

# Preliminary Study on Stromal Nodule Detection Using B-Mode Ultrasound Images and Deep Learning

Izabella Chaverra<sup>a,b</sup>, Nati Nawawithan<sup>a,b</sup>, Teja R. Pathour<sup>a,b</sup>, Ling Ma<sup>a,b</sup>,  
Douglas W. Strand<sup>c</sup>, Xiaosong Meng<sup>c</sup>, Brett A. Johnson<sup>c</sup>, Baowei Fei<sup>a,b,d,\*</sup>

<sup>a</sup> Center for Imaging and Surgical Innovation, University of Texas at Dallas, Richardson, TX

<sup>b</sup> Department of Bioengineering, University of Texas at Dallas, Richardson, TX

<sup>c</sup> Department of Urology, University of Texas Southwestern Medical Center, Dallas, TX

<sup>d</sup> Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX

\* Corresponding author: bfei@utdallas.edu, Website: <https://fei-lab.org>

## ABSTRACT

Benign prostatic hyperplasia (BPH) is a common condition among aging men. Currently, transrectal ultrasonography (TRUS) is used for detecting BPH but frequently falls short in accuracy. In this study, we investigated the use of a deep learning network for stromal nodule detection in prostate specimens exhibiting BPH using B-mode ultrasound (US) imaging. B-mode US data was captured from *ex-vivo* prostate specimens of human patients. Expert annotations differentiating between stromal and non-stromal regions for each prostate specimen were correlated with corresponding B-mode data and used as the reference for validation. The Segment Anything Model 2 (SAM2) network was trained and applied to segment stromal versus non-stromal tissue within B-mode ultrasound images. An 80/20 patient-level train/test split was used to evaluate model performance. Preliminary results demonstrate the feasibility of using SAM2 to identify stromal nodules directly from B-mode ultrasound data, achieving a Dice score of 0.76. Ongoing work focuses on increasing the dataset and improving segmentation accuracy. The study provides an approach to validate the capability of ultrasound imaging for BPH detection. More image data are needed to improve the segmentation performance.

**Keywords:** Ultrasound (US) imaging, prostate, benign prostatic hyperplasia (BPH), deep learning, segmentation

## 1. INTRODUCTION

Benign prostatic hyperplasia (BPH) is highly prevalent and can cause lower urinary tract symptoms that significantly affect quality of life [1-3]. Accurate identification of prostate stromal and non-stromal tissue is essential for distinguishing BPH from prostate cancer in aging men. Although transrectal ultrasound (TRUS) is commonly used for BPH diagnosis, its limited accuracy motivates the development of improved methods for detecting stromal tissue in the prostate [4]. Prostate cancer is a common malignancy that can be life-threatening if detected in the late stage, often progressing slowly and remaining asymptomatic in the early stage [5]. Prostate cancer is typically associated with glandular tissue [5], whereas BPH involves both stromal and glandular hyperplasia [6]. Therefore, distinguishing stromal and glandular regions is critical for accurate diagnosis, understanding BPH progression [2], and supporting personalized treatment planning, ultimately improving clinical decision-making and patient outcomes.

Moreover, accurately detecting stromal regions in the prostate can significantly reduce the time and resources lost to overtreatment. Because prostate cancer often progresses slowly, not all diagnosed cases require immediate or aggressive intervention. Precise segmentation of different tissue types enables clinicians to better assess disease aggressiveness and determine the most appropriate management strategy for each patient, including offering active surveillance for cancers considered low risk [7].

For detection of stromal and glandular nodules in the prostate, the current standard relies on histopathological examination, typically obtained through biopsy or prostatectomy [8,9]. While this method provides highly detailed information for tissue characterization, it is invasive and may be associated with patient discomfort, sampling errors, and procedural risks. TRUS is widely used in clinical practice and remains the most common imaging modality due to its cost-effectiveness and accessibility, despite its limitations in accurately distinguishing BPH from prostate cancer. Consequently, there is a pressing need for a reliable, non-invasive diagnostic method that can improve the accuracy of prostate disease detection and enhance clinical decision-making for conditions such as BPH and prostate cancer.

