

Cancer-Net PCa-Seg: Benchmarking Deep Learning Models for Prostate Cancer Segmentation Using Synthetic Correlated Diffusion Imaging

Jarett Dewbury¹, Chi-en Amy Tai¹, and Alexander Wong¹

¹Vision and Image Processing Group, Systems Design Engineering,
University of Waterloo
`{jdewbury, amy.tai, alexander.wong}@waterloo.ca`

January 13, 2025

Abstract

Prostate cancer (PCa) is the most prevalent cancer among men in the United States, accounting for nearly 300,000 cases, 29% of all diagnoses and 35,000 total deaths in 2024. Traditional screening methods such as prostate-specific antigen (PSA) testing and magnetic resonance imaging (MRI) have been pivotal in diagnosis, but have faced limitations in specificity and generalizability. In this paper, we explore the potential of enhancing PCa gland segmentation using a novel MRI modality called synthetic correlated diffusion imaging (CDI^s). We employ several state-of-the-art deep learning models, including U-Net, SegResNet, Swin UNETR, Attention U-Net, and LightM-UNet, to segment prostate glands from a 200 CDI^s patient cohort. We find that SegResNet achieved superior segmentation performance with a Dice-Sørensen coefficient (DSC) of 76.68 ± 0.8 . Notably, the Attention U-Net, while slightly less accurate (DSC 74.82 ± 2.0), offered a favorable balance between accuracy and computational efficiency. Our findings demonstrate the potential of deep learning models in improving prostate gland segmentation using CDI^s to enhance PCa management and clinical support.

1 Introduction

Prostate cancer (PCa) is the most prevalent cancer among men in the United States, accounting for nearly 300,000 cases, 29% of all diagnoses and 35,000 total deaths in 2024 [1]. Although a serious disease, early detection and accurate diagnosis have helped reduce the PCa death rate by half from 1993 to 2013 [2].

Traditional screening methods, such as prostate-specific antigen (PSA) testing, involve measuring the levels of PSA, a protein produced by both cancerous and non-cancerous cells, in the blood [3]. Elevated PSA levels can indicate the presence of PCa, prompting additional diagnostic procedures to confirm the diagnosis, such as a prostate biopsy [3]. However, PSA testing has limitations in specificity and can lead to overdiagnosis and unnecessary interventions that can be harmful for patients [4].

In recent years, magnetic resonance imaging (MRI) has emerged as a powerful tool for prostate cancer detection. Multi-parametric MRI (mpMRI) offers detailed visualization of the prostate tissue, enabling more reliable identification and characterization of potentially clinically significant regions [5]. The integration of deep learning techniques has further enhanced the potential of MRI-based prostate cancer diagnosis. Convolutional Neural Networks (CNNs) have shown promising results in automated prostate cancer detection and segmentation tasks, achieving statistically comparable performance metrics to those of experienced radiologists [6].

Despite these advancements, there are inherent limitations in traditional MRI techniques, including variability in imaging modalities and challenges in distinguishing between cancerous and non-cancerous tissues. These issues often result in poor generalization of artificial intelligence (AI) models when applied to diverse patient populations and imaging conditions, making them obsolete in practice. Several studies have explored the use of segmentation models for PCa lesion detection with a relatively low degree of success. Gunashekhar et al. investigated the use of a 3D-UNet model on a 122 patient mpMRI dataset, achieving a Dice-Sørensen coefficient (DSC) of 0.32 when trained on 150 epochs [7]. Similarly, Pellicer-Valero et al. demonstrated the use of a 3D Retina-UNet model across two different datasets (with 204 and 221 patients respectively), achieving a best PCa lesion DSC of 0.276 [8]. These results underscore the complexity of PCa lesion segmentation across considerable patient cohorts. Accurate segmentation of the prostate gland is a critical prerequisite for effective lesion detection, as it defines the anatomical context necessary for localization. Poor gland segmentation can propagate errors throughout the evaluation pipeline, leading to decreased model performance and limiting clinical utility.

Synthetic correlated diffusion imaging (CDI^s) has shown superior performance in addressing the limitations of other MRI techniques by providing more distinct visualization of clinically relevant regions of interest [9]. CDI^s enhances the contrast between cancerous and healthy tissue, leading to clearer delineation and reduced false positives. While AI has been applied to CDI^s for breast cancer classification and demonstrated improved performance over traditional MRI modalities for grade and post-treatment response prediction [10], its application for prostate region segmentation has yet to be explored.

