ProstNFound: Integrating Foundation Models with Ultrasound Domain Knowledge and Clinical Context for Robust Prostate Cancer Detection

Paul F. R. Wilson[1(✉)], Minh Nguyen Nhat To[2], Amoon Jamzad[1], Mahdi Gilany[1], Mohamed Harmanani[1], Tarek Elghareb[2], Fahimeh Fooladgar[2], Brian Wodlinger[3], Purang Abolmaesumi[2], and Parvin Mousavi[1]

[1] Queen's University, Kingston, ON, Canada
1pfrw@queensu.ca
[2] University of British Columbia, Vancouver, BC, Canada
[3] Exact Imaging, Markham, ON, Canada

Abstract. Analysis of high-resolution micro-ultrasound data using deep learning presents a promising avenue for the accurate detection of prostate cancer (PCa). While previous efforts have focused on designing specialized architectures and training them from scratch, they are challenged by limited data availability. Medical foundation models, pretrained on large and diverse datasets, offer a robust knowledge base that can be adapted to downstream tasks, reducing the need for large task specific datasets. However, their lack of specialized domain knowledge hinders their success: our initial research indicates that even with extensive fine-tuning, existing foundation models falls short of surpassing specialist models' performance for PCa detection. To address this gap, we propose ProstNFound, a method that empowers foundation models with domain-specific knowledge pertinent to ultrasound imaging and PCa. In this approach, while ultrasound images are fed to a foundation model, specialized auxiliary networks embed high-resolution textural features and clinical markers which are then presented to the network as prompts. Using a multicenter micro-ultrasound dataset with 693 patients, we demonstrate significant improvements over the state-of-the-art in PCa detection. ProstNFound achieves 90% sensitivity at 40% specificity, performance that is competitive with that of expert radiologists reading multi-parametric MRI or micro-ultrasound images, suggesting significant promise for clinical application. Our code is available at github.com/pfrwilson/prostNfound.

Keywords: Ultrasound Imaging · Computer Assisted Interventions · Prostate Cancer · Foundation Models · Transfer Learning

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-72089-5_47.

M. G. Linguraru et al. (Eds.): MICCAI 2024, LNCS 15006, pp. 499–509, 2024.
https://doi.org/10.1007/978-3-031-72089-5_47

1 Introduction

Ultrasound (US) based tissue characterization, using B-mode and raw radio frequency (RF) data, is a long-standing and complex challenge with extensive clinical applications [18,20,21]. Advanced US modalities, including contrast-enhanced US, Doppler US, micro-US, and temporal-enhanced US have been proposed in the literature, combined with quantitative approaches to analyze the complex interactions between US waves and tissue structure. In particular, micro-US uses a high frequency clinical transducer to obtain spatial resolutions as low as 70μm, and has shown promising performance in tissue characterization [14]. While traditional methods have used handcrafted features from US data such as texture [15] and spectrum [21] for cancer detection, deep learning offers a more powerful and versatile approach for feature learning. However, its reliance on large, task and modality specific datasets poses major roadblocks to its full utility. This challenge is exemplified in prostate cancer (PCa) diagnosis, where accurate tissue characterization using transrectal US could improve biopsy targeting or provide non-invasive risk assessment, highly desirable clinical goals. The dominant learning paradigm has been to train *specialist* models using various US modalities and to apply them for tissue classification. These specialist models have been hand designed and are mainly CNN architectures that examine patches of prostate images from conventional US [22], contrast-enhanced US [10], and micro-US [23] data. Though specialist models leverage strong, task-specific inductive biases conferred by their network architecture design choices, their success is hindered by the availability of task-specific training data.

To alleviate the bottleneck in data availability, we explore the growing field of *foundation models* [3], specifically medical foundation models which have been trained on large, multi-modal image datasets [8,16]. In principle, this training offers a knowledge base that can be fine-tuned to specific applications, reducing reliance on large task-specific datasets. In practice, multiple challenges limit their direct application in PCa detection. First, substantial retraining is often needed to apply foundation models to new data modalities [12]. Second, existing models are predominantly developed for segmentation. They are, hence, not immediately well suited to tissue characterization, which depends on identification of subtle differences in tissue appearance without clear boundaries. Third, foundation models lack inductive biases proven to be useful for tissue characterization, such as local texture features, which may not be adequately modeled by the vision transformer backbones of these models [4]. Finally, current models do not explicitly incorporate clinical context. An ideal solution should leverage pre-trained knowledge of foundation models while adapting them to the new domain and task and adding ultrasound-specific inductive bias and clinical context.

