Original Article

Evaluation of AI and radiologist contouring in prostate MRI for targeted MRI/US fusion biopsy

Danai Manorom

National Cancer Institute, Bangkok, Thailand

Keywords:
AI contouring prostate gland, MRI fusion biopsy of prostate, MRI prostate

Abstract

Objective: Prostate cancer is an increasingly prevalent public health issue, particularly in aging populations such as Thailand. While traditional diagnostic methods like systematic transrectal ultrasound-guided biopsy are widely used, they can result in overdiagnosis and unnecessary treatment. MRI/Ultrasound (MRI/US) Fusion Biopsy offers greater precision by targeting suspicious areas detected in MRI scans. However, manual contouring of the prostate and lesion locations by radiologists or urologists is time-consuming and subject to variability, potentially delaying diagnosis and treatment.

Materials and Methods: This retrospective study developed and evaluated an AI-based prostate segmentation model using 125 annotated prostate MRI cases (3,193 images) from a public dataset for training, and then it was tested on 109 clinical cases (2,952 images) from the National Cancer Institute. The model combined a YOLO-based bounding box detection with the segment anything model (SAM) for prostate segmentation. Model performance was compared to radiologist-drawn contours using dice similarity coefficient (DSC) and % relative percent difference (RPD) in prostate volume estimation.

Results: For cases not requiring post-processing, the AI model achieved a mean DSC of 0.72 and an RPD of 8.90% in comparison to radiologist contours. For cases requiring post-processing, the DSC dropped to 0.66 and the RPD increased to 13.45%. These results indicate a high level of agreement between the AI and expert annotations, particularly in standard cases.

Conclusion: The AI-based model demonstrated promising accuracy with regard to segmentation of the prostate gland on MRI scans, comparable to radiologist performance. This approach has the potential to reduce diagnostic delays and lessen the workload of radiologists in prostate cancer workflows. Future improvements should focus on enhancing model precision, incorporating prostate imaging-reporting and data system (PI-RADS) scoring, and validating the system across diverse clinical settings to support safe and effective integration into routine diagnostic practice.

Insight Urol 2025;46(2):104-11. doi: 10.52786/isu.a.111

Corresponding author: Danai Manorom
Address: National Cancer Institute, Bangkok 10400, Thailand
E-mail: danaimanorom@gmail.com
Manuscript received: March 26, 2025
Revision received: October 15, 2025
Accepted after revision: November 21, 2025

Introduction

Cancer remains a major public health issue in many countries around the world, with a steadily increasing incidence rate. This rise can be attributed to several factors, including the aging global population, population growth, and notably, lifestyle changes among Asians, particularly dietary habits that are becoming more Westernized. According to statistics, prostate cancer has been the most common cancer in men in the United States for many years. In Thailand, the country is transitioning into an aging society, and age is one of the most significant risk factors for prostate cancer hence the incidence of prostate cancer is increasing.

The primary goal of cancer treatment is to achieve a cure while minimizing treatment-related side effects. It is therefore essential to have up-to-date knowledge of treatments and surgical techniques that are both appropriate and in alignment with current standards. According to the publication Cancer in Thailand 2019-2021[1], prostate cancer ranks as the fourth most common cancer among Thai males, with an incidence rate of 8.7 per 100,000 population. This rate continues to increase over time.

Traditionally, the diagnosis of prostate cancer has relied on systemic transrectal ultrasound-guided biopsy (TRUS biopsy), which is widely accepted. However, this approach may lead to overdiagnosis and overtreatment, along with the risk of biopsy-related complications. In recent years, MRI of the prostate has been increasingly used prior to biopsy. If suspicious lesions are detected, a targeted biopsy using magnetic resonance imaging-ultrasound fusion-guided prostate biopsy (MRI/US fusion biopsy) can be performed. This technique uses three-dimensional imaging in conjunction with real-time ultrasound, enabling physicians to clearly visualize and localize suspicious areas within the prostate, allowing for precise, targeted biopsies rather than random sampling.

