Prostate Cancer Screening with Artificial Intelligence–Enhanced Micro-Ultrasound: A Comparative Study with Traditional Methods

Muhammad Imran[a], Wayne G. Brisbane[b], Li-Ming Su[c], Jason P. Joseph[c], Wei Shao[a,*]

[a]*Department of Medicine, University of Florida, , Gainesville, 32611, FL, USA*
[b]*Department of Urology, University of California, , Los Angeles, 90095, CA, USA*
[c]*Department of Urology, University of Florida, , Gainesville, 32611, FL, USA*

Abstract

Background and objective: Micro-ultrasound (micro-US) is a novel ultrasound modality with diagnostic accuracy comparable to magnetic resonance imaging (MRI) for detecting clinically significant prostate cancer (csPCa). This study investigated whether interpretation of micro-US by artificial intelligence (AI) can outperform clinical screening methods using prostate-specific antigen (PSA) and digital rectal examination (DRE).

Methods: We retrospectively studied 145 men who underwent micro-US guided biopsy (79 with csPCa and 66 without). A self-supervised convolutional autoencoder was trained to extract deep image features from 2D micro-US slices. Random forest classifiers were developed using five-fold cross-validation to predict csPCa at the slice level. A patient was classified as csPCa-positive if ≥ 8 consecutive slices were predicted positive. Model performance was compared with a classifier trained on common clinical screening variables (PSA, DRE, prostate volume, and age).

Key findings and limitations: The AI-enhanced micro-US model and the clinical screening model achieved AUROCs of 0.871 and 0.753, respectively. Using a fixed classification threshold for both models, the micro-US model achieved a sensitivity of 92.5% and specificity of 68.1%, while the clinical model achieved a sensitivity of 96.2% but with a lower specificity at 27.3%. Limitations of this study include its retrospective single-center design and lack of external validation.

Conclusions and clinical implications: AI-interpreted micro-US significantly improves specificity for csPCa while maintaining high sensitivity. This approach may reduce unnecessary biopsies and offers a low-cost, point-of-care alternative to PSA-based screening. Future prospective studies are needed to validate these findings.

Patient summary: We developed an artificial intelligence system to analyze micro-ultrasound images of the prostate. In this study, it detected aggressive prostate cancer more accurately than traditional screening methods such as PSA blood tests and digital rectal exams. This approach may help reduce unnecessary biopsies in the future.

Keywords: Prostate cancer screening, Micro-ultrasound, Artificial intelligence

1. Introduction

Prostate cancer is one of the most commonly diagnosed malignancies and a leading cause of cancer-related death worldwide (James et al., 2024). Early detection of clinically significant prostate cancer (csPCa) is critical, as it increases the 5-year survival rate from 37% to nearly 100% (Society, 2025). Clinically, screening often relies on prostate-specific antigen (PSA) testing and digital rectal examination (DRE), but both methods have notable limitations. PSA lacks specificity and may be elevated in benign conditions such as benign prostatic hyperplasia or prostatitis. DRE, while inexpensive and easy to perform, suffers from poor sensitivity and high inter-observer variability. As a result, these traditional tools can lead to both overdiagnosis and missed diagnoses, contributing to unnecessary biopsies or delayed detection of aggressive disease.

Multiparametric magnetic resonance imaging (mpMRI) improves the detection of csPCa and is commonly used to guide targeted biopsies (Ahmed et al., 2017). However, its role in routine screening is limited by high cost, long acquisition times, limited availability, and the need for specialized radiological expertise. These constraints make mpMRI impractical for large-scale or point-of-care screening. Micro-ultrasound (micro-US), a high-resolution (29 MHz) imaging modality, provides real-time visualization of prostate microarchitecture with spatial resolution three to four times greater than conventional transrectal ultrasound (Klotz et al., 2020). The OPTIMUM randomized trial confirmed that micro-US–guided biopsy is noninferior to MRI-targeted biopsy for

*Corresponding author.
E-mail address: weishao@ufl.edu (W. Shao).
ORCID: 0000-0003-4931-4839 (W. Shao).

detecting csPCa (Kinnaird et al., 2025). With its portability, lower cost, and suitability for outpatient use, micro-US is well positioned as a potential screening tool. However, interpretation remains challenging and highly operator-dependent (Zhou et al., 2024), limiting consistent performance and widespread adoption.