We introduce ProstNFound, a micro-US based model for PCa detection, which leverages US-specific inductive biases and domain knowledge along with the large-scale medical foundation models through a new architecture design based on the conditional prompt [25] paradigm. Our key contributions are:

1. We present the first method (to the best of our knowledge) that successfully leverages foundation models for US tissue characterization by integrating foundational knowledge with specialized domain knowledge for the tasks.
2. We propose a novel architecture to integrate local texture features and clinical context with foundation models. This involves incorporating a CNN with self-supervised in-domain pretraining to extract patch-based texture features, alongside an auxiliary network that learns embeddings of clinical markers, all passed to the foundation model as prompts.
3. Using a large, multi-center dataset, we demonstrate that our method significantly outperforms previous SOTA in micro-US PCa detection, achieving 90% sensitivity at 40% specificity, comparable with analysis of multi-parametric MRI and micro-US images by expert clinicians.

2 Materials and Method

2.1 Data

We use proprietary data from 693 patients across five medical centers who underwent systematic TRUS-guided biopsy as part of a clinical trial (clinicaltrials.gov NCT02079025), following patients' consent. Biopsies were performed using the ExactVu micro-ultrasound system (Exact Imaging, Markham, Canada). Before activating the biopsy gun, RF and B-Mode images were captured with dimensions of 28 mm depth and 46.06 mm width in the sagittal plane. Patients underwent retrieval of 10 to 12 biopsy cores from standardized sextant biopsy locations. Histopathology reports indicated the presence of Gleason Score (GS), and length of cancer in cores. Of 6607 total cores, 5727 are benign (defined as GS$\leq$6), 675 are GS7, 134 are GS9, 60 are GS9, and 11 are GS10. Additional patient metadata includes prostate specific antigen (PSA), age, and family history of prostate cancer but racial/ethnic data are not available. In addition to US images, masks representing the approximate area of the biopsy needle trace and the prostate gland are provided (see Supplementary Table 1 for more details on data).

B-Mode images are resized to 1024×1024 pixels via bilinear interpolation, individually normalized to the pixel range $(0, 1)$, and repeated along the channel axis to conform to the RGB input format of the foundation model [16]. Grades are converted to binary labels of 0 (no cancer) and 1 (cancer). PSA and age values are normalized by subtracting the mean and dividing by the range. PSA density is approximated by dividing the PSA by the area of the largest prostate section in the US images of a patient. Between 20 and 40 patches of 5 mm $\times$ 5 mm of tissue are extracted from the needle trace region of corresponding RF and B-Mode images. Compared to B-mode, RF captures not only the intensity but also the frequency of backscattered US waves, providing additional subresolution information on tissue microstructure [17], justifying its possible utility.