Recent foundation segmentation models have demonstrated strong generalization across diverse imaging domains, including medical ultrasound imaging. Segment Anything Model 2 (SAM2) [10], a transformer-based and promptable segmentation framework, leverages hierarchical vision transformers and attention mechanisms to capture global context and long-range dependencies, enabling robust segmentation in complex and noisy B-mode ultrasound images [10-12]. Compared to traditional convolutional neural networks, transformer-based models have shown superior performance in preserving structural boundaries and global shape coherence across medical imaging modalities, including ultrasound [13,14]. Given the heterogeneous texture and spatial variability of prostate tissue, SAM2 offers a promising approach for distinguishing stromal from non-stromal regions through integrated global reasoning and localized prompt guidance, with potential to enhance prostate tissue characterization and support more precise diagnostic workflows [15,16].

In this study, we present a prostate tissue segmentation framework using SAM2 applied to B-mode ultrasound images [17]. The model was fine-tuned to segment stromal versus non-stromal regions using bounding box prompts derived from ground truth annotations. By combining ultrasound-specific preprocessing with SAM2's transformer-based architecture, the approach captures both global context and fine structural detail to improve tissue boundary delineation. This method demonstrates the feasibility of using promptable foundation models for ultrasound-based prostate tissue segmentation and highlights the potential of SAM2 as a noninvasive and clinically relevant tool for improved prostate tissue characterization.

## 2. MATERIALS AND METHODS

### 2.1 Ultrasound Imaging Setup

Our US data consists of B-mode images of *ex-vivo* prostate specimens that were captured using the setup in Figure 1. The prostate specimen was placed on a foam stage to minimize backscatter in a transparent water bath filled with deionized (DI) water. The GEM5ScD (Verasonics, Redmond, WA) clinical phased array transducer was paired with the Verasonics Vantage 256 system (Verasonics, Redmond, WA) for the acquisition of the B-mode data. A customized wide beam imaging script reproduced by Verasonics setup script was used that employed 64 angled pulses, facilitating the collection of ultrasound data. A total of 200 frames were captured at a rate of 8 frames per second (fps) for 25 seconds and the transducer moved in the z direction, where z is the breadth of water, x is the width, and y is the height as seen in Figure 1. An automatic stage mover was used to reposition the transducer with a fixed step size and ensure the position of the transducer remained at a fixed position for even data collection. The dataset was comprised of B-mode images for all specimens, with experiments conducted on ten *ex-vivo* prostate specimens. In our study, we decided to use 2D image slices in the x:z direction so that the entire prostate was covered.

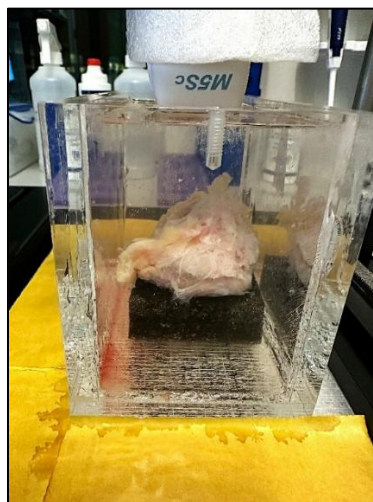


Figure 1. Ultrasound imaging setup for the *ex-vivo* prostate specimens. A clinical grade phased array transducer is positioned above an *ex-vivo* prostate specimen, stabilized on a foam stage within a transparent water tank for optimal ultrasound imaging.

Each specimen featured a single slice exhibiting BPH, characterized by either glandular nodules or a combination of stromal and glandular nodules. The combination of stromal and glandular nodules was named as “stromal” and the slices with only glandular nodules were named as “non-stromal” class in this study.

### 2.2 B-mode Image Pre-processing

B-mode ultrasound images acquired from the Verasonics system were pre-processed to reduce noise, normalize intensity values, and enhance image quality, as seen in Figure 2. Images were cropped to isolate the prostate and remove background water artifacts, followed by contrast and brightness adjustments to improve visualization similar to Pathour *et al.* [18]. Each acquisition was divided into 200 individual frames, which were captured for 25 seconds at 8 fps for analysis. The top-down view (Figure 1), with the transducer positioned above the specimen, was used for all evaluations. Reference annotations were performed to define prostate zones and stromal nodules in each specimen.

