracy (%) are reported as mean ± standard deviation across 5-fold cross-validation.

Method	AUROC	SEN@60SPE	BACC
TimesNet [26]	69.3 ± 3.2	71.2 ± 3.0	65.4 ± 3.2
InceptionT [4]	73.1 ± 2.5	75.2 ± 2.8	68.4 ± 3.0
UNet [16]	75.2 ± 3.0	78.3 ± 3.6	70.2 ± 3.2
SAM [11]	77.9 ± 2.9	78.2 ± 4.2	70.0 ± 3.8
MedSAM [12]	81.0 ± 2.6	81.2 ± 3.8	73.0 ± 3.5
ProstNFound [25]	83.4 ± 2.1	86.2 ± 3.4	75.7 ± 2.0
Cinepro [7]	85.8 ± 1.9	88.0 ± 3.2	77.0 ± 1.7
ProTeUS	$\mathbf{86.9 \pm 1.1}$	$\mathbf{90.1 \pm 3.3}$	$\mathbf{77.9 \pm 1.5}$

trained solely on time-series RF data. Finally, we train two SOTA foundation model methods for PCa detection, ProstNFound [25] and Cinepro [7], currently regarded as the SOTA in PCa detection.

Training and Evaluation. We perform an 80/20 data split, allocating 80% of patients for training and validation through a 5-fold cross-validation scheme, and reserving 20% as an independent test set. The splits ensure no patient-level overlap, maintaining evaluation integrity.. We use AdamW as the optimizer, applying a learning rate of 1×10^{-5} for the image encoder, mask decoder, and prompt encoder, and 1×10^{-4} for the time-series encoder and projection layers. A cosine-annealing schedule governs the learning rate throughout training. All experiments are conducted on a single NVIDIA RTX 6000 GPU with a batch size of 1, and each training epoch completes in approximately 30 minutes.

3 Results and Discussion

Quantitative Results. Table 1 presents our findings in comparison to other competing methods. We evaluate performance using AUROC, sensitivity, and specificity. AUROC is computed by comparing the predicted continuous involvement scores against ground truth annotations. Sensitivity and specificity are derived from the ROC curve, with sensitivity reported at 60% specificity—i.e., the true positive rate at the point where specificity equals 60%. Our results demonstrate that whole-image methods outperform time-series models, with standard architectures like UNet exceeding InceptionTime by 2.1%. The gap widens with foundation models, where SAM and MedSAM surpass InceptionTime by 4.8% and 7.9%, respectively. Encoding additional information further enhances performance: ProstNFound outperforms MedSAM by 2.4%, while Cinepro surpasses MedSAM by 4.8% and ProstNFound by 2.4% due to its robust loss function and temporal cine-series augmentation.

By integrating time-series RF signals with a foundation model encoder, ProTeUS achieves state-of-the-art performance, outperforming Cinepro by 1.1% and

