Radiology: Artificial Intelligence

ORIGINAL RESEARCH

# Deep Learning Prostate MRI Segmentation Accuracy and Robustness: A Systematic Review

*Mohammad-Kasim Fassia, MD, MS* • *Adithya Balasubramanian, MD* • *Sungmin Woo, MD, PhD* • *Hebert Alberto Vargas, MD* • *Hedvig Hricak, MD, PhD* • *Ender Konukoglu, PhD* • *Anton S. Becker, MD, PhD*

From the Departments of Radiology (M.K.F.) and Urology (A.B.), New York-Presbyterian Weill Cornell Medical Center, 525 E 68th St, New York, NY 10065-4870; Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY (S.W., H.A.V., H.H., A.S.B.); and Department of Biomedical Imaging, ETH-Zurich, Zurich Switzerland (E.K.). Received April 26, 2023; revision requested June 14; revision received February 24, 2024; accepted March 19. **Address correspondence to** M.K.F. (email: *mhf9004@nyp.org*).

Authors declared no funding for this work.

Conflicts of interest are listed at the end of this article.

*Radiology: Artificial Intelligence* 2024; 6(4):e230138 • https://doi.org/10.1148/ryai.230138 • Content codes: AI GU MR

***Purpose:*** To investigate the accuracy and robustness of prostate segmentation using deep learning across various training data sizes, MRI vendors, prostate zones, and testing methods relative to fellowship-trained diagnostic radiologists.

***Materials and Methods:*** In this systematic review, Embase, PubMed, Scopus, and Web of Science databases were queried for English-language articles using keywords and related terms for prostate MRI segmentation and deep learning algorithms dated to July 31, 2022. A total of 691 articles from the search query were collected and subsequently filtered to 48 on the basis of predefined inclusion and exclusion criteria. Multiple characteristics were extracted from selected studies, such as deep learning algorithm performance, MRI vendor, and training dataset features. The primary outcome was comparison of mean Dice similarity coefficient (DSC) for prostate segmentation for deep learning algorithms versus diagnostic radiologists.

***Results:*** Forty-eight studies were included. Most published deep learning algorithms for whole prostate gland segmentation (39 of 42 [93%]) had a DSC at or above expert level (DSC ≥ 0.86). The mean DSC was 0.79 ± 0.06 (SD) for peripheral zone, 0.87 ± 0.05 for transition zone, and 0.90 ± 0.04 for whole prostate gland segmentation. For selected studies that used one major MRI vendor, the mean DSCs of each were as follows: General Electric (three of 48 studies), 0.92 ± 0.03; Philips (four of 48 studies), 0.92 ± 0.02; and Siemens (six of 48 studies), 0.91 ± 0.03.

***Conclusion:*** Deep learning algorithms for prostate MRI segmentation demonstrated accuracy similar to that of expert radiologists despite varying parameters; therefore, future research should shift toward evaluating segmentation robustness and patient outcomes across diverse clinical settings.

Systematic review registration link: osf.io/nxaev



Prostate MRI is integral for prostate cancer detection, staging, and disease surveillance because of its exceptional resolution, contrast, and ability to extract in vivo functional information (1). However, the current diagnostic approach relies on qualitative or semiquantitative assessments (2,3). Fully quantitative diagnostic assessment relies on prostate segmentation, which is routinely done for focal or radiation therapy planning. Unfortunately, performing this task manually is prohibitively time-consuming and harbors substantial interreader variability (4). Thus, automated prostate segmentations may not only unlock novel and more accurate prognostic biomarkers but also allow for more efficient therapy planning (Fig 1).

Despite these promising benefits, automated prostate segmentation remained challenging because of ambiguous prostatic boundaries (in particular at the apex), wide glandular morphologic variability, and technical interscanner differences. However, in recent years deep learning algorithms have emerged as a potential solution to automate this difficult task. Since 2016, an increasing body of scientific literature has been published to improve on this task (Fig 2).