To address the interpretive limitations of micro-ultrasound, we developed an artificial intelligence (AI) model to automatically detect csPCa from micro-US images. Using a retrospective cohort of 145 men who underwent micro-US–guided prostate biopsy, we trained a self-supervised convolutional autoencoder to extract deep imaging features from 2D micro-US slices. These features were used to train a random forest classifier for slice-level prediction, and patient-level prediction was determined by aggregating predictions across consecutive slices. We compared this model to a classifier trained on standard clinical screening variables including PSA, DRE, prostate volume, and age. To our knowledge, this is the first study to align micro-US imaging with biopsy-confirmed pathology at the slice level and to perform patient-level csPCa screening predictions using AI. This work evaluates the potential of AI-enhanced micro-US as a practical and accurate tool for prostate cancer screening.

2. Patients and methods

2.1. Patient population and data description

This retrospective study was approved by the University of Florida Institutional Review Board and included 145 men who underwent micro-ultrasound (micro-US)–guided prostate biopsy. All patients had clinical indications for biopsy, such as elevated PSA and/or an abnormal digital rectal examination (DRE). Most patients also underwent a systematic 12-core biopsy, including those without visible micro-US lesions. During biopsy, the operator recorded needle trajectories and target locations using micro-US images, enabling retrospective mapping of cores to corresponding regions. All patients provided informed consent. Ground truth for csPCa was established by histopathological analysis of all biopsy cores. Patients were classified as csPCa-positive if any core contained Gleason score $\geq 3+4$. Baseline patient characteristics are summarized in Table 1.

2.1.1. Micro-ultrasound imaging

Pre-biopsy micro-US scans were acquired using a 29 MHz transrectal system (ExactVu, Exact Imaging, Markham, Canada) by an experienced urologist (WGB) with four years of micro-US interpretation experience. Scans were recorded at 10 frames per second for up to 30 seconds, producing 200 to 300 2D micro-US slices per scan.

2.1.2. Clinical biomarkers

For each patient, we collected the following clinical biomarkers: PSA, DRE findings, age, and prostate volume. Prostate volume was estimated from the pre-biopsy micro-US scan. We used the MicroSegNet model (Jiang et al., 2024) to segment the prostate capsule on each 2D slice, and reconstructed the segmentations into a 3D volume using the method described in (Imran et al., 2024). The final prostate gland volume (in mL) was computed from the reconstructed 3D model.

2.1.3. Slice-level labeling for model training

Model training required slice-level annotations indicating the presence or absence of csPCa on each 2D micro-ultrasound (micro-US) slice. For csPCa-negative patients, all slices were labeled as negative. For csPCa-positive patients, operator-recorded needle trajectories were used to cognitively map each biopsy core to the corresponding region on the pre-biopsy micro-US scan. An expert urologist (WGB) manually reviewed slices surrounding these trajectories to assess csPCa extent, using sonographic features defined in the PRI-MUS protocol (Maffei et al., 2024). Slices with suspicious features were labeled as positive. All other slices in csPCa-positive cases were excluded from training, as their cancer status could not be confidently determined in the absence of histopathological confirmation. In total, 2,062 positive and 14,769 negative slices were included. Model evaluation was conducted at the patient level, using biopsy-confirmed csPCa status to reflect clinically meaningful outcomes.

2.2. Model development and evaluation

2.2.1. Micro-US image feature extraction

We trained a convolutional autoencoder with self-supervised learning to extract high-level features from micro-US images. The autoencoder (Figure 1) consists of two symmetric components: a convolutional encoder g_ϕ and a decoder f_ϕ. The encoder compresses the input image $\mathbf{x}$ into a lower-dimensional latent representation $\mathbf{z}$, while the decoder attempts to reconstruct the original image from $\mathbf{z}$. The encoder includes five convolutional layers with increasing channel dimensions (from 3 to 256), interleaved with ReLU activations and strided convolutions for spatial downsampling. The decoder mirrors this structure with transposed convolutions and corresponding upsampling layers to produce the reconstructed image. The autoencoder was trained to minimize the mean squared error between the input image $\mathbf{x}$ and the reconstructed output $\mathbf{x}'$. After training, we used the encoder as a fixed feature extractor. Each 2D micro-US slice was passed through the encoder to generate a feature map, which was then reduced via adaptive average pooling to obtain a 256-dimensional feature vector.

Table 1: Baseline characteristics of our study cohort. Values are medians with interquartile ranges (IQRs) or counts with percentages.