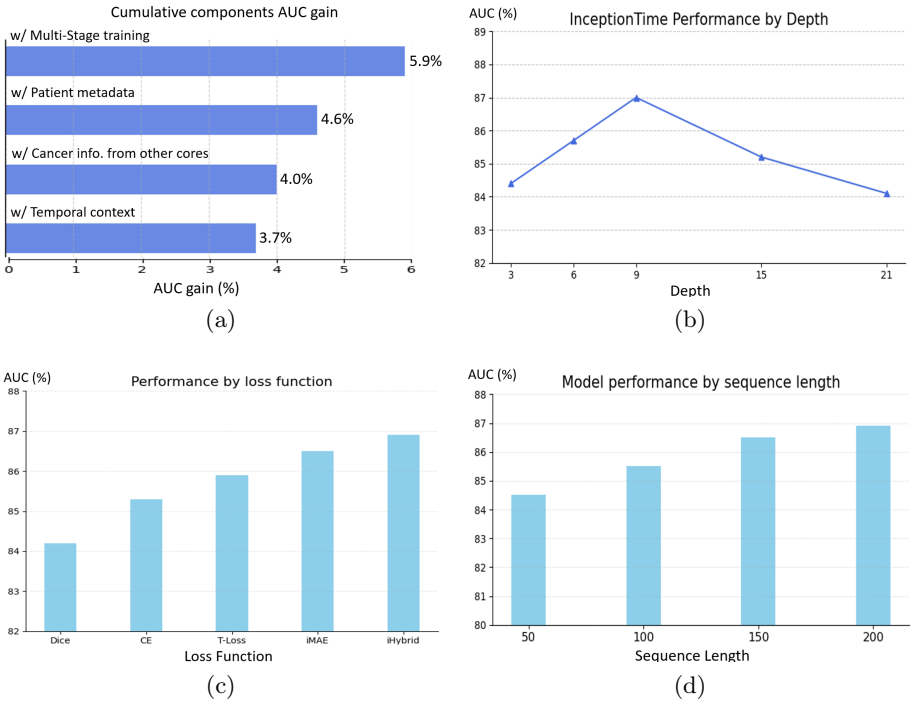

Fig. 3: (a) The effect of cumulative performance improvement to MedSAM after adding different components. (b) The effect of increasing the depth of InceptionTime network. (c) Performance with different loss functions. (d) The effect of increasing the time-series sequence length.

ProstNFound by 3.5%. It also surpasses the current SOTA by 0.9% in balanced accuracy and 2% in sensitivity (at 60% specificity). Statistical significance over CinePro is confirmed (p=0.003), and clinically, ProTeUS identified three additional cancerous cores, including two high-risk Gleason score 9 cores with small (<10%) involvement, demonstrating the power of spatio-temporal learning for PCa detection.

Ablation Studies. We conducted a series of ablation experiments to assess the impact of each component in our framework. Figure 3(a) illustrates the incremental contributions of (i) temporal RF embeddings from InceptionNet, (ii) additional core-level cancer information (during training only), (iii) patient metadata (PSA and PSA density), and (iv) progressive training strategy. The largest performance gain arises from the time-series data, indicating that temporal cues play a crucial role. Cancer information from other cores provides broader context, enabling more discriminative visual features, and patient metadata refines these representations further. Gradually introducing lower-involvement cores via progressive training then stabilizes learning and boosts overall performance.

Figure 3(b) shows that while moderate network depth enhances performance, going too deep eventually leads to overfitting and a decline in AUROC. In Figure 3(d), varying the time-series sequence length shows that longer sequences capture richer temporal information, yielding higher performance.

We also compared multiple loss functions, including dice, cross-entropy, T-Loss [5], and our *involvement-aware hybrid loss*, $\mathcal{L}_{\text{iHybrid}}$. As illustrated in Figure 3(c), $\mathcal{L}_{\text{iHybrid}}$ outperforms single-term alternatives by combining strong penalties for large errors (iMSE) with consistent error reduction (iMAE), leading to balanced gradients and improved robustness against coarse labels.

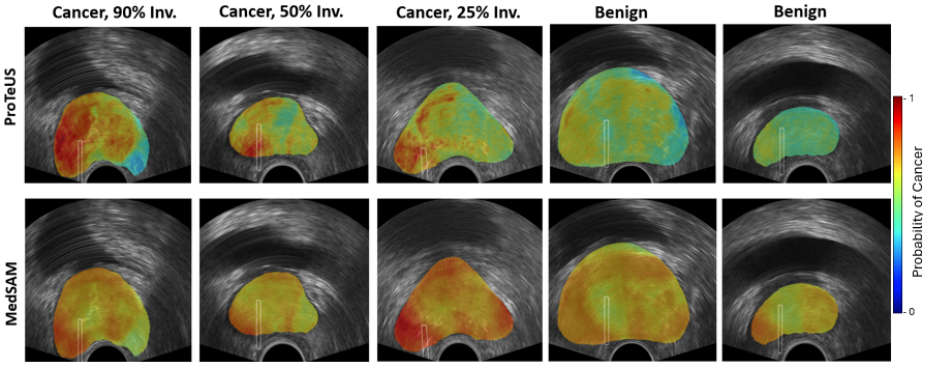

Fig. 4: Qualitative comparison of predicted cancer heatmaps from ProTeUS and the fine-tuned MedSAM baseline, against the histopathology labels. Red regions signify higher cancer likelihood. The biopsy samples are overlaid on each image with white rectangles. The labels are derived from the biopsy samples. An alternative version of this figure with an accessibility-friendly color palette is available in our GitHub repository.