Despite their growing popularity, questions remain about the robustness and accuracy of deep learning algorithms for prostate gland segmentation. Deep learning algorithms are often trained on small, highly curated datasets. In addition, training data are often acquired from a single MRI vendor (eg, General Electric, Philips, or Siemens) or a specific geographic location or narrow patient population. These study limitations inspire multiple questions surrounding deep learning performance for prostate anatomy segmentation: Can deep learning algorithms for prostate anatomy segmentation be applied to different MRI vendors, training dataset sizes, and different prostate zones? Is the testing performed on each algorithm (eg, internal vs external testing) adequate to yield accurate results? Is the accuracy of prostate anatomy segmentation performed by deep learning algorithms sufficient when compared with that of board-certified radiologists, or is greater investment in algorithm performance necessary to improve them?

Hence, the aim of this systematic review was to summarize the performance of prostate anatomy segmentation algorithms in the published biomedical literature. With additional subgroup analyses, we also aimed to investigate

### Abbreviations

ANOVA = analysis of variance, CPU = central processing unit, DSC = Dice similarity coefficient, GPU = graphics processing unit


### Summary

Deep learning algorithms outperformed expert radiologists in prostate anatomy segmentation agreement, but mere accuracy gains may lack clinical impact; therefore, future focus should be on robustness and real-world patient outcomes.


### Key Points

- Ninety-three percent of studies on deep learning for prostate gland segmentation had a Dice similarity coefficient (DSC) at or above expert level (DSC ≥ 0.86).
- Performance for whole prostate gland segmentation (0.90 ± 0.04 [SD]) was similar to that for transition zone segmentation (0.87 ± 0.05) but greater than that for peripheral zone segmentation (0.79 ± 0.06).
- Given the high performance of deep learning on the task of prostate gland segmentation, future research should shift toward evaluating robustness and clinical outcomes.


### Keywords

MRI, Genital/Reproductive, Prostate Segmentation, Deep Learning


any association between performance and algorithm architecture, experimental design, sample size, and variability of MRI scanner data with regard to different vendors.

## Materials and Methods

The study was designed to conform with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (5). The systematic review was registered with OSF online registries *(https://osf.io/2bz7t?view_only=cc234cfb711d4e078fbea7496f551750)*.

### Literature Search

Embase, PubMed, Scopus, and Web of Science databases were queried for English-language articles dated to July 31, 2022. The following search query was used for all databases: "(magnetic resonance imaging OR MRI) AND (prostate segmentation) AND (deep learning OR machine learning OR artificial intelligence) AND (automated OR automatic)". The search query was constructed using keywords and their related terms for prostate MRI segmentation and deep learning algorithms. After the potential articles were identified, citations from each article were further evaluated to find any related articles that may have been missed by the initial search query.

### Study Selection

Inclusion criteria for articles were as follows: *(a)* whole prostate gland segmentation and/or peripheral zone segmentation and/or transition zone segmentation, *(b)* convolutional neural networks, and *(c)* multilayer convoluted neural networks (deep learning). Exclusion criteria consisted of the following: *(a)* prostate lesion segmentation, *(b)* prostate cancer staging, *(c)* machine learning algorithms other than deep learning (eg, Bayesian inference, support vector machine), *(d)* CT or US segmentation, *(e)* conference abstract, or *(f)* review article. The purpose of this systematic review was aimed at evaluating the performance of deep learning on prostate gland segmentation by comparing results from original research; therefore, review articles, conference abstracts, and applications for lesion detection and prostate cancer staging were not included. In addition, deep learning models are generally the highest scoring algorithms in web-based competitions for prostate gland segmentation; thus, non–deep learning algorithms were excluded from this study.

The study selection, summarized in Figure 3, was performed by two reviewers (one postgraduate year 3 interventional and diagnostic radiology resident [M.K.F.] and one postgraduate year 3 urology resident [A.B.]) and was confirmed by one subspecialized genitourinary radiologist (A.B., with 4 years of experience since board certification).

### Data Extraction

The following characteristics were extracted from each article that met the inclusion criteria: *(a)* deep learning parameters: neural network architecture (eg, U-Net, ResNet), number of neural network layers, epochs, initial learning rate, image input dimension, and batch size; *(b)* MRI: sequence, field strength, vendor (ie, General Electric, Philips), endorectal coil; *(c)* datasets: segmentation software, data source, sample size, test size, validation test size, external testing, and internal testing; *(d)* hardware and software: central processing unit (CPU) type, CPU number, graphic processing unit (GPU) type, GPU number, and public availability of source code; and *(e)* deep learning algorithm performance as measured by the Dice similarity coefficient (DSC).