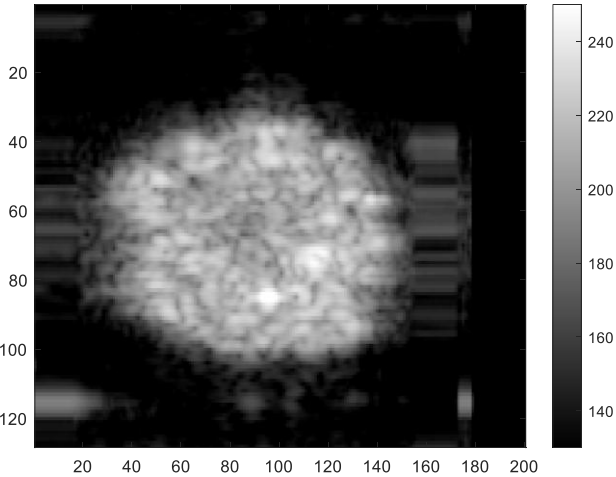


Figure 2. Representative processed B-mode ultrasound frame of a prostate specimen.

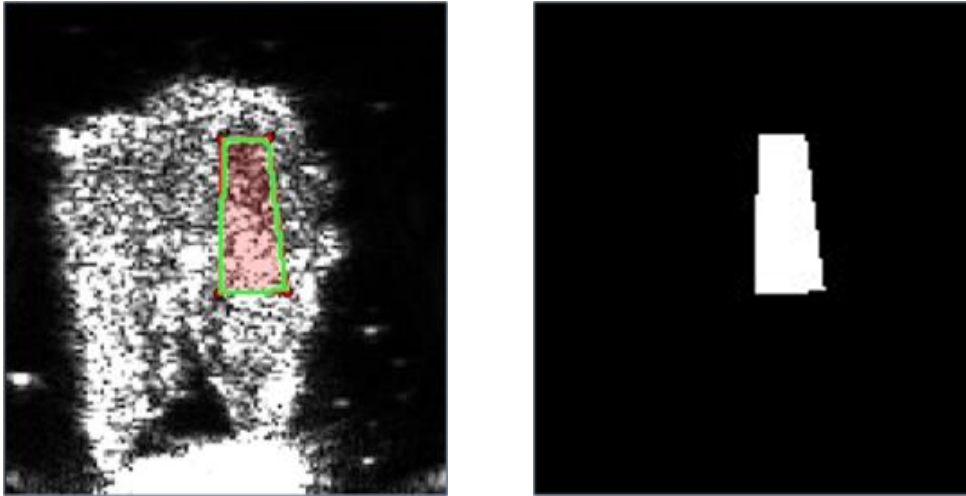


Figure 3. Images of manual mask (right) and corresponding B-mode US data (left). On the mask image (right), the white region is the stromal area of the prostate which was determined by the manual annotation on the B-mode ultrasound image (left).

### 2.3 B-mode Masking

Reference labels were established by manual annotation of prostate zones and stromal areas within the specimens. The transition zone, which commonly contains stromal tissue, is particularly relevant for BPH detection. To generate regions of interest (ROIs) corresponding to stromal areas, the top-view B-mode ultrasound images were spatially aligned and overlaid onto their annotated photograph counterparts. This overlay allowed the stromal annotations drawn on the photograph images to be transferred to the corresponding B-mode frames. Using these aligned image pairs, ROIs were extracted and converted into binary masks in MATLAB, where stromal regions were assigned a value of 1 and the background a value of 0, as shown in Figure 3. These masks were then used as training labels for the SAM2 model.

### 2.4 SAM2 Training and Evaluation

We trained a fine-tuned SAM2 to segment stromal tissue in 845 B-mode ultrasound images using a supervised learning framework, as shown in Figure 4. Image data were split at the subject level into training (80%) and testing (20%) sets to prevent data leakage across individuals. Each dataset sample consisted of paired ultrasound images and binary segmentation masks, with both resized to  $256 \times 256$  pixels. Data augmentation during training included horizontal flipping, while test data underwent resizing only. For each training instance, a bounding box prompt was automatically derived from the reference mask and provided to SAM2 as spatial guidance. The image encoder was frozen to reduce overfitting given the limited dataset size, and optimization was performed using AdamW [19] with a learning rate of  $1 \times 10^{-5}$  and mixed-precision training. The model was trained for 100 epochs with a batch size of 8. Segmentation loss combined binary cross-entropy with Dice loss to balance pixel-wise accuracy and region overlap. Model training performance was monitored using average epoch loss across the training set.