Qualitative Results. Figure 4 compares the cancer heatmaps predicted by ProTeUS and the fine-tuned MedSAM baseline. ProTeUS demonstrates a more accurate cancer distribution confined to the biopsy (needle) region, consistent with pathology findings. In benign samples, ProTeUS exhibits markedly lower activation levels, indicating enhanced tissue characterization and fewer false-positive regions compared to the baseline.

4 Conclusion

In this work, we presented a comprehensive framework for PCa detection that combines the strengths of foundation models, RF time-series data, patient metadata, and additional core-level cancer information. By unifying these distinct data modalities, our approach offers a richer representation of prostate tissue,

enabling more accurate lesion characterization. Furthermore, our involvement-aware loss function and progressive training strategy alleviate challenges posed by coarse labels, ensuring robust performance under real-world conditions. Collectively, these contributions pave the way for more targeted biopsy procedures, thereby reducing unnecessary interventions and optimizing clinical outcomes in PCa diagnosis.

Acknowledgments. This work was supported in part by the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), Vector AI Institute, and through computational resources and services provided by Advanced Research Computing at the University of British Columbia. P Mousavi is supported in part by Canada CIFAR AI Chair and Canada Research Chair.

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

References

1. Ahmed, H.U., Bosaily, A.E.S., Brown, L.C., Gabe, R., Kaplan, R., Parmar, M.K., Collaco-Moraes, Y., Ward, K., Hindley, R.G., Freeman, A., et al.: Diagnostic accuracy of multi-parametric mri and trus biopsy in prostate cancer (promis): a paired validating confirmatory study. The Lancet **389**(10071), 815–822 (2017)
2. Drost, F.J.H., Osses, D., Nieboer, D., Bangma, C.H., Steyerberg, E.W., Roobol, M.J., Schoots, I.G.: Prostate magnetic resonance imaging, with or without magnetic resonance imaging-targeted biopsy, and systematic biopsy for detecting prostate cancer: a cochrane systematic review and meta-analysis. European urology **77**(1), 78–94 (2020)
3. Elghareb, T., Jamzad, A., Nhat To, M.N., Fooladgar, F., Wilson, P.F., Sojoudi, S., Reznik, G., Leveridge, M., Siemens, R., Chang, S., Black, P., Mousavi, P., Abolmaesumi, P.: Self-supervised prototype learning for spatio-temporal enhanced ultrasound-based prostate cancer detection. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2025)
4. Fawaz, I., Lucas, B., Forestier, G., et al.: Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery **34**(6), 1936–1962 (2020)
5. Gonzalez-Jimenez, A., Lionetti, S., Gottfrois, P., Gröger, F., Pouly, M., Navarini, A.: Robust t-loss for medical image segmentation (2023)
6. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., Poon, H.: Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare **3**(1) (2021)
7. Harmanani, M., Jamzad, A., Nhat To, M.N., Wilson, P.F., Guo, Z., Fooladgar, F., Sojoudi, S., Gilany, M., Chang, S., Black, P., Leveridge, M., Siemens, R., Abolmaesumi, P., Mousavi, P.: Cinepro: Robust training of foundation models for cancer detection in prostate ultrasound cineloops. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2025)
8. Jahanandish, H., Vesal, S., Bhattacharya, I., Li, C.X., Fan, R.E., Sonn, G.A., Rusu, M.: A deep learning framework to assess the feasibility of localizing prostate cancer