Characteristic	Positive Cases	Negative Cases
Age (yr), median (IQR)	70 (66–74)	69 (63–71)
PSA (ng/ml), median (IQR)	8.2 (5.8–13.1)	5.7 (3.3–7.5)
DRE = 1, n (%)	39 (49.4%)	6 (9.1%)
DRE = 0, n (%)	40 (50.6%)	60 (90.9%)
Prostate volume (ml), median (IQR)	37.5 (31.5–49.4)	47.1 (39.0–55.4)
Total	79	66
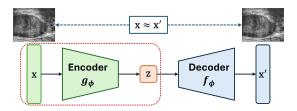
Figure 1: Architecture of the convolutional autoencoder used for feature extraction.

2.2.2. Micro-US image classification

We trained random forest classifiers to classify 2D micro-US slices as csPCa-positive or negative using 256-dimensional feature vectors extracted by the autoencoder's encoder, which captured key imaging characteristics such as texture, intensity, and shape. Slice-level predictions were aggregated to produce patient-level classifications. A patient was considered csPCa-positive if at least eight consecutive slices were predicted positive. This rule was based on retrospective analysis of lesion length, which showed that csPCa typically spanned eight adjacent slices on average. Requiring spatially contiguous predictions helped reduce false positives and improved specificity without compromising sensitivity.

2.2.3. Classification with clinical biomarkers

To assess the predictive value of commonly used screening tools, we trained a random forest model using only non-imaging features: patient age, PSA level, prostate volume, and the binary outcome of DRE. Since the `scikit-learn` implementation supports internal out-of-bag (OOB) validation, a separate validation set was not required for hyperparameter tuning.

2.2.4. Cross-validation strategy

We used five-fold cross-validation to ensure robust and unbiased model evaluation. The dataset, consisting of 79 csPCa-positive and 66 csPCa-negative patients, was partitioned into five mutually exclusive folds. Each patient appeared exactly once in the test set, once in the validation set, and in the training set for the remaining three folds. This approach ensured that all cases contributed to both training and evaluation while preventing data leakage. For each fold, training was monitored on the validation set to prevent overfitting, and the model checkpoint with the lowest validation loss was retained.

2.2.5. Implementation details

All autoencoder models were implemented in `PyTorch (v1.13)` and trained on an NVIDIA A100 GPU using the Adam optimizer (learning rate = 0.001, batch size = 32). After training, the decoder was discarded and the encoder was used as a fixed feature extractor. Each random forest model was trained with 1,000 trees, class-balanced weights, and stratified sampling to preserve the distribution of positive and negative slices. Records with missing clinical values or duplicate patient entries were removed to ensure data integrity.

2.2.6. Peformance metrics

We evaluated model performance at the patient level using the following metrics: area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, precision, and F1-score. All metrics were averaged across the five cross-validation folds. AUROC served as the primary evaluation metric, as it captures overall discriminatory performance across all classification thresholds. The remaining metrics were computed using a fixed probability threshold of 0.15, which was empirically selected to balance sensitivity and specificity in the training data.

3. Results

3.1. ROC-based comparison of imaging and clinical models

Figure 2 shows the ROC curves comparing the performance of the imaging-based model and the clinical biomarker model. The clinical model, trained on age, PSA, DRE, and prostate volume, achieved a mean AUROC of 0.753, indicating moderate discriminative ability. In contrast, the imaging-based model achieved a higher AUROC of 0.871, reflecting a stronger ability to distinguish csPCa from non-csPCa based on deep features extracted from micro-US slices.

3.2. Comparison of classification metrics

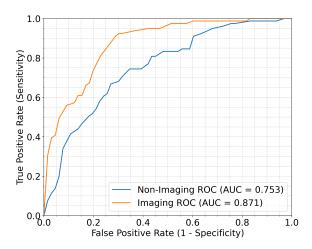Table 2 summarizes the average classification metrics across five cross-validation folds. The clinical

Figure 2: ROC curves comparing the imaging-based model and clinical biomarker-based model. The imaging model achieved a higher AUROC (0.871) than the clinical model (0.753).

model achieved high sensitivity (96.2%) but low specificity (27.3%), resulting in a high false-positive rate. Its precision (61.4%), F1-score (74.9%), and accuracy (64.8%) were moderate. In contrast, the imaging model maintained high sensitivity (92.5%) while substantially improving specificity (68.1%). It also achieved higher precision (77.8%), F1-score (84.5%), and accuracy (81.4%), demonstrating a better overall balance between true-positive detection and false-positive control.

