This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Automated Micro-Ultrasound Prostate Volume for MRI-Free PSA-Density: Reproducibility and Agreement with MRI

Nektarios Winter[1,2,3], Balint Kovacs[1,2], Maurin Helen Mangold[1,3,4], Maurice Stephan Michel[4], Caelan Max Haney-Aubert[1,3,4], and Karl-Friedrich Kowalewski[1,3,4]

[1] Division of Intelligent Systems and Robotics in Urology (ISRU), German Cancer Research Center (DKFZ), Heidelberg, Germany
[2] Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
[3] DKFZ Hector Cancer Institute at the University Medical Center Mannheim, Mannheim, Germany
[4] Department of Urology and Urosurgery, University Medical Center Mannheim, Medical Faculty Mannheim at Heidelberg University, Mannheim, Germany

Abstract. Prostate-specific antigen (PSA) density is a key determinant in biopsy decision-making, but its calculation usually depends on MRI-based prostate volumetry, which is both reader-dependent and costly. Micro-ultrasound (microUS) provides high-resolution imaging of the entire prostate, enabling bedside gland volumetry and offering a practical MRI-free alternative for PSA-density assessment prior to biopsy. However, free-hand microUS sweeps are uncalibrated through-plane, meaning frame counts reflect hand speed rather than true anatomical length. In this study of 455 men, we introduce an automated 2D nnU-Net pipeline with two key design choices to overcome this: (1) using the mean cross-sectional area rather than the sum to bypass hand-speed variance, and (2) applying a geometric filter to remove truncated or irregular sweeps. This mean-area estimator improved volume agreement with MRI compared to the naive sum (Concordance Correlation Coefficient (CCC) 0.83 vs. 0.58, $p < 0.001$). The geometric filter further tightened within-patient agreement and abstained on large glands with field-of-view truncation. Clinically, the microUS-derived PSA-density agreed with MRI in 85% of cases, matching standard MRI inter-reader variability. In predicting clinically significant prostate cancer, the microUS volume was equivalent to MRI within a pre-specified ± 0.05 AUROC margin (0.76 vs. 0.75; two one-sided tests (TOST) $p = 0.002$). We demonstrate that automated microUS volumetry is a reproducible, mathematically robust, MRI-free alternative for risk stratification.

Keywords: Micro-ultrasound · Prostate volumetry · Freehand ultrasound · PSA density · Point-of-care End-to-End pipeline

1 Introduction

Prostate-specific antigen (PSA) density refines the biopsy decision by adjusting PSA for gland size, with a threshold of $\sim$0.15 ng/mL/cc commonly triggering biopsy [1]. Volume is conventionally measured via mpMRI, which adds cost and is subject to inter-reader variability [2, 3]. However, high-resolution micro-ultrasound (microUS) has emerged as a viable alternative, demonstrating non-inferiority to mpMRI for detecting clinically significant prostate cancer in the OPTIMUM trial [4]. Because microUS is increasingly utilized in biopsy pathways and can potentially replace mpMRI in selected patients [4, 5], automating volumetry from microUS sweeps could provide a practical, bedside, MRI-free PSA-density measurement.

However, extracting volumes from 2D free-hand sweeps presents a structural challenge: the through-plane acquisition is uncalibrated. The number of frames depends on the physician's hand speed, making a standard 3D integral (sum of areas) highly susceptible to acquisition noise. We propose a pipeline that frames volume estimation as a linear regression of the *mean* cross-sectional area, fundamentally decoupling the volume prediction from frame count.

Rather than claiming microUS represents perfect anatomical truth (which requires ex-vivo specimens), we evaluate whether this technical approach yields a reproducible MRI-calibrated volume proxy with clinically comparable PSA-density classification to MRI-derived volume at the decision threshold.

2 Methods

Patients. As part of routine clinical practice at a single center (Jan 2024–Jun 2026), 1,430 consecutive patients were evaluated with microUS. Of these, 784 underwent a transperineal pre-biopsy evaluation. After excluding cases without usable microUS DICOMs and those lacking an independent mpMRI volume (to prevent carried-forward bias), 455 formed the agreement cohort; of these, 450 with complete clinical data and determinable csPCa formed the risk cohort (Table 1).

Table 1. Baseline characteristics ($N = 450$), by significant-cancer status. Median (IQR). Gland volume differs markedly by cancer status, motivating its inclusion in the risk model.