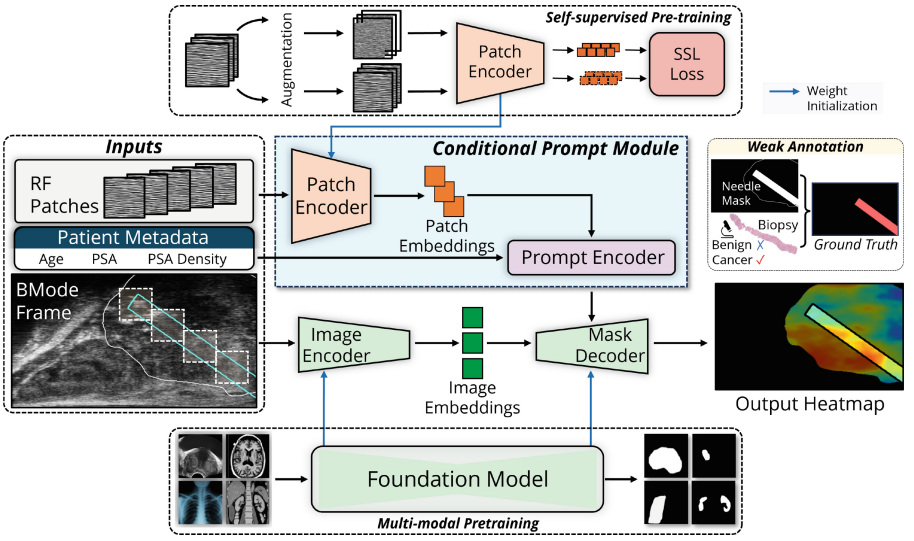

Fig. 1. ProstNFound combines a B-mode image encoder and a conditional prompt module embedding raw RF patch data (via a CNN patch encoder) and patient metadata. These embeddings feed to a mask decoder to produce a heatmap of cancer likelihood. Training leverages histopathology labels in the needle region. The patch encoder is initialized from self-supervised pretraining and the image encoder and mask decoder are initialized from a medical image foundation model.

2.2 Model

Figure 1 depicts an overview of our proposed approach and its components. **Foundation Model:** Our model borrows the image encoder and mask decoder components from the MedSAM [16] medical segmentation foundation model. The encoder ($\approx$ 90M parameters) is a vision transformer used to extract $256 \times 64 \times 64$ embeddings of B-mode images. The mask decoder ($\approx 6M$ parameters) accepts the image embeddings along with zero or more 256-dimensional prompt embeddings to generate a 256×256 "heatmap", and uses two attention blocks between the image and prompt embeddings followed by two deconvolution layers.

Conditional Prompting: To combine the foundation model with US specific inductive biases and clinical context we design two additional networks. This is inspired by the conditional prompts [25] paradigm, which uses auxiliary networks to generate learnable, data-dependant "prompts" passed to the foundation model. (i) The first network uses a self-supervised CNN encoder to extract patch-level texture features, following successful approaches in the literature [23]. Specifically, we train ResNet10 ($\approx 5M$ parameters) to extract a 512-dimensional embedding from a 256×256 (RF or B-Mode) patch using the VICReg [5] self-supervised learning framework. VICReg forces invariance between features extracted from the same patch following two different augmentations, and allows

learning of rich features without labels.[1] We extract one embedding per patch and max-pool all patch embeddings for a given image. These pooled embeddings are projected to 256 dimensions using a 2-layer MLP. (ii) The second network embeds relevant clinical markers – age, PSA, and approximate PSA density – using three 2-layer MLPs to project the markers' scalar values to 256 dimensions. Overall, the prompt encoder outputs four 256-dimensional prompt embeddings.

Loss Function: To train the network, we frame cancer detection as a weakly supervised, pixel-level classification task. The output of the network is pixel-wise predictions of cancer versus benign. However, pixel-wise gold-standard labels are unavailable. Instead, we use weak labeling by applying the label from a biopsy core to all pixels within the needle trace region of that core. We use a masked cross-entropy loss function to force the model outputs to match the binary label of cancer or benign within the region. Concretely, let $(X, N, P, \mu, \{\rho_1, ..., \rho_k\}, y)$ denote the image, needle mask, prostate mask, metadata, patches and label, respectively. Let $f(\cdot)$ denote our network, and $\mathcal{L}_{\text{CE}}$ denote the standard cross-entropy loss. The loss function for the sample is:

$$\mathcal{L}_{\text{MaskedCE}} = \sum_{i=1}^{1024} \sum_{j=1}^{1024} \mathbb{1}_{N[i,j]\neq 0} \mathbb{1}_{P[i,j]\neq 0} \mathcal{L}_{\text{CE}}\big(f(X, \mu, \{\rho_1, ...\rho_n\}), y\big),$$

Training Strategy: To adapt the foundation model backbone, we experiment with several strategies from the literature [24]. We implement prompt tuning (training only prompt components), partial finetuning (tuning mask decoder and prompt modules but not the image encoder), full finetuning (training every weight), two-stage tuning (partial followed by full finetuning) and adapter tuning (training mask decoder and prompt modules while freezing image encoder except for added trainable "adapter" layers). The latter option has the best performance, likely as it allows the image encoder to adapt to micro-US data without forgetting the pretraining of the foundation model, and is selected for ProstNFound. The patch encoder CNN module is also trained, but at a lower learning rate than the mask decoder and prompt decoder to avoid forgetting its SSL pretraining.