Table 2: Threshold-based classification metrics (averaged over five folds) using a fixed decision threshold of 0.15.

Model	Sensitivity	Specificity	Accuracy	Precision	F1-Score
Clinical	**96.2%**	27.3%	64.8%	61.4%	74.9%
Micro-US	92.5%	**68.1%**	**81.4%**	**77.8%**	**84.5%**

These results highlight that the AI-based imaging model offers a more favorable trade-off between sensitivity and specificity, reducing unnecessary false positives while maintaining robust detection of clinically significant disease. This supports its potential as a more accurate and efficient screening tool for csPCa.

4. Discussion

This study demonstrates that an AI-enhanced micro-US model can significantly improve prostate cancer screening performance compared to traditional biomarker-based approaches. While the clinical model, which incorporated PSA, DRE, prostate volume, and age, achieved high sensitivity (96.2%), it exhibited poor specificity (27.3%), consistent with well-documented limitations of PSA-based screening (Thompson et al., 2004; Schröder et al., 1998). In contrast, the imaging-based model maintained high sensitivity (92.5%) while substantially improving specificity (68.1%), resulting in better overall accuracy and precision. This improved balance is particularly important in a screening context, where reducing false positives can lower the burden of unnecessary biopsies, overtreatment, and patient anxiety.

Micro-US has been proposed as a lower-cost, point-of-care alternative to multiparametric MRI for prostate cancer detection (Klotz et al., 2020; Lughezzani et al., 2019). Prior studies, including the OPTIMUM randomized trial, have shown that micro-US is non-inferior to MRI for csPCa detection in biopsy settings (Kinnaird et al., 2025). However, its broader use in screening has been limited by high inter-operator variability and a steep learning curve (Zhou et al., 2024). Our results suggest that artificial intelligence can help overcome these barriers by enabling consistent, objective interpretation of micro-US images. The self-supervised autoencoder used in this study learned imaging features correlated with csPCa, which were then aggregated using a slice-level prediction framework to produce patient-level classifications. This approach effectively eliminates reliance on subjective interpretation and may help standardize micro-US for widespread clinical adoption.

Previous models, such as TRUSformer (Gilany et al., 2023) and TRUSWorthy (Harmanani et al., 2025), have focused on patch-level classification using weak labels, limiting their clinical interpretability. In contrast, our model aligned individual micro-US slices with biopsy-confirmed pathology and used an empirically chosen aggregation rule to generate patient-level predictions. This method mirrors clinical decision-making and supports actionable, real-time screening decisions. The improved diagnostic performance, particularly in specificity and F1-score, underscores the potential of AI-enhanced micro-US to function as a frontline screening modality that complements or even outperforms existing PSA- and DRE-based strategies.

Despite encouraging results, this study has limitations. The cohort was drawn from a single academic center and consisted of patients already referred for biopsy, which may introduce selection bias. As such, performance in general screening populations remains to be validated. While five-fold cross-validation was used to minimize overfitting, external validation on independent cohorts is necessary to assess generalizability. Additionally, the threshold for patient-level classification (eight consecutive positive slices) was empirically defined and may require adjustment in future prospective settings. Future work should include multicenter clinical trials, decision-curve analysis, and cost-effectiveness studies to evaluate the real-world impact of integrating AI-enhanced micro-US into routine prostate cancer screening. If validated, this approach could provide a scalable, affordable, and interpretable solution for early detection of clinically significant prostate cancer, bridging the gap between low-specificity biomarker screening and high-cost MRI-based diagnostics.

5. Conclusions

This study indicates that AI-augmented micro-ultrasound can outperform traditional PSA- and DRE-based methods for screening clinically significant prostate cancer. These results highlight the potential of micro-US, when interpreted by AI, to serve as a point-of-care screening tool that detects clinically significant cancers more accurately and reduces unnecessary biopsies. If validated in prospective multi-center settings, AI-enhanced micro-US could transform early prostate cancer detection by enabling more precise, accessible, and cost-effective screening, ultimately improving patient outcomes while minimizing harm.

References

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., 2017. Diagnostic accuracy of multi-parametric mri and trus biopsy in prostate cancer (promis): a paired validating confirmatory study. The Lancet 389, 815–822.