Characteristic	csPCa ($n = 280$)	Non-csPCa ($n = 170$)	p
Age (years)	68 (63–74)	65 (60–70)	< 0.001
PSA (ng/mL)	6.8 (5.2–10.3)	7.4 (5.3–11.4)	0.41
DRE suspicious (%)	25	8	< 0.001
MRI gland volume (mL)	43 (33–56)	61 (42–79)	< 0.001

microUS Volume Estimator and Sweep Filtering. The prostate capsule was segmented per frame using a 2D nnU-Net ensemble. The model was trained via 5-fold cross-validation on a curated dataset of 1,549 micro-ultrasound slices from 74 independent patients, specifically excluding ambiguous out-of-field frames to ensure boundary precision. This production-quality model achieved a mean out-of-fold Dice score of 0.943 ($\pm$0.004), comparing favorably to recent dedicated micro-ultrasound segmentation models such as MicroSegNet (Dice 0.939) [6].

Because the through-plane spacing of a free-hand sweep is uncalibrated, with the frame count reflecting hand speed rather than anatomy, a geometric area-integral does not yield a physically calibrated volume without reliable through-plane spacing. We instead use the **mean cross-sectional area $\bar{A}$ as a single linear predictor** of gland volume ($\hat{V} = a\bar{A} + b$), fit out-of-fold against MRI. We compared this to a naive sum-area (ΣA) approach.

Before pooling, a predefined eight-stage geometric filter assessed sweep quality, discarding sweeps with field-of-view truncation, lateral off-centring, or irregular motion (flagged by high frame-to-frame variance). Rejected sweeps generally represent large glands cropped by the sector edge.

Evaluation. We evaluated the technical impact of the mean-area estimator and geometric filtering via the Concordance Correlation Coefficient (CCC) and Mean Absolute Percentage Error (MAPE) against MRI. Clinical interchangeability was assessed by PSA-density reclassification rates at the 0.15 ng/mL/cc threshold. A logistic risk model for csPCa (Grade Group $\geq$ 2) using age, PSA, DRE, and volume was evaluated via repeated cross-validation AUROC. Splitting was performed at the patient level, with all sweeps from an individual assigned to the same fold. The segmentation-training cohort was disjoint from the evaluation cohorts, and the volume-calibration parameters were fitted using training-fold patients only before application to held-out patients. For the risk model, the microUS volume was computed from all acquired sweeps (unfiltered), so every patient retained a value and no case was excluded ($n = 450$). Sweep filtering was applied only to the technical volume-agreement analysis, where it served as a geometric adequacy check; the unfiltered risk analysis was retained to evaluate clinical performance without selective patient exclusion. Difference from the MRI model was tested by DeLong, and equivalence by two one-sided tests (TOST) against a pre-specified margin of $\pm$0.05 AUROC.

3 Results

Estimator Ablation and Design. Table 2 presents a comprehensive ablation of volume estimators. As established, a naive Σ-area estimator yields a baseline CCC of 0.58, making normalization necessary. Comparing normalized per-frame statistics, our linear mean-area estimator significantly outperformed the maximum (peak) area (CCC 0.83 vs. 0.80, $p = 0.012$) and performed comparably

to the median area ($p = 0.058$). Non-linear calibrations, specifically the power-law and isotonic estimators, yielded a higher aggregate CCC ($p = 0.008$ and $p = 0.037$, respectively).

Sweep Cleaning and Abstention. The geometric filter retained a median of 4 valid sweeps per patient and produced a usable volume in 94% of patients (428/455). Within these 428 patients, removing irregular sweeps improved within-patient agreement (CCC 0.77 to 0.83, $p < 0.001$; MAPE 18.5% to 17.0%, $p = 0.044$). The filter abstained from volume-agreement analysis in 27 patients (6%) whose glands exceeded the field of view (median MRI volume 72 mL vs. 47 mL for retained, $p = 0.005$). The automated microUS volume on the remaining sweeps was reproducible (within-patient ICC 0.95, CV 8%).

Agreement with MRI and Size Class Balance. Performance across the prostate size spectrum varied among estimators (Table 3). Non-linear calibration models (e.g., power-law) achieved a higher pooled CCC by reducing error in $>$ 80 mL outliers (13% MAPE), but they did so unevenly across size groups: their error rose sharply in the small-gland cohort ($<$ 30 mL), reaching 39% MAPE. Mean cross-sectional area is not intended to uniquely recover three-dimensional gland anatomy; rather, it provides a frame-count-invariant sweep summary. By contrast, the linear mean-area estimator maintained a balanced error profile across all size classes, achieving the lowest error in the $<$ 30 mL cohort (28% MAPE).