In this study, we benchmark several state-of-the-art segmentation models using CDI^s data to evaluate their efficacy for prostate gland segmentation in PCa patients. We aim to evaluate these models based on DSC [11], inference time per patient volume, and parameter size, offering a systematic comparison of their performance and viability in a clinical setting. By identifying strengths

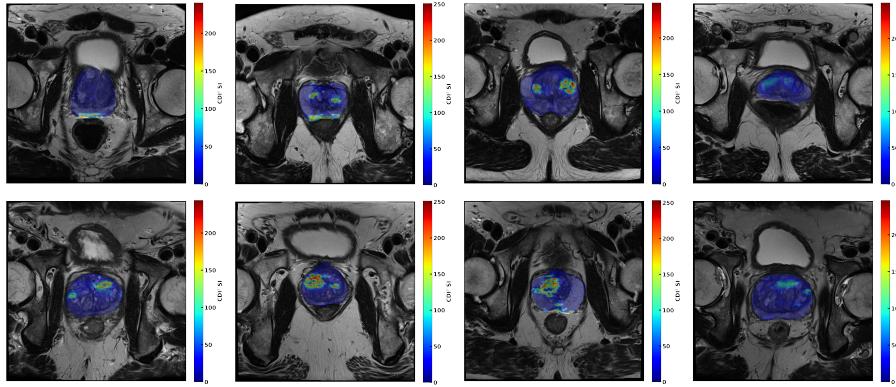


Figure 1: Visualization of sample data from the patient cohort in Cancer-Net PCa-Data dataset. T2-weighted (T2w) images are shown with CDI^s prostate gland region boundary in color overlay. For the purposes of our paper, we consider both clinically significant (csPCa) and clinically insignificant (insPCa) prostate cancer segmentation.

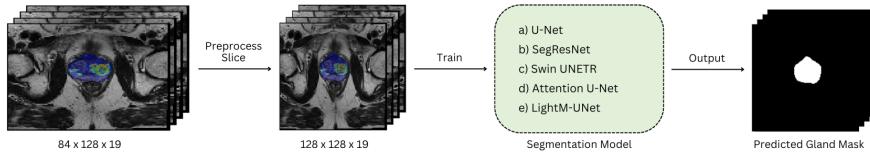


Figure 2: Prostate cancer segmentation pipeline using Cancer-Net PCa-Data. The workflow includes (left to right): input CDI^s volume slice, preprocessing, segmentation across five different deep learning models, and the final predicted prostate gland mask output.

and weaknesses in existing models, our findings can help guide future model development and selection to enhance PCa diagnostic capabilities and promote better patient outcomes.

2 Methodology

2.1 Dataset

This study leveraged the Cancer-Net PCa-Data dataset [12], a novel and publicly available benchmark dataset for PCa research. This data was derived from the SPIE-AAPM-NCI PROSTATEx Challenges, PROSTATEx.masks, and The Cancer Imaging Archive (TCIA) datasets [13, 14, 15, 16]. The dataset contains volumetric CDI^s imaging data for 200 patients along with their associated

prostate gland and tumor lesion mask (see Figure 1 for sample patient data).

2.2 Data Preparation

A training, validation, and test set split of 70%, 15%, and 15% was used respectively. The data was split at the patient level, ensuring the entire set of slices from a single patient volume were assigned to the same data split to prevent data leakage. As illustrated in Figure 2, the preprocessing pipeline involved standardizing the image volumes and corresponding masks using min-max normalization and resizing to 128x128x19, with a batch size of 16. This resulted in a distribution of 2736 training slices, 592 validation slices, and 592 test slices.

2.3 Training

We evaluated the baseline performance on the dataset using commonly used model architectures in medical image segmentation tasks including U-Net [17], SegResNet [18], SwinUNETR [19], Attention U-Net [20], and LightM-UNet [21]. All models were implemented using the MONAI framework [22], except for LightM-UNet, which utilized the PyTorch implementation provided in their paper [21]. Each model was trained for 200 epochs using Binary Cross Entropy with Logits Loss as the primary loss metric. A learning rate of 0.001 was used for all models except LightM-UNet, which was trained with a learning rate of 0.0001. This adjustment significantly improved the performance of LightM-UNet on the dataset. A StepLR scheduler was employed with a step size of 0.2 adjusting the learning rate every 50 epochs. The vanilla Adam optimizer was utilized across all experiments, and no pre-training was conducted on any of the models. Model performance was primarily evaluated using DSC, with the best validation DSC determining the model weights used for testing. To ensure robustness of model performance, each model was trained and evaluated across five different data fold seeds 79, 46, 39, 76, and 42.