This MRI/US fusion technique helps reduce unnecessary biopsies and minimizes biopsy-related complications, especially when MRI findings suggest a low risk of prostate cancer.

In Thailand, MRI/US fusion-guided prostate biopsy has become increasingly widespread. However, the procedure requires delineation (segmentation) of the prostate gland and identification of suspicious lesions to ensure accurate fusion with ultrasound images during biopsy. This task is typically performed by radiologists or urologists. In routine clinical practice, patients who show abnormal findings from prostate-specific antigen (PSA) testing or digital rectal examination are referred for an MRI of the prostate. The use of this technique has reduced the incidence of patients who will need to proceed to biopsy. Instead, lesions are reported using the prostate imaging-reporting and data system (PI-RADS), graded from 1 to 5. If a patient has a PI-RADS score of 3-5 or if malignancy is suspected, urologists usually advise patients to return to a radiologist for prostate and lesion contouring prior to fusion biopsy. In some hospitals, this contouring is done by the urologists themselves (Fig. 1).

The process of contouring the prostate typically takes about 15-20 minutes. Having the patient return to the radiologist for this step often causes delays in the biopsy workflow.

The implementation of artificial intelligence (AI) for automatic prostate segmentation has the potential to reduce the time required for prostate biopsy procedures and alleviate the workload of radiologists and urologists involved in manual prostate contouring. This advancement may expedite diagnosis and treatment for patients. In recent years, there has been growing interest in research into the application of AI technologies to this domain.

For instance, Ghafoor et al. in 2023[2] conducted a study comparing the delineation of index lesions on prostate MRI between radiologists and urologists collaborating in MRI/US Fusion Prostate Biopsy procedures. Their findings indicated

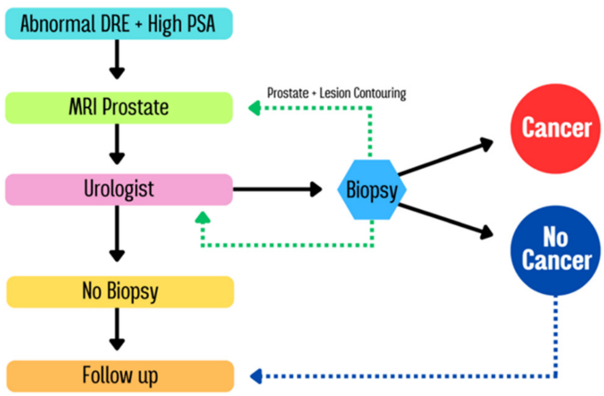

Figure 1. Workflow of prostate cancer diagnostic evaluation using prostate MRI

substantial inter-observer variability in lesion contouring between the two specialties, potentially reducing the accuracy of targeted biopsies. Schelb et al. in 2021[3] evaluated the consistency and accuracy of lesion delineation among multiple radiologists versus a deep learning-based AI model trained for automatic segmentation. The study showed that AI-based segmentation could improve the precision of lesion localization on prostate MRI. Similarly, Nachbar et al. in 2020[4] investigated the use of AI for prostate lesion contouring on MRI for online adaptive radiotherapy planning. The study demonstrated that AI-based contouring is a promising tool for radiotherapy, allowing for rapid and accurate treatment plan adjustments, thereby enhancing treatment efficacy and safety while reducing the workload of radiologists. In addition, in 2023 Palazzo et al.[5] compared manual and AI-based auto-contouring on CT scans, reporting that AI-based method achieved high accuracy and were clinically viable. Their use significantly reduced radiation treatment planning time, highlighting the benefit of AI as a supportive tool for radiation oncologists.

The aim of this study was to evaluate and compare AI-based prostate segmentation with radiologist-delineated contours on prostate MRI. The AI model used in this study was trained on a public dataset comprising 125 MRI cases (3,193 images), which included expert-labeled prostate segmentations. The trained AI model was then tested on an internal dataset of 109 prostate MRI cases (2,952 images) from the National Cancer Institute. In this study prostate volume measurements derived from AI-generated contours were compared to those manually delineated by radiologists.