Agreement at the Biopsy Decision. microUS- and MRI-derived PSA-density agreed in 85% of men. At the 0.15 ng/mL/cc decision threshold, microUS reclassified 15% of patients. This rate aligns with published inter-reader MRI variability using the standard ellipsoid method (17.5% [7]).

Table 2. Estimator-design ablation. The linear mean-area estimator is compared against alternative per-frame statistics (median, max) and non-linear calibration models (power-law, isotonic). Non-linear fits yield a higher aggregate CCC but alter the linear calibration.

Estimator	CCC	MAPE	LoA (±mL)	ΔCCC vs ref (95% CI)	p
Mean-area, linear (ref)	0.83	17%	28	reference	—
Median	0.81	19%	29	+0.013 (−0.00, +0.03)	0.058
Max (peak)	0.80	17%	30	+0.023 (+0.00, +0.04)	0.012
Power-law	0.85	16%	27	−0.020 (−0.03, −0.01)	0.008
Isotonic	0.85	16%	27	−0.018 (−0.03, −0.00)	0.037

Table 3. Mean Absolute Percentage Error (MAPE) stratified by prostate size class. Non-linear models (power-law, isotonic) exhibit increased error in the small-gland cohort (< 30 mL), whereas the linear estimator maintains a balanced profile.

Estimator	< 30 mL	30–50 mL	50–80 mL	> 80 mL
Mean-area, linear (ours, ref)	**28%**	**17%**	**13%**	**15%**
Median	29%	18%	15%	17%
Max (peak)	31%	18%	13%	18%
Power-law	39%	13%	15%	13%
Isotonic	35%	14%	14%	16%

Risk Stratification Models. In predicting csPCa, microUS performed comparably to MRI (Table 4). Adding the microUS volume raised the baseline clinical AUROC from 0.66 to 0.76, while MRI raised it to 0.75. The two volumes met the pre-specified ±0.05 AUROC equivalence margin ($\Delta = +0.011$, 95% CI -0.016 to $+0.038$; TOST $p = 0.002$; DeLong $p = 0.42$ for the difference). Restricting the risk feature to patients with ≥ 1 quality-passed sweep ($n = 423$) yielded a similar result (AUROC 0.77). Because $\hat{V}$ is a linear transformation of $\bar{A}$, the classifier absorbs the calibration coefficients, resulting in an identical AUROC whether using mL units or raw mean-area. Combining both volumes yielded an AUROC of 0.77 ($p = 0.07$).

4 Discussion

The core technical finding of this study is that relying on mean cross-sectional area rather than a traditional volume integral counteracts the variable speed of free-hand acquisitions. Coupled with geometric filtering, this approach yields a highly stable metric (ICC 0.95). This formulation does not explicitly model prostate length, local morphology, or temporal/geometric relationships between consecutive frames. Consequently, anatomically different glands with similar mean cross-sectional area may receive similar estimates; the estimator should

Table 4. Clinical risk models for csPCa ($n = 450$). MicroUS matches MRI within a pre-specified ±0.05 AUROC equivalence margin (TOST $p = 0.002$; DeLong $p = 0.42$). The risk feature uses the unfiltered volume, so all 450 patients are retained.

Model Features	AUROC (95% CI)
Clinical only (age, PSA, DRE)	0.66 (0.60–0.71)
Clinical + MRI volume	0.75 (0.71–0.80)
Clinical + microUS (Calibrated Mean-Area)	**0.76** (0.72–0.81)
Clinical + microUS (Naive Σ-Area)	0.75 (0.70–0.79)
Clinical + MRI + microUS (Mean-Area)	0.77 (0.72–0.81)