2.3 Experiments

Our primary experiments are as follows: (i) *SOTA comparison.* We compare the performance of our model with prior art for PCa detection in US, including patch classification with ResNet [23] with and without self-supervised pretraining (SSL), and TRUSFormer [11] which uses patch feature extraction with transformer-based aggregation. For comparison with other foundation model strategies, we implemented SAM-UNETR [2], an extension of SAM for MRI-based PCa detection that we reimplemented for micro-US, and its variant Med-SAM-UNETR, which we implemented by swapping the SAM backbone with

[1] The choice of VICReg follows the literature and works well, but studying the specific algorithm for self-supervised feature extraction is beyond the scope of this work.

MedSAM, alongside adapter-tuned MedSAM and SAM-Med2d [8]; (ii) *Ablation studies.* We perform three types of ablations. First, we evaluate the contribution of each type of conditional prompt. Second, we compare the different adaptation strategies with and without conditional prompts. Third, we test the effects of training on increasing fractions of data with and without conditional prompts.

Model Evaluation: To evaluate the models, we use a leave-one-center-out strategy, holding all data from a given center for testing in each experiment. Meanwhile, the remaining four centers are stratified into training patients and validation patients. Experiments are repeated with the three largest centers as test data and metrics are averaged. Model performance is tested at "core" level by measuring the average heatmap value within the intersection of needle and prostate regions compared to histopathology (cancer versus benign) for that core, and at "patient" level by measuring the average heatmap intensity in the prostate across all cores of a given patient compared to the overall finding of clinically significant (at least one core with Gleason score $> 4{+}3$) or no clinically significant disease. AUROC and sensitivity at several specificity levels are measured. Additionally, core-level predictions are compiled from all three centers and used to generate a ROC curve for detailed analysis. Hyperparameters are tuned to maximize the average validation core AUROC during the last five epochs of training, and, once the best hyperparameters are found, average test performance during the last five epochs is reported.

Implementation Details: Python 3.11 and PyTorch 2.1 are used. Model training takes approximately 8 h on a single NVIDIA A40 GPU. Inference on a single example takes 2.5 s on CPU or 150ms on GPU. The hyperparameters are tuned via manual search. Training parameters include a learning rate of $1e$-5 (main model and image encoder) and $1e$-6 (patch encoder) for 35 epochs with 5 epochs of linear warmup followed by cosine annealing, a batch size of 8 and Adam optimizer with 0 weight decay. Training loss did not improve at higher learning rates or without learning rate warm up. Data augmentation using random translations is found to improve performance. MedSAM was chosen due to its superior performance compared to other applicable models with publicly available weights (SAM [13] and SAMMed-2d [8]) that we identified. For SSL training, 5 epochs, a batch size of 64 and learning rate of 1e-4 with default VICReg loss parameters are used. Details of hyperparameter search are provided in supplementary Table 2. Full code and configurations are provided in our repository at github.com/pfrwilson/prostNfound.

3 Results and Discussion

Comparison to Prior State of the Art: Results of SOTA comparison are shown in Table 1. Overall, ProstNFound outperforms all baselines, and by a substantial margin in core AUROC, patient AUROC and sensitivity at 40% and 60% specificity. This includes prior work in micro-US PCa detection (rows 1–3) in particular the prior SOTA TRUSFormer (improved by 7.2% core AUROC).

Since previous approaches were specialist models, this indicates the benefits of foundation model integration. Furthermore, ProstNFound substantially outperforms foundation-based methods that do not leverage conditional prompts (rows 4–7), even those with the same MedSAM backbone. This highlights the advantage of integrating ultrasound and PCa specific priors through our conditional prompt module over straightforward application of foundation models without this component. While we achieve strong performance using only B-mode patches (second last row), leveraging RF patches results in SOTA core AUROC of 78.3% and 90% sensitivity at 40% specificity (last row). We hypothesize that the additional benefits of RF are due to its ability to resolve richer backscattering features associated with tissue micro-architecture, although more work is needed to investigate this potential benefit. The competitive performance of B-Mode is of practical importance since RF is not always available clinically. Overall, the dominance of our model strongly justifies our design choices.

Table 1. Prostate cancer detection performance in Micro-ultrasound.