on b-mode transrectal ultrasound images. In: Medical Imaging 2024: Ultrasonic Imaging and Tomography. vol. 12932, pp. 168–173. SPIE (2024)
9. Javadi, G., Samadi, S., Bayat, S., et al.: Training deep networks for prostate cancer diagnosis using coarse histopathological labels. In: MICCAI. pp. 680–689. Springer (2021)
10. Karimi, D., Dou, H., Warfield, S.K., Gholipour, A.: Deep learning with noisy labels: exploring techniques and remedies in medical image analysis (2020)
11. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything (2023)
12. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15** (2024)
13. Mate, K., Nedjim, S., Bellucci, S., Boucault, C., Ghaffar, N., Constantini, T., Marvanykovi, F., Vestris, P.G., Sadreux, Y., Laguerre, M., et al.: Prostate biopsy approach and complication rates. Oncology Letters **26**(3), 1–7 (2023)
14. Panzone, J., Byler, T., Bratslavsky, G., Goldberg, H.: Transrectal ultrasound in prostate cancer: Current utilization, integration with mpmri, hifu and other emerging applications. Cancer Management and Research pp. 1209–1228 (2022)
15. Rawla, P.: Epidemiology of prostate cancer. World Journal of Oncology **10**(2), 63 (2019)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015)
17. Rusu, M., Jahanandish, H., Vesal, S., Li, C.X., Bhattacharya, I., Venkataraman, R., Zhou, S.R., Kornberg, Z., Sommer, E.R., Khandwala, Y.S., et al.: Procusnet: Prostate cancer detection on b-mode transrectal ultrasound using artificial intelligence for targeting during prostate biopsies. European Urology Oncology (2025)
18. Secasan, C., Onchis, D., Bardan, R., et al.: Artificial intelligence system for predicting prostate cancer lesions from shear wave elastography measurements. Current Oncology **29**(6), 4212–4223 (2022)
19. Sedghi, A., Mehrtash, A., Jamzad, A., et al.: Improving detection of prostate cancer foci via information fusion of mri and temporal enhanced ultrasound. IJCARS **15**, 1215–1223 (2020)
20. Song, H., Kim, M., Park, D., Shin, Y., Lee, J.G.: Learning from noisy labels with deep neural networks: A survey (2022)
21. Sun, Y., Fang, J., Shi, Y., et al.: Machine learning based on radiomics features combing b-mode transrectal ultrasound and contrast-enhanced ultrasound to improve peripheral zone prostate cancer detection. Abdominal Radiology **49**(1), 141–150 (2024)
22. To, M., Fooladgar, F., Wilson, P., et al.: Lensepro: label noise-tolerant prototype-based network for improving cancer detection in prostate ultrasound with limited annotations. IJCARS **19**, 1121–1128 (2024)
23. Vesal, S., Bhattacharya, I., Jahanandish, H., Li, X., Kornberg, Z., Zhou, S.R., Sommer, E.R., Choi, M.H., Fan, R.E., Sonn, G.A., Rusu, M.: Prosdectnet: Bridging the gap in prostate cancer detection via transrectal b-mode ultrasound imaging (2023)
24. Wilson, P.F., Harmanani, M., To, M.N.N., Gilany, M., Jamzad, A., Fooladgar, F., Wodlinger, B., Abolmaesumi, P., Mousavi, P.: Toward confident prostate cancer detection using ultrasound: a multi-center study. International Journal of Computer Assisted Radiology and Surgery **19**(5), 841–849 (2024)

25. Wilson, P.F., To, M.N.N., Jamzad, A., Gilany, M., Harmanani, M., Elghareb, T., Fooladgar, F., Wodlinger, B., Abolmaesumi, P., Mousavi, P.: Prostnfound: Integrating foundation models with ultrasound domain knowledge and clinical context for robust prostate cancer detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 499–509. Springer (2024)
26. Wu, H., Hu, T., Liu, Y., et al.: Timesnet: Temporal 2D-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186 (2022)