Materials and Methods

Study population

This retrospective study included 109 patients who underwent MRI fusion-guided prostate biopsy between 2020 and 2023. Inclusion criteria were: (1) patients who had MRI fusion prostate biopsy during the specified period, and (2) MRI images that were interpreted and segmented by the same radiologist, including both prostate gland and suspicious lesion contours. The prostate gland contouring was performed by radiologists with expertise in prostate MRI interpretation.6 Exclusion criteria included patients whose MRI studies were not reviewed and contoured by a radiologist (Fig. 2).

AI model evaluation and comparison metrics

The quality of AI-generated segmentations was evaluated by comparing them with manual contours created by radiologists. The comparison was based on the dice similarity coefficient (DSC) to assess spatial overlap and the relative percent difference (%RPD) in calculated prostate

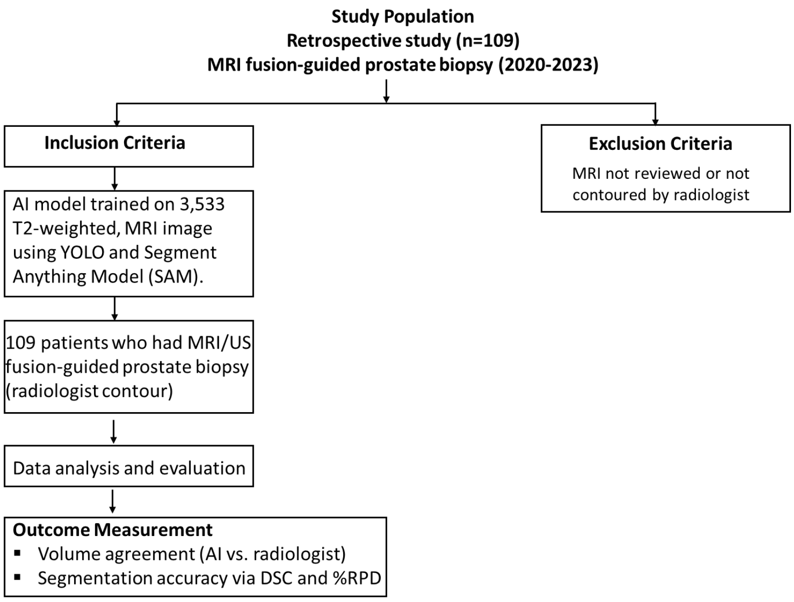

Figure 2. Methodological workflow of AI-based prostate segmentation study

volumes. Statistical significance of the differences in DSC values was assessed using the Wilcoxon signed-rank test, with a primary focus on evaluation of the volume agreement between AI and radiologist contours.

Post-processing refers to the additional refinement of AI-generated prostate contours using image-cleaning and smoothing techniques after the initial segmentation. The 'without post-processing' results represent the raw output directly from the AI model, while the 'with post-processing' results include morphological adjustments to remove noise and enhance contour continuity. However, in this analysis, post-processing slightly reduced segmentation accuracy, indicating that excessive smoothing may have altered true prostate boundaries.

MRI data of the prostate used for training the AI model

In this study, a public dataset named "Prostate158" was used for training the AI model. Although the dataset contains multiple imaging sequences, only 3T prostate T2-weighted sequences were selected for model training. The ground truth segmentations provided in the dataset distinguish the prostate into the peripheral zone (PZ) and the transitional zone (TZ), but these were combined into a single whole-prostate segmentation for training purposes. A total of 3,533 paired images and prostate labels were used to train the model to localize the prostate in MRI scans.

Development of the AI model for automatic prostate segmentation

The AI model for automatic prostate segmentation was developed using a combination of two models. First, the YOLO (You Only Look Once) model was trained to detect the bounding box of the prostate in MRI images. Once the prostate region was localized, the segmentation of the prostate boundary was refined using a fine-tuned segment anything model (SAM) (Fig. 3).

Outcome Measurement and data analysis

In this study, the prostate volume was calculated using the predictions from the trained AI model, which generated prostate contours on T2-weighted MRI (T2w MR) images for every slice. These AI-generated contours were compared with the volumes calculated from manual segmentations performed by radiologists. The prostate volume was calculated using the equation described in [1].