Method	AUROC %		Sensitivity (Core) %		
	Core	Patient	20% SPE	40% SPE	60% SPE
ResNet10-Patch [23]	65.8	-	91.0	80.9	63.7
ResNet10-Patch (SSL) [23]	67.4	-	92.9	81.8	66.3
TRUSFormer [11]	71.1	-	93.8	83.4	71.6
SAM-UNETR [2]	66.7	62.8	93.0	82.1	64.5
MedSAM-UNETR	70.0	63.7	93.1	82.8	69.9
MedSAM+Adapter Tuning	71.2	68.8	94.3	85.2	70.9
SAM-Med2D+Adapter Tuning	67.0	64.9	92.9	81.1	64.9
ProstNFound *(ours)*	77.6	73.8	95.4	89.9	79.7
ProstNFound + RF*(ours)*	**78.3**	**76.1**	**96.4**	**90.0**	**80.1**

Ablation Studies: Figure 2 shows the results of our ablation studies. For prompt ablation (Fig. 2A) we observe a trend: the incorporation of each successive prompt – PSA, age, patch features, and PSA density – improves model performance, cumulatively resulting in a substantial increase of 7.1% core AUROC over baseline, highlighting the efficacy of each additional feature to improve PCa detection. Our prompting method is effective and flexible, enabling the integration of arbitrary clinical data. Emerging biomarkers such as PCA3, 4K score and PHI, have been demonstrated to be effective for PCa diagnosis and treatment planning [7] - incorporating these as prompts into our method is a promising direction for future work. Regarding the efficacy of various adaptation strategies (Fig. 2B), adapter tuning has the best overall performance. The introduction of our conditional prompt module significantly boosts performance across all adaptation strategies. In particular, it still enables good performance

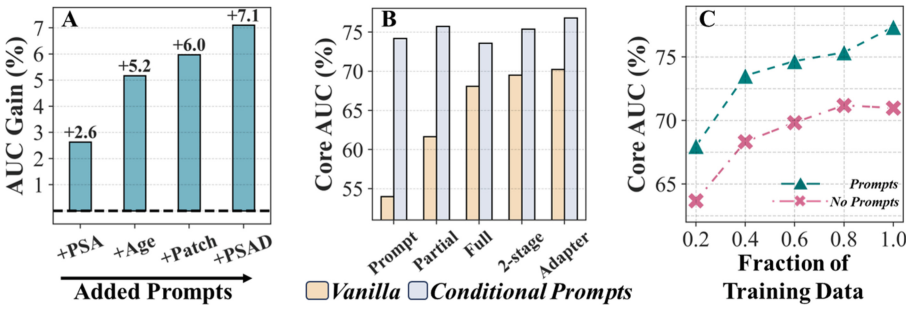

Fig. 2. (a) Comparison of different combinations of conditional prompts; (b) Comparison of different adaptation strategies; (c) Adapter tuning with and without prompts.

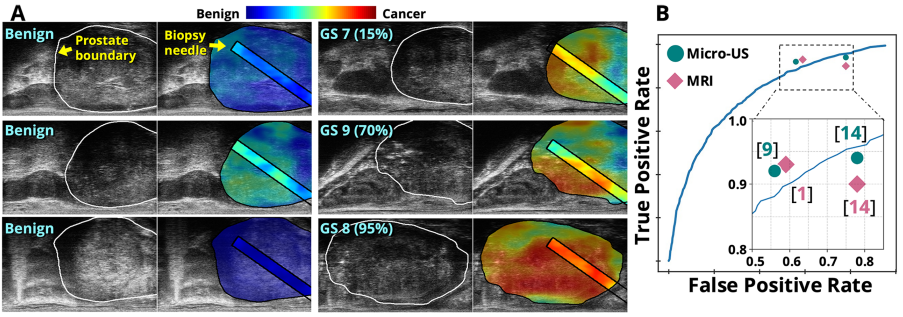

Fig. 3. (a) Visualization of model outputs as cancer heatmaps. Left column: model has no or low cancer activations for benign cores. Right column: model identifies suspicious cancerous lesions (red areas), confirmed by histopathology; (b) Our model (blue ROC curve) has competitive performance with expert readers in clinical studies cited. (Color figure online)

with only minimal tuning approaches, such as prompt tuning or partial tuning, indicating that it can diminish the need for extensive retraining of the foundation model. Finally, analysing the impact of limited training data (Fig. 2C) reveals that the utilization of the conditional prompt module improves performance. In particular, it allows reasonable (73% AUROC) performance with only 40% of training data, but shows continued improvements as the volume of training data increases, indicating the dual benefits of our approach: enabling effective learning with reduced data while also scaling well with increased data availability.