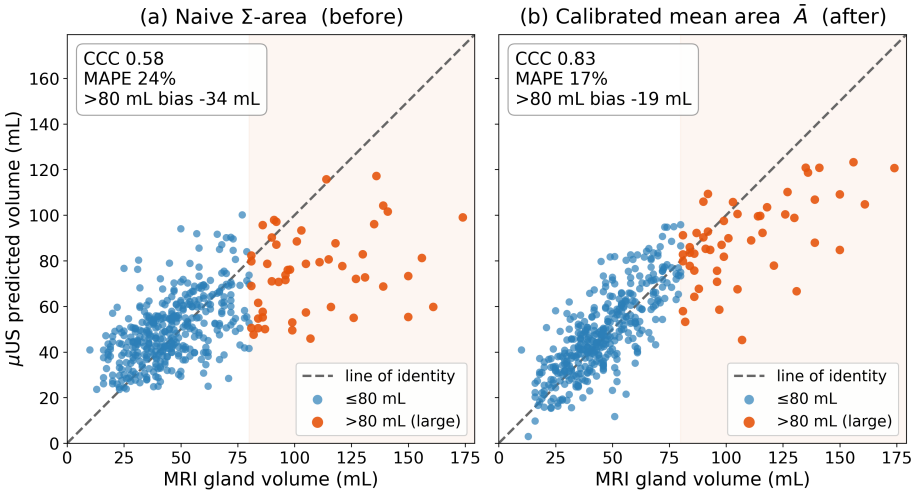

Fig. 1. Scatter plots comparing microUS volumetry methods against the MRI reference standard. (a) The naive Σ-area exhibits a size-dependent under-read in large prostates (highlighted red zone, $> 80\,\mathrm{mL}$), as the cross-section outgrows the field-of-view. (b) The calibrated mean-area ($\bar{A}$) maintains a linear relationship with MRI across all gland sizes, tightening overall agreement (CCC 0.58 vs. 0.83) and reducing the large-gland bias.

therefore be interpreted as an MRI-calibrated volume proxy rather than a direct three-dimensional anatomical reconstruction.

Our ablation of volume estimators demonstrates that while non-linear calibration models achieve a statistically higher pooled CCC, they do so at the expense of small-gland accuracy. Clinically, small prostates are where the PSA-density calculation is most volatile and decision-sensitive. We therefore favor the linear mean-area estimator; it provides a balanced operationalization across the clinical range, protects the small-gland cohort (28% MAPE vs. 39% for a power-law fit), and ensures the volume feature remains a strict linear transformation of the raw mean-area for downstream risk models. Furthermore, transrectal probes inherently compress the gland during acquisition, a physical deformation that our linear calibration model implicitly absorbs when fitting against the uncompressed MRI reference.

In clinical modeling, microUS volume provided risk stratification statistically equivalent to mpMRI within a pre-specified ± 0.05 AUROC margin. It reclassifies only 15% of patients at the critical PSA-density threshold, matching the inherent noise of human MRI readers, and provides comparable csPCa discrimination. The primary limitation remains the absence of ex-vivo specimen volumes to confirm absolute anatomical accuracy [8], and a slight abstention bias against very large (often benign) glands. In addition, calibration to MRI-derived volume was performed at a single center, and its transportability across scanners, institutions, and acquisition protocols requires external validation. The present 2D

pipeline also does not exploit sequential information within the free-hand sweep; incorporating temporal or geometric sweep modeling is a direction for future work. The equivalence is also bounded by power: at $n = 450$ the study can only exclude AUROC differences ≥ 0.04, so a small true advantage of either modality cannot be ruled out.

5 Conclusion

This study demonstrates that an automated, 2D pipeline can extract a reproducible prostate volume from uncalibrated free-hand microUS sweeps. As a bedside tool, automated microUS volumetry offers a robust, MRI-free pathway to accurate PSA-density profiling.

References

1. Mottet, N. et al.: EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancer. European Association of Urology (2024)
2. Thimansson, E. et al.: High inter-reader agreement of prostate volume measurements with the ellipsoid formula on MRI. Eur. Radiol. 33, 5606–5613 (2023)
3. Wallström, J. et al.: National quality-assessment survey of prostate MRI reading: agreement and diagnostic accuracy among 43 radiologists. Eur. J. Radiol. Open 12, 100552 (2024)
4. Klotz, L. et al.: Microultrasonography-Guided vs MRI-Guided Biopsy for Prostate Cancer Diagnosis. JAMA (2025)
5. Imran, H. et al.: Artificial-intelligence micro-ultrasound for clinically significant prostate cancer detection. BJUI Compass (2026)
6. Ghosh, S. et al.: MicroSegNet: A deep learning approach for prostate segmentation on micro-ultrasound images. Comput Methods Programs Biomed. 248, 107851 (2024)
7. Hamzaoui, D. et al.: Prostate volume measurement on MRI: comparison of ellipsoid formula and planimetry in prediction of clinically significant prostate cancer. Eur. Radiol. 32, 4522–4530 (2022)
8. Baco, E. et al.: Prostate volume by transrectal ultrasound and multiparametric MRI compared to true volume in prostatectomy specimens. BMC Urol. 23, 70 (2023)