Table 1: Experimental results of DSC on the testing set (standard deviation indicated after \pm), inference time in seconds for a single patient volume, and number of parameters in millions reported across the evaluated models.

Model	DSC	Inference Time (s)	Params (M)
U-Net [17]	71.35 ± 1.5	0.0013	1.62
SegResNet [18]	76.68 ± 0.8	0.0016	6.29
SwinUNETR [19]	73.46 ± 1.8	0.0021	6.30
Attention U-Net [20]	74.82 ± 2.0	0.0015	1.99
LightM-UNet [21]	71.38 ± 1.5	0.0016	2.30

3 Results

Table 1 presents a comparative analysis of the evaluated segmentation models and performance on the testing set across DSC, inference time per patient volume, and total parameter count.

Model results are presented as mean \pm standard deviation ($n=5$), where n denotes the total number of independent data folds evaluated on. The THOP library was used to calculate the total number of parameters for each model [23].

The baseline U-Net achieved a DSC of 71.35 ± 1.5 with a significantly smaller parameter count of 1.62 million and the fastest inference time of 0.0013 seconds. Its lightweight architecture and fast inference time make the U-Net particularly suitable for applications where real-time processing speed and resource efficiency is critical over the cost of a marginal improvement in segmentation accuracy.

SegResNet achieved the highest test set DSC of 76.68 ± 0.8 , indicating superior prostate tumor segmentation capabilities on the Cancer-Net PCa-Data dataset. This suggests that SegResNet is particularly effective in capturing the prostate gland boundaries, making it a strong candidate for precise prostate region segmentation applications. However, SegResNet’s superior performance comes at a tradeoff of increased computational demands, with the second largest number of parameters and inference time amongst the architectures of 6.29 million and 0.0016 seconds per patient volume, respectively. This suggests that while SegResNet offers optimal prostate gland segmentation performance out of the evaluated models, its computational demands may limit its applicability in high-throughput clinical scenarios.

Swin UNETR, despite its transformer-based architecture, demonstrated moderate segmentation performance improvements over the U-Net with a DSC of 73.46 ± 1.8 relative to its higher parameter count (6.30 million) and inference time (0.0021 seconds). This limits its applicability in real-time or resource constrained scenarios for PCa segmentation.

Attention U-Net exhibited a notable improvement over the baseline U-Net, with the second highest DSC of 74.82 ± 2.0 and a slight increase in parameters (1.99 million) and inference time (0.0015 seconds). This suggests that the attention mechanism effectively enhances feature selection for identifying the prostate gland without substantially increasing complexity.

The LightM-UNet’s performance was comparable to the U-Net with a DSC of 71.38 ± 1.5 , 2.30 million total parameters, and inference time of 0.0016 seconds per patient volume. Despite its architectural innovations, its marginal performance difference relative to the baseline U-Net suggests that the design advantages of the LightM-UNet may not be fully leveraged by the training pipeline.

4 Conclusion

In this paper, we investigate the performance of various deep learning models to enhance prostate gland segmentation for patients with prostate cancer

(PCa) using a novel MRI modality called synthetic correlated diffusion imaging (CDI^s). SegResNet demonstrated superior segmentation performance with a Dice-Sørensen coefficient (DSC) of 76.68 ± 0.8 , showcasing its potential for precise prostate gland delineation. However, its higher computational demands may limit its applicability in real-time clinical scenarios. Notably, Attention U-Net emerged as a promising compromise between performance and efficiency. With the second highest DSC of 74.82 ± 2.0 and relatively low computational requirements, it represents a viable option for clinical implementation. Given the promising results of the deep learning models on the CDI^s dataset, future work involves exploring additional model architectures that could further optimize segmentation performance and computational efficiency. The transferability of these models on additional CDI^s patient cohorts should also be explored to assess the robustness of the models in clinical practice across diverse patient populations.