Qualitative Evaluation and Clinical Impact: Given our significant breakthrough in performance compared to prior PCa detection studies, it is important to evaluate the potential clinical impact of our model. To simulate clinical deployment, heatmaps generated by our model are overlaid on the corresponding ultrasound input images (Fig. 3A). The model has low-to-no activations for benign cores, while highlighting suspicious lesions in red which approximately

match the biopsy-reported cancer percentage in cores. This simulation suggests that our model could enable *targeted biopsy*, where the clinician acquires additional samples from regions flagged by the model. This could substantially reduce the sampling error (missing a tumour) during the biopsy procedure. Our model could complement existing visual protocols (e.g., PRIMUS for micro-US, PI-RADS for MRI), which require extensive training [6,19], by offering objective, user-independent feedback to complement the clinicians' judgement.

To evaluate our model's performance relative to established visual protocols, we compare the ROC curve of our model against the performance metrics of trained experts using PI-RADS or PRIMUS, as documented in recent literature [1,9,14]. As shown in Fig. 3B, at matched false positive rates, our model surpasses expert performance in two out of four scenarios (with increases in sensitivity of 1% and 6%), equals expert performance in another, and falls short in one scenario (with a 5% decrease in sensitivity). The overall competitive performance of our model against that of human experts highlights its promising potential for clinical application. Prospective validation and clinical translation studies should therefore be the subject of immediate future work.

4 Conclusion

As foundation models open up a new horizon in medical imaging research, domain knowledge remains of central importance in mastering specialized clinical applications. We propose ProstNFound, a method which empowers medical foundation models with ultrasound and prostate specific priors using prompts, substantially improving the state of the art in micro-US PCa detection. The success of this model, and its apparent ability to match performance of human experts analyzing micro-US and MRI, represents a major step forward in computer-assisted PCa diagnosis and opens up new possibilities for clinical applications.

Acknowledgments. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canadian Institutes of Health Research (CIHR). Parvin Mousavi is supported by Canada CIFAR AI Chair and the Vector Institute.

Disclosure of Interests. Brian Wodlinger is Vice President of Clinical and Engineering at Exact Imaging, and provided access to the dataset used in this study. No other author has any potential conflict of interest to disclose.