Anteroposterior * Transverse * Longitudinal * π/6 [1]

Note: The results are presented in grams.

Additionally, the accuracy of the AI-generated contours (contour_AI) was evaluated against the contours manually drawn by radiologists using the DSC, a metric that quantifies the level of agreement between the AI-based and radiologist-based segmentations.

$$DCS = \frac{2|contour\_AI \cap contour\_Radiologist|}{|contour\_AI| + |contour\_Radiologist|}$$

Note: A DSC value close to 1 indicates a high similarity between the AI-generated contour and the radiologist's contour.

The difference in volume between the AI-based and radiologist-based prostate contours was assessed using the %RPD. This value was calculated as the percentage difference in prostate volume (in grams) between the AI and radiologist measurements. The volume was derived from the prostate area in each MRI slice using the equation described in [2].

Area Contour * pixel spacing * Slice Thickness [2]

To evaluate the efficacy of the use of AI in interpreting MRI slides for the identification of prostate cancer locations, this study compared AI-based interpretation with that of diagnostic radiologists, aiming to assess the level of agreement between the two approaches. A study by Zhaonan et al.[7] employed AI technology to interpret prostate cancer findings and compared the results with those of diagnostic radiologists. It was found that AI and radiologists had a disagreement rate of 20% across 98 samples, with a 12% difference in interpretation between AI and radiologists. With a statistical power (β) of 80% and a significance level (α) of 0.05, the estimated sample size was approximately 107 cases. To account for potential data loss during collection, an additional 10% of the calculated sample size was included. Therefore, data from a total of 109 cases were collected in this study. The analysis included:

(1) A comparison between the prostate volume calculated from the formula and the

Table 1. Baseline demographic and clinical characteristics of the patients

Parameter	Mean / n (%)	Range
Age (years)	68.6	45-85
PSA (ng/ml)	18.3	0.59-341
PI-RADS 3	18 (16.5)	-
PI-RADS 4	42 (38.5)	-
PI-RADS 5	49 (45.0)	-

PSA = prostate specific antigen

volume determined by the manual contouring by the radiologists;

(2) A comparison between the prostate volume calculated from the formula and the volume derived from AI model-generated contours;

(3) A direct comparison between prostate volumes obtained from radiologist-drawn contours and those from the AI model.

Results

Total of 109 patients were analyzed. Their demographic and clinical characteristics are summarized in Table 1. The age of the patients ranged from 54 to 85 years, with a mean age of 68.6 years. PSA levels ranged from 0.59 to 341 ng/ml, with a mean of 18.26 ng/ml. PI-RADS scores were distributed as follows: PI-RADS 3 in 18 cases, PI-RADS 4 in 42 cases, and PI-RADS 5 in 49 cases. Among the 109 patients, 83 were diagnosed with prostate cancer, with the following staging: Stage 1 (n = 12), Stage 2 (n = 49), Stage 3 (n = 15), and Stage 4 (n = 7). The remaining 26 patients were not diagnosed with cancer.

Prostate volumes calculated using the formula described in equation [1] ranged from 14.3 to 203.9 g, with a mean of 44.51 g.

The mean DSC between radiologist-drawn and AI-generated contours was 0.72 without post-processing and 0.66 with post-processing. The %RPD between radiologist and AI volumes increased from 8.90% to 13.45% after post-processing. The statistical analysis is presented in Table 2 and the results suggest that post-processing did not substantially alter segmentation performance.

The mean %RPD of prostate volume comparisons were as follows:

- Between the calculated volume from Equation [1] and the radiologist-drawn contours:
  - 20.48% (without post-processing)
  - 39.04% (with post-processing)
- Between the calculated volume from Equation [1] and the AI model-generated contours:
  - 20.11% (without post-processing)
  - 55.56% (with post-processing)
- Between the radiologist-drawn contours and the AI model-generated contours:
  - 8.9% (without post-processing)
  - 13.45% (with post-processing

Discussion

This study selected cases of patients who underwent prostate biopsy based on the prevalence of prostate cancer among the sample population of 109 cases. The variations in prostate volume and cancer stage were also included as these are key factors in the development and training of AI models. The goal was to enable an accurate comparison between AI-assisted diagnosis and radiologist interpretation.