Gilany, M., Wilson, P., Perera-Ortega, A., Jamzad, A., To, M.N.N., Fooladgar, F., Wodlinger, B., Abolmaesumi, P., Mousavi, P., 2023. Trusformer: improving prostate cancer detection from micro-ultrasound using attention and self-supervision. International Journal of Computer Assisted Radiology and Surgery 18, 1193–1200.

Harmanani, M., Wilson, P.F., To, M.N.N., Gilany, M., Jamzad, A., Fooladgar, F., Wodlinger, B., Abolmaesumi, P., Mousavi, P., 2025. Trusworthy: toward clinically applicable deep learning for confident detection of prostate cancer in micro-ultrasound. International Journal of Computer Assisted Radiology and Surgery , 1–9.

Imran, M., Nguyen, B., Pensa, J., Falzarano, S.M., Sisk, A.E., Liang, M., DiBianco, J.M., Su, L.M., Zhou, Y., Joseph, J.P., et al., 2024. Image registration of in vivo micro-ultrasound and ex vivo pseudo-whole mount histopathology images of the prostate: A proof-of-concept study. Biomedical Signal Processing and Control 96, 106657.

James, N.D., Tannock, I., N'Dow, J., Feng, F., Gillessen, S., Ali, S.A., Trujillo, B., Al-Lazikani, B., Attard, G., Bray, F., et al., 2024. The lancet commission on prostate cancer: planning for the surge in cases. The Lancet 403, 1683–1722.

Jiang, H., Imran, M., Muralidharan, P., Patel, A., Pensa, J., Liang, M., Benidir, T., Grajo, J.R., Joseph, J.P., Terry, R., et al., 2024. Microsegnet: A deep learning approach for prostate segmentation on micro-ultrasound images. Computerized Medical Imaging and Graphics 112, 102326.

Kinnaird, A., Luger, F., Cash, H., Ghai, S., Urdaneta-Salegui, L.F., Pavlovich, C.P., Brito, J., Shore, N.D., Struck, J.P., Schostak, M., et al., 2025. Microultrasonography-guided vs mri-guided biopsy for prostate cancer diagnosis: The optimum randomized clinical trial. JAMA .

Klotz, L., Lughezzani, G., Maffei, D., Sánchez, A., Pereira, J.G., Staerman, F., Cash, H., Luger, F., Lopez, L., Sanchez-Salas, R., et al., 2020. Comparison of micro-ultrasound and multiparametric magnetic resonance imaging for prostate cancer: A multicenter, prospective analysis. Canadian Urological Association Journal 15, E11.

Lughezzani, G., Saita, A., Lazzeri, M., Paciotti, M., Maffei, D., Lista, G., Hurle, R., Buffi, N.M., Guazzoni, G., Casale, P., 2019. Comparison of the diagnostic accuracy of micro-ultrasound and magnetic resonance imaging/ultrasound fusion targeted biopsies for the diagnosis of clinically significant prostate cancer. European urology oncology 2, 329–332.

Maffei, D., Avolio, P.P., Moretto, S., Piccolini, A., Aljoulani, M., Dagnino, F., De Carne, F., Fasulo, V., Marco, P., Saita, A.R., et al., 2024. Mp49-15 evaluating the role of pri-mus protocol in identifying clinically significant prostate cancer: A high-volume experience on microultrasound. Journal of Urology 211, e788.

Schröder, F.H., Kruger, A.B., Rietbergen, J., Kranse, R., Maas, P.v.d., Beemsterboer, P., Hoedemaeker, R., 1998. Evaluation of the digital rectal examination as a screening test for prostate cancer. Journal of the National Cancer Institute 90, 1817–1823.

Society, A.C., 2025. Cancer facts & figures 2025. atlanta: American cancer society; 2025.

Thompson, I.M., Pauler, D.K., Goodman, P.J., Tangen, C.M., Lucia, M.S., Parnes, H.L., Minasian, L.M., Ford, L.G., Lippman, S.M., Crawford, E.D., et al., 2004. Prevalence of prostate cancer among men with a prostate-specific antigen level ≤ 4.0 ng per milliliter. New England Journal of Medicine 350, 2239–2246.

Zhou, S.R., Choi, M.H., Vesal, S., Kinnaird, A., Brisbane, W.G., Lughezzani, G., Maffei, D., Fasulo, V., Albers, P., Zhang, L., et al., 2024. Inter-reader agreement for prostate cancer detection using micro-ultrasound: a multi-institutional study. European Urology Open Science 66, 93–100.