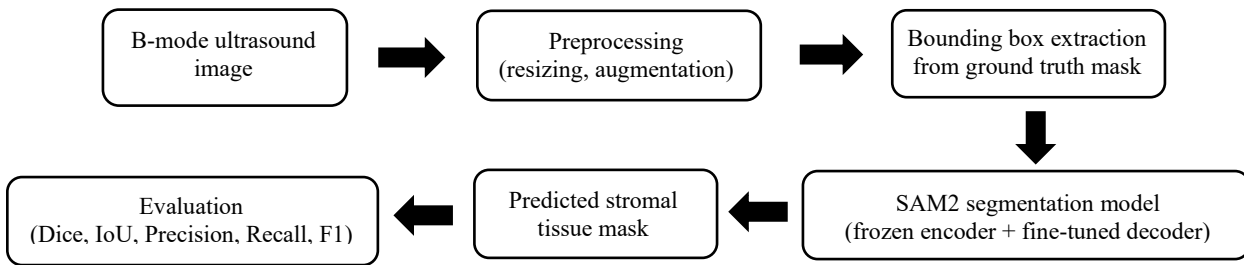


Figure 4. Pipeline for SAM2-based stromal tissue segmentation in B-mode ultrasound images.

## 3. RESULTS

Gross examination of excised prostate tissue, as seen in Figure 5, reveals substantial heterogeneity in stromal composition and spatial distribution, with stromal-rich regions appearing as pale, irregular, and nodular structures embedded within surrounding tissue. This variability highlights the complexity of stromal architecture and underscores the challenge of accurately identifying stromal regions using imaging alone. The nonuniform and localized nature of stromal tissue motivates the use of ROI-based analysis in ultrasound segmentation, as global, full-field evaluation can obscure clinically relevant stromal features due to background dominance and class imbalance.

Qualitative assessment as seen in Figure 6 shows that the SAM2 model reliably localized stromal tissue in B-mode ultrasound images, producing segmentation masks that closely matched expert annotations. The model preserved the overall shape and spatial extent of stromal nodules despite ultrasound-specific challenges such as speckle noise, low contrast, and heterogeneous tissue texture. Errors primarily occurred in regions with ambiguous echogenic boundaries, where predictions occasionally extended into adjacent non-stromal tissue. Overall, these results demonstrate SAM2's ability to extract meaningful tissue features from ultrasound and generate anatomically plausible segmentation outputs.

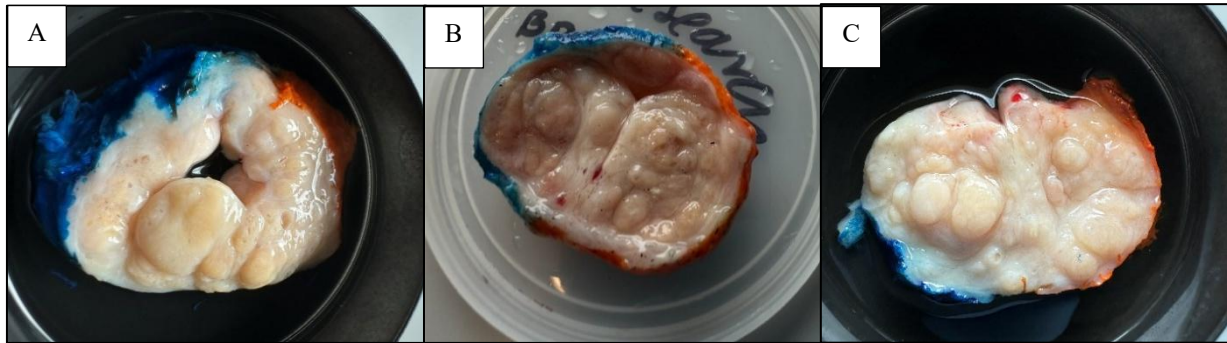


Figure 5. Three representative gross prostate tissue specimens. Pale, nodular stromal-rich regions are visible within excised prostate samples, with heterogeneous texture and irregular spatial distribution across the gland. Colored surgical inks are used to mark the orientation of the specimens.