The performance of an AI model for automatic prostate gland segmentation on MRI was evaluated and compared to manual segmentation

Table 2. The average DSC and average RPD of prostate volume

Case type	The average DSC comparing prostate contouring between the radiologist and the AI model	The average of the RPD of the prostate volume			P-value (two-tailed)
		Prostate size from Equation [1] vs contour_ Radiologist	Prostate size from Equation [1] vs contour_AI	contour_ Radiologist vs contour_ AI	
Without post-processing n = 56	0.72	20.48	20.11	8.90	< 0.001
With post-processing n = 53	0.66	39.04	55.56	13.45	0.815

DSC = dice similarity coefficient, RPD = relative percent difference, AI = artificial intelligence

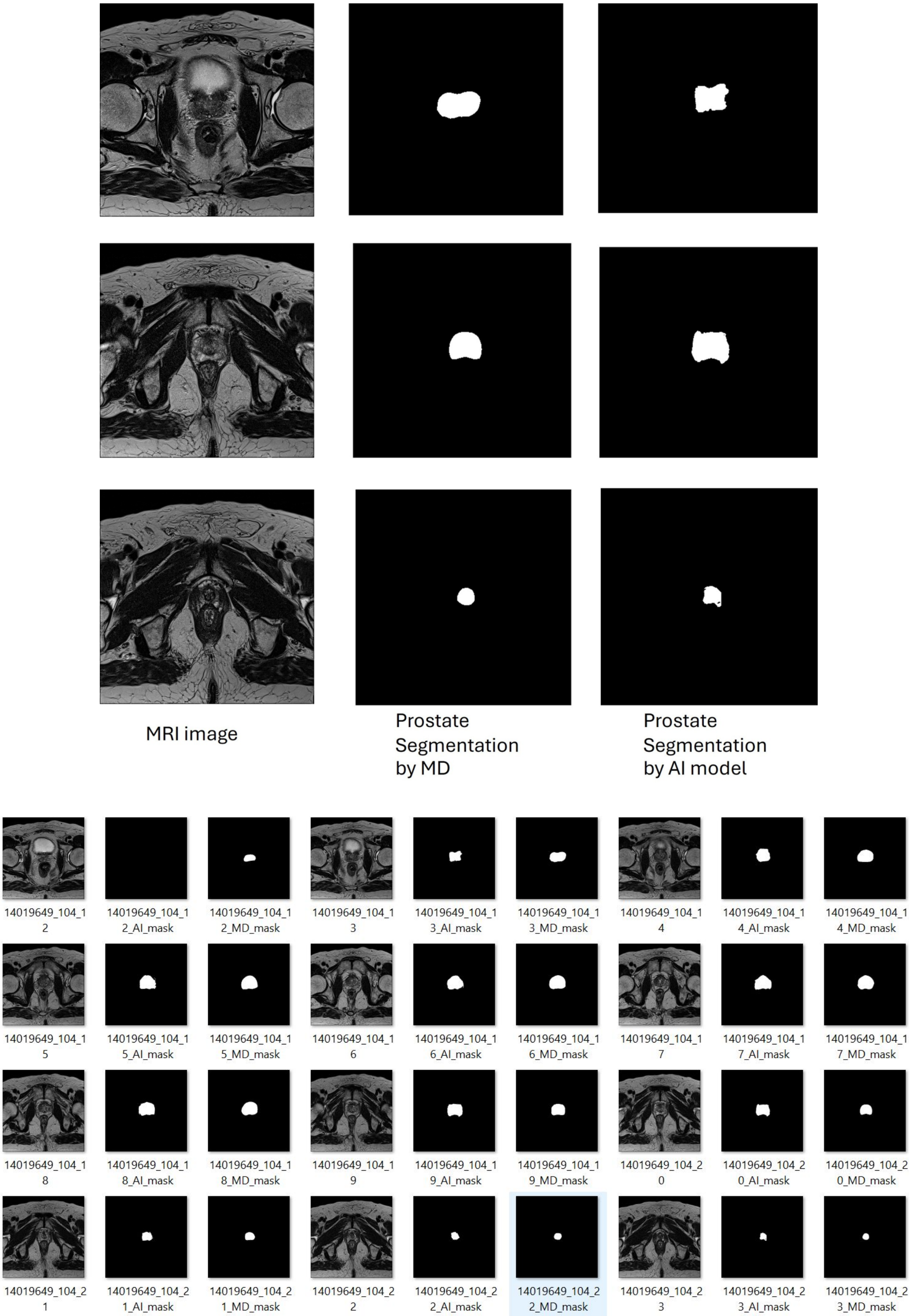

Figure 3. Example of prostate MRI images comparing prostate gland contours manually drawn by a radiologist and those generated by the AI model

by diagnostic radiologists. The aim was to support the accuracy of MRI-guided fusion biopsy procedures. The results demonstrated that AI was effective in producing prostate contours closely resembling those drawn by radiologists, as indicated by the DSC, a standard metric for assessing segmentation accuracy. The findings suggest that AI has strong potential for reducing the workload and processing time for radiologists and urologists in prostate cancer diagnostics, hence improving the experience for patients.

The rationale behind the selection of biopsy-confirmed cases lies in the higher likelihood of detecting cancer, as evidenced by 83 out of 109 cases being positive, encompassing a wide range of prostate sizes and cancer stages. This diversity enhanced the training of the AI mode land allowed for more meaningful comparisons with radiologist-drawn contours.

These findings are consistent with related studies, including that by Ghafoor et al. in 2023[2], which emphasized the value of AI in MRI fusion biopsy workflows. In that study the ability of AI was highlighted with regard to the reduction of inter-reader variability between radiologists and urologists, a common issue in traditional diagnostic pathways. However, in contrast, Thimansson et al. in 2024[8] reported low agreement levels between AI and both local and expert radiologists, suggesting that real-world implementation of AI for prostate cancer screening requires additional model training and validation in specific populations for reliable outcomes.