Data were extracted by one reviewer (M.K.F.) and independently verified by a second reviewer (A.B.).

### Outcomes

The primary outcome from this systematic review was comparison of the mean DSC between two or more radiologists and comparison of DSC between a deep learning algorithm and one or more genitourinary radiologists. The DSC is defined as the agreement or overlap of a segmented prostate by two or more different interpreters (eg, one radiologist and one deep learning algorithm). A DSC of 1 signifies that the segmentation between the two interpreters is exactly the same. Conversely, a DSC of 0 signifies that there is no overlap between the two segmented prostate volumes.

### Statistical Analysis

Changes of algorithm performance (DSC) over time (qualitative absolute differences, no statistical testing), variations of DSC between anatomic prostate zones, and the mean DSC estimates were compared between studies using internal versus external testing sets using the Student *t* test, and Spearman correlation was estimated to determine whether there was an association between algorithm performance, as measured by DSC, and sample size. A *P* value of .05 was used to indicate

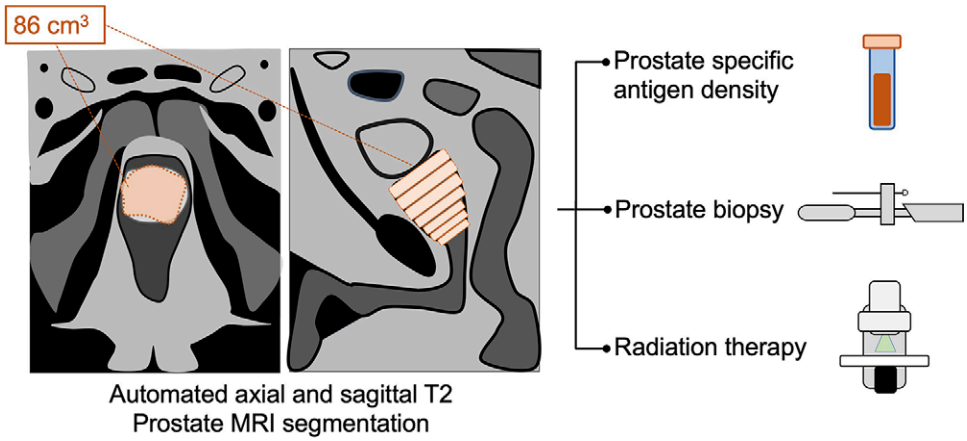


**Figure 1:** Automated prostate segmentation of T2-weighted MR images used to estimate three-dimensional prostate size. Prostate volume measurements have multiple applications, including calculation of prostate-specific antigen density, prostate biopsy planning, and calculation of radiation therapy doses.

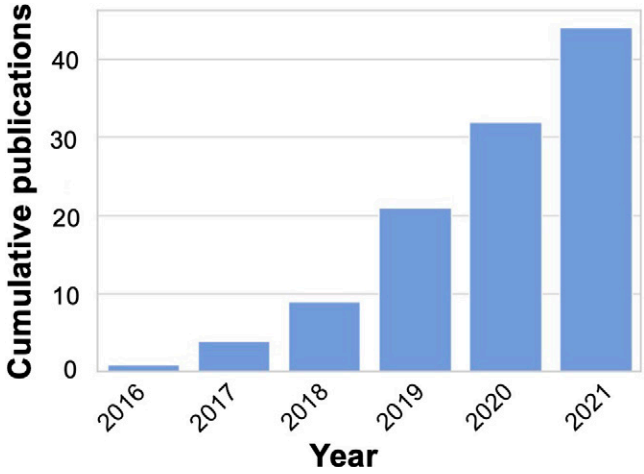


**Figure 2:** Bar graph shows that the total number of publications for deep learning MRI prostate segmentation meeting criteria for this systematic review has increased each subsequent year.

statistical significance. Analysis of variance (ANOVA) was performed on the different combinations of model architectures, MRI vendors, and DSC