References

- [1] Rebecca L Siegel, Angela N Giaquinto, and Ahmedin Jemal. Cancer statistics, 2024. *CA: a cancer journal for clinicians*, 74(1), 2024.
- [2] American Cancer Society. Prostate cancer facts. <https://www.cancer.org/cancer/types/prostate-cancer/about/key-statistics.html>, 2024. Accessed: 2024-08-30.
- [3] National Cancer Institute. What is the psa test? <https://www.cancer.gov/types/prostate/psa-fact-sheet#:~:text=the%20PSA%20test%3F-,What%20is%20the%20PSA%20test%3F,to%20a%20laboratory%20for%20analysis.>, 2024. Accessed: 2024-08-30.
- [4] Stacy Loeb, Marc A Bjurlin, Joseph Nicholson, Teuvo L Tammela, David F Penson, H Ballantine Carter, Peter Carroll, and Ruth Etzioni. Overdiagnosis and overtreatment of prostate cancer. *European urology*, 65(6):1046–1055, 2014.
- [5] Armando Stabile, Francesco Giganti, Andrew B Rosenkrantz, Samir S Taneja, Geert Villeirs, Inderbir S Gill, Clare Allen, Mark Emberton, Caroline M Moore, and Veeru Kasivisvanathan. Multiparametric mri for prostate cancer diagnosis: current status and future directions. *Nature reviews urology*, 17(1):41–61, 2020.
- [6] Jason C Cai, Hirotugu Nakai, Shiba Kuanar, Adam T Froemming, Candice W Bolan, Akira Kawashima, Hiroaki Takahashi, Lance A Mynderse, Chandler D Dora, Mitchell R Humphreys, et al. Fully automated deep learning model to detect clinically significant prostate cancer at mri. *Radiology*, 312(2):e232635, 2024.
- [7] Deepa Darshini Gunashekhar, Lars Bielak, Leonard Hägele, Benedict Oerther, Matthias Benndorf, Anca-L Grosu, Thomas Brox, Constantinos Zamboglou, and Michael Bock. Explainable ai for cnn-based prostate tumor segmentation in multiparametric mri correlated to whole mount histopathology. *Radiation Oncology*, 17(1):65, 2022.
- [8] Oscar J Pellicer-Valero, Jose L Marenco Jimenez, Victor Gonzalez-Perez, Juan Luis Casanova Ramon-Borja, Isabel Martín García, María Barrios Benito, Paula Pelechano Gomez, José Rubio-Briones, María José Rupérez, and José D Martín-Guerrero. Deep learning for fully automatic detection, segmentation, and gleason grade estimation of prostate cancer in multiparametric magnetic resonance images. *Scientific reports*, 12(1):2975, 2022.
- [9] Alexander Wong, Hayden Gunraj, Vignesh Sivan, and Masoom A Haider. Synthetic correlated diffusion imaging hyperintensity delineates clinically significant prostate cancer. *Scientific Reports*, 12(1):3376, 2022.
- [10] Chi-en Amy Tai, Hayden Gunraj, Nedim Hodzic, Nic Flanagan, Ali Sabri, and Alexander Wong. Enhancing clinical support for breast cancer with deep learning models using synthetic correlated diffusion imaging. In *International Workshop on Applications of Medical AI*, pages 83–93. Springer, 2023.
- [11] Thorvald Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Biologiske skrifter*, 5:1–34, 1948.

- [12] Hayden Gunraj, Chi en Amy Tai, and Alexander Wong. Cancer-net pca-data: An open-source benchmark dataset for prostate cancer clinical decision support using synthetic correlated diffusion imaging data. *NeurIPS Workshops*, 2023.
- [13] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. Computer-aided detection of prostate cancer in mri. *IEEE Transactions on Medical Imaging*, 33(5):1083–1092, 2014.
- [14] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. Prostatax challenge data [data set], 2017.
- [15] Renato Cuocolo, Arnaldo Stanzione, Anna Castaldo, Davide Raffaele De Lucia, and Massimo Imbriaco. Quality control and whole-gland, zonal and lesion annotations for the prostatax challenge public dataset. *European Journal of Radiology*, 138:109647, 2021.
- [16] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
- [18] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, pages 311–320. Springer, 2019.
- [19] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop*, pages 272–284. Springer, 2021.
- [20] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [21] Weibin Liao, Yinghao Zhu, Xinyuan Wang, Chengwei Pan, Yasha Wang, and Liantao Ma. Lightm-unet: Mamba assists in lightweight unet for medical image segmentation. *arXiv preprint arXiv:2403.05246*, 2024.
- [22] MONAI. Monai: Medical open network for ai. <https://monai.io/>, 2024. Accessed: 2024-08-30.
- [23] Ligeng Zhu. Thop: Pytorch-opcounter. <https://github.com/Lyken17/pytorch-OpCounter>, 2022. Accessed: 2024-07-21.