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. Alzate-Grisales, J.A., Mora-Rubio, A., García-García, F., Tabares-Soto, R., De La Iglesia-Vayá, M.: Sam-unetr: Clinically significant prostate cancer segmentation using transfer learning from large model. IEEE Access **11**, 118217–118228 (2023)
3. Azad, B., Azad, R., Eskandari, S., Bozorgpour, A., Kazerouni, A., Rekik, I., Merhof, D.: Foundational models in medical imaging: A comprehensive survey and future vision. arXiv preprint arXiv:2310.18689 (2023)
4. Azad, R., Kazerouni, A., Azad, B., Khodapanah Aghdam, E., Velichko, Y., Bagci, U., Merhof, D.: Laplacian-former: Overcoming the limitations of vision transformers in local texture detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 736–746. Springer (2023)
5. Bardes, A., Ponce, J., Lecun, Y.: Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In: ICLR 2022-International Conference on Learning Representations (2022)
6. Cash, H., Hofbauer, S., Shore, N., Pavlovich, C.P., Bulang, S., Schostak, M., Planken, E., Jaspars, J.J., Luger, F., Klotz, L., et al.: Prostate cancer detection by novice micro-ultrasound users enrolled in a training program. Société Internationale d'Urologie Journal **3**(2), 62–68 (2022)
7. Chen, J.Y., Wang, P.Y., Liu, M.Z., Lyu, F., Ma, M.W., Ren, X.Y., Gao, X.S.: Biomarkers for prostate cancer: from diagnosis to treatment. Diagnostics **13**(21), 3350 (2023)
8. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. arXiv preprint arXiv:2308.16184 (2023)
9. Dias, N., Colandrea, G., Botelho, F., Rodriguez-Sanchez, L., Lanz, C., Macek, P., Cathelineau, X.: Diagnostic accuracy and clinical utility of micro-ultrasound guided biopsies in patients with suspected prostate cancer. Central European Journal of Urology **76**(1), 25 (2023)
10. Feng, Y., Yang, F., Zhou, X., Guo, Y., Tang, F., Ren, F., Guo, J., Ji, S.: A deep learning approach for targeted contrast-enhanced ultrasound based prostate cancer detection. IEEE/ACM Transactions on Computational Biology and Bioinformatics **16**(6), 1794–1801 (2018)
11. Gilany, M., Wilson, P., Perera-Ortega, A., Jamzad, A., To, M.N.N., Fooladgar, F., Wodlinger, B., Abolmaesumi, P., Mousavi, P.: Trusformer: improving prostate cancer detection from micro-ultrasound using attention and self-supervision. International Journal of Computer Assisted Radiology and Surgery pp. 1–8 (2023)
12. Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., et al.: Segment anything model for medical images? Medical Image Analysis **92**, 103061 (2024)
13. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
14. Klotz, L., Lughezzani, G., Maffei, D., Sánchez, A., Pereira, J.G., Staerman, F., Cash, H., Luger, F., Lopez, L., Sanchez-Salas, R., et al.: Comparison of micro-ultrasound and multiparametric magnetic resonance imaging for prostate cancer: A multicenter, prospective analysis. Canadian Urological Association Journal **15**(1), E11 (2021)
15. Li, X., Sang, Y., Ma, X., Cai, Y.: Quantitative feature classification for breast ultrasound images using improved naive bayes. IET Image Processing **17**(5), 1417–1426 (2023)
16. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)

17. Oelze, M.L.: Quantitative ultrasound techniques and improvements to diagnostic ultrasonic imaging. In: 2012 IEEE International Ultrasonics Symposium. pp. 232–239. IEEE (2012)
18. Oelze, M.L., Mamou, J.: Review of quantitative ultrasound: Envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound. IEEE transactions on ultrasonics, ferroelectrics, and frequency control **63**(2), 336–351 (2016)
19. Purysko, A.S., Tempany, C., Macura, K.J., Turkbey, B., Rosenkrantz, A.B., Gupta, R.T., Attridge, L., Hernandez, D., Garcia-Tomkins, K., Bhargavan-Chatfield, M., et al.: American college of radiology initiatives on prostate magnetic resonance imaging quality. European Journal of Radiology p. 110937 (2023)
20. Rohrbach, D., Wodlinger, B., Wen, J., Mamou, J., Feleppa, E.: High-frequency quantitative ultrasound for imaging prostate cancer using a novel micro-ultrasound scanner. Ultrasound in medicine & biology **44**(7), 1341–1354 (2018)
21. Sadeghi-Naini, A., Sannachi, L., Tadayyon, H., Tran, W.T., Slodkowska, E., Trudeau, M., Gandhi, S., Pritchard, K., Kolios, M.C., Czarnota, G.J.: Chemotherapy-response monitoring of breast cancer patients using quantitative ultrasound-based intra-tumour heterogeneities. Scientific reports **7**(1), 10352 (2017)
22. Sun, Y.K., Zhou, B.Y., Miao, Y., Shi, Y.L., Xu, S.H., Wu, D.M., Zhang, L., Xu, G., Wu, T.F., Wang, L.F., et al.: Three-dimensional convolutional neural network model to identify clinically significant prostate cancer in transrectal ultrasound videos: a prospective, multi-institutional, diagnostic study. Eclinicalmedicine **60** (2023)
23. Wilson, P.F., Gilany, M., Jamzad, A., Fooladgar, F., To, M.N.N., Wodlinger, B., Abolmaesumi, P., Mousavi, P.: Self-supervised learning with limited labeled data for prostate cancer detection in high frequency ultrasound. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2023)
24. Xing, J., Liu, J., Wang, J., Sun, L., Chen, X., Gu, X., Wang, Y.: A survey of efficient fine-tuning methods for vision-language models-prompt and adapter. Computers & Graphics (2024)
25. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Conditional prompt learning for vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16816–16825 (2022)