The benefits of using AI for prostate segmentation are substantial. One major advantage is the reduction in processing time as AI can rapidly identify suspicious regions, minimizing patient waiting time for biopsy procedures. Additionally, AI alleviates the workload of radiologists, allowing them to focus on more complex diagnostic tasks. The use of AI also reduces human error, particularly in cases of fatigue or inexperience, improving the accuracy of cancer detection. In terms of clinical implementation, our team has linked the hospital's PACS system with the developed AI model. This integration allows automatic prostate contour generation whenever a prostate MRI is performed, demonstrating the potential for real-world clinical applications.

This study demonstrated that AI can achieve a satisfactory level of accuracy in prostate contouring. The average DSC between radiologist contours and AI-generated contours was 0.72 in the without post-processing group, considered a good level of agreement, while the with post-processing group showed a slightly lower average DSC of 0.66. This decline suggests that the post-processing step may have introduced alterations that reduced the similarity between AI and radiologist contours.

In terms of volume accuracy, the mean %RPD between mathematically calculated prostate volume (from equation [1]) and radiologist contours was 20.48% in the without post-processing group and increased to 39.04% in the with post-processing group. Similarly, the RPD between the calculated volume and AI contours was 20.11% and 55.56%, respectively. When comparing radiologist and AI-drawn volumes directly, the RPD increased from 8.90% to 13.45% after post-processing. These findings suggest that post-processing may have negatively impacted segmentation precision. Moreover, prostate volume estimation using mathematical formulas showed the lowest reliability compared to radiologist and AI-based measurements, which were more consistent with each other.

The current AI model was trained using axial T2-weighted imaging (T2WI) for prostate gland segmentation. However, precise delineation of targeted lesions would require the inclusion of additional MRI sequences, particularly diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) maps. Future work is planned to extend model training using these sequences to improve lesion-level segmentation accuracy and enhance diagnostic performance.