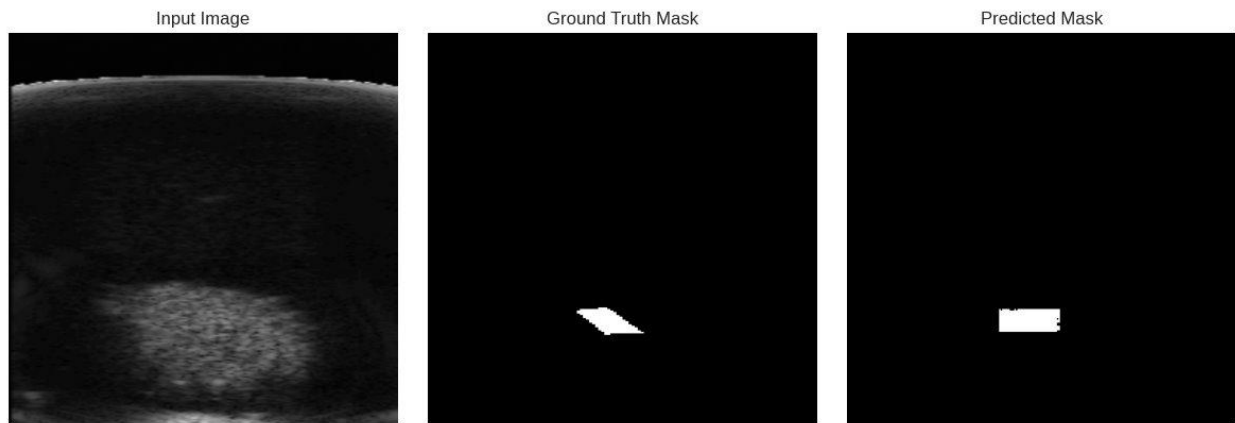


Figure 6. Prostate B-mode ultrasound image (left), reference annotation of stromal area (middle), and SAM2 segmentation (right).

Table 1 shows the performance of the SAM2 model, which was evaluated on a held-out patient-level test set for stromal versus non-stromal tissue segmentation in B-mode ultrasound images. A Dice similarity coefficient of 0.761 and an Intersection over Union (IoU) of 0.622 were achieved between predicted masks and expert annotations. The model achieved a precision of 0.683 and a recall of 0.878, where the high recall value reflects a high sensitivity to stromal tissue regions while maintaining reasonable control over false positive predictions, as indicated by the precision value. The resulting F1 score of 0.761 further confirms a balanced tradeoff between detection sensitivity and segmentation accuracy. For comparison, Dice scores of 0.755 were obtained at both 20 and 50 epochs, while performance slightly decreased to 0.748 at 150 epochs, suggesting diminishing returns beyond 100 epochs. Qualitatively, the SAM2 model produced coherent segmentation masks that aligned well with visually identifiable stromal structures, successfully capturing the shape and spatial extent of stromal regions across heterogeneous specimens. The relatively high recall suggests that the model effectively identifies most stromal tissue regions, though the slightly lower precision indicates a tendency toward mild over-segmentation in challenging or low-contrast regions. Overall, these findings demonstrate that SAM2 can extract stromal tissue features from ultrasound images despite inherent noise, texture variability, and limited dataset size.

Table 1. Performance of SAM2 for stromal tissue segmentation.

| Metric     | Value |
|------------|-------|
| Dice Score | 0.761 |
| IoU        | 0.622 |
| Precision  | 0.683 |
| Recall     | 0.878 |
| F1 Score   | 0.761 |

#### 4. DISCUSSION & CONCLUSION

In this study, we demonstrate the feasibility of using the SAM2 model to segment stromal tissue in B-mode ultrasound images of *ex-vivo* prostate specimens of human patients. Our results indicate that the deep learning model can identify stromal regions consistently with expert annotations, supporting the premise that clinically meaningful tissue feature information may be embedded in ultrasound data. While segmentation performance remains constrained by limited dataset size and specimen heterogeneity, these findings provide a proof-of-concept of using deep learning-based segmentation for stroma detection on ultrasound images.

Clinically, improved stromal tissue identification and segmentation may enhance the assessment of BPH, where stromal hyperplasia is a key pathological feature. A SAM2-based segmentation framework could support more accurate differentiation between stromal and glandular tissue, improve diagnostic confidence, and contribute to more personalized patient management while reducing unnecessary invasive procedures. Nevertheless, this study has several limitations. The small number of *ex-vivo* specimens limited the generalizability. Further validation on a larger, more diverse *in-vivo* datasets is needed. Future work will focus on expanding the dataset, refining prompting strategies, and improving segmentation robustness to facilitate eventual clinical translation.