Although this study highlights the potential of AI in prostate segmentation, certain limitations must be acknowledged. The sample size was relatively small, and larger, more diverse populations are necessary to validate the performance of the AI model across different clinical environments. In addition, the accuracy of the AI model may be influenced by the quality of input images sourced from multiple locations, as the model was trained using public MRI datasets. This could affect the generalizability and reliability of diagnostic results. A potential limitation is that all manual prostate contours were performed by a single radiologist. This may introduce observer bias and limit the diversity of the ground truth, potentially affecting the generalizability of the performance of the AI model compared to multi-radiologist consensus.

Another challenge lies in the tendency of the AI model to mistakenly include surrounding anatomical structures, such as the urethra or rectum—especially near the apex region—as part of the prostate. This necessitates a post-processing step, which, as shown, may inadvertently degrade accuracy. Further refinement of the model is needed to improve its precision in the identification of cancer-suspected regions, including the ability to distinguish PI-RADS scores and differentiate prostate zones for more effective clinical implementation.

To enhance clinical applicability, future development should focus on expansion of the training dataset using diverse imaging data from multiple institutions. This would improve the efficacy of the AI model with regard to generalization and accurate analysis of MRI scans. Additionally, incorporation of advanced machine learning techniques such as deep learning would enable the AI to adapt better to various clinical contexts, including PI-RADS score classification and prostate zone segmentation.

Real-world testing in different hospitals and cancer centers is also essential to ensure reliable AI performance under practical conditions. If AI is to be integrated into clinical workflows, clear medical standards and guidelines should be established to ensure its safe and effective use alongside human radiologists.

Conclusion

This study demonstrates the high potential of AI in automatic prostate segmentation from MRI, with promising results with regard to a reduction in radiologist and urologist workload and processing time. While some limitations remain, further development and clinical validation could enable AI to enhance the efficiency and accuracy of prostate cancer diagnostics, ultimately contributing to faster and more precise patient care.

Conflict of Interest

The author declares no conflict of interest.

References

1. Thanasitthichai S, Ingsirorat R, Chairat C, Chiawiriyabunya I, Wongsena M, Srpitak K, et al. Cancer in Thailand Vol. X, 2019-2021. Bangkok: National Cancer Institute. 2025 [cited 2025 Jan 1]. Available from: https://www.nci.go.th/th/File_download/Nci%20Cancer%20Registry/Cancer%20in%20Thailand%20Vol.XI.pdf
2. Ghafoor S, Steinebrunner F, Stocker D, Hötker AM, Schmid FA, et al. Index lesion contouring on prostate MRI for targeted MRI/US fusion biopsy - Evaluation of mismatch between radiologists and urologists. Eur J Radiol 2023;162:110763.
3. Schelb P, Tavakoli AA, Tubtawee T, Hielscher T, Radtke JP, et al. Comparison of prostate MRI Lesion segmentation agreement between multiple radiologists and a fully automatic deep learning system. Rofo 2021;193:559-73.
4. Nachbar M, Russo ML, Gani C, Boeke S, Wegener D, Paulsen F, et al. Automatic AI-based contouring of prostate MRI for online adaptive radiotherapy. Z Med Phys 2024;34:197-207.
5. Palazzo G, Mangili P, Deantoni C, Fodor A, Broggi S, Castriconi R, et al. Real-world validation of Artificial Intelligence-based computed tomography auto-contouring for prostate cancer radiotherapy planning. Phys Imaging Radiat Oncol 2023;28:100501.
6. Phongkitkarun S. MRI of Prostate Cancer. 1st ed. Bangkok: Ideol Digital Print; 2022.
7. Sunoqrot MRS, Saha A, Hosseinzadeh M, Elschot M, Huisman H. Artificial intelligence for prostate MRI: open datasets, available applications, and grand challenges. Eur Radiol 2022;6:35.
8. Thimansson E, Zackrisson S, Jäderling F, Alterbeck M, Jiborn T. A pilot study of AI-assisted reading of prostate MRI in Organized Prostate Cancer Testing. Acta Oncol 2024;63:816-21.