In conclusion, this study presents a SAM2-based framework for segmenting stromal tissue in prostate B-mode ultrasound images, demonstrating the feasibility of applying deep learning segmentation models to ultrasound-based prostate tissue analysis. By leveraging SAM2's ability to integrate global contextual reasoning with prompt-guided segmentation, our approach enables automated identification of stromal regions associated with benign prostatic hyperplasia. These findings highlight the potential of deep learning-based models to support noninvasive prostate tissue characterization. Continued validation on larger and more diverse datasets will be essential to advance this framework toward real-world diagnostic and decision-support applications.

#### ACKNOWLEDGEMENTS

Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA288379 and the Cancer Prevention and Research Institute of Texas (CPRIT) under Award Number RP240289 and R240542. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

#### REFERENCES

- [1] J. E. McNeal, "Origin and evolution of benign prostatic enlargement," *Invest. Urol.*, 15(4), 340–345 (1978).
- [2] S. J. Berry et al., "The development of human benign prostatic hyperplasia with age," *J. Urol.*, 132(3), 474–479 (1984). doi: 10.1016/s0022-5347(17)49698-4
- [3] C. G. Roehrborn, "Pathology of benign prostatic hyperplasia," *Int. J. Impot. Res.*, 20(3), S11–S18 (2008). doi: 10.1038/ijir.2008.55
- [4] R. A. David et al., "Diagnostic performance of transrectal ultrasound for prostate volume estimation in men with benign prostate hyperplasia," *Int. J. Clin. Pract.*, 74(11), e13615 (2020). doi: 10.1111/ijcp.13615
- [5] J. T. Isaacs, "Prostatic structure and function in relation to the etiology of prostatic cancer," *The Prostate*, 4(4), 351–366 (1983). doi: 10.1002/pros.2990040405

- [6] E. Shapiro et al., “The Relative Proportion of Stromal and Epithelial Hyperplasia is Related to the Development of Symptomatic Benign Prostate Hyperplasia,” *J. Urol.*, 147(5), 1293–1297 (1992). doi: 10.1016/S0022-5347(17)37546-8
- [7] T. J. Wilt *et al.*, “Radical Prostatectomy or Observation for Clinically Localized Prostate Cancer: Extended Follow-up of the Prostate Cancer Intervention Versus Observation Trial (PIVOT),” *Eur. Urol.*, 77(6), 713–724 (2020).
- [8] D. S. S. Surasi, et al., “Imaging and Management of Prostate Cancer,” *Semin. Ultrasound CT MRI*, 41(2), 207–221 (2020). doi: 10.1053/j.sult.2020.02.001
- [9] M. Sayan *et al.*, “Histologically Overt Stromal Response and the Risk of Progression after Radical Prostatectomy for Prostate Cancer,” *Cancers*, 16(10) (2024). doi: 10.3390/cancers16101871
- [10] A. Kirillov et al., “Segment Anything,” *arXiv preprint*, arXiv:2304.02643 (2023).
- [11] N. Ravi, et al., “SAM 2: Segment Anything in Images and Videos,” *arXiv preprint*, arXiv:2408.00714 (2024).
- [12] Dosovitskiy, A. et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale” *arXiv preprint arXiv:2010.11929* (2020).
- [13] A. Hatamizadeh et al., “UNETR: Transformers for 3D Medical Image Segmentation,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 574–584 (2022).
- [14] J. Chen et al., “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,” *arXiv preprint*, arXiv:2102.04306 (2021).
- [15] J. A. Noble et al., “Ultrasound Image Segmentation: A Survey,” *IEEE Trans. Med. Imaging*, 25(8), 987–1010 (2006). doi: 10.1109/TMI.2006.877092
- [16] G. Litjens, et al., “A Survey on Deep Learning in Medical Image Analysis,” *Med. Image Anal.*, 42, 60–88 (2017). doi: 10.1016/j.media.2017.07.005
- [17] J. Ma et al., “Segment Anything in Medical Images,” *arXiv preprint*, arXiv:2304.12306 (2023).
- [18] T. Pathour *et al.*, “Feature Extraction of Ultrasound Radiofrequency Data for the Classification of the Peripheral Zone of Human Prostate,” *Proc. SPIE* 12932, 129321F (2024). doi: 10.1117/12.3008643
- [19] Loshchilov, I. and Hutter, F., “Decoupled Weight Decay Regularization” *arXiv preprint arXiv:1711.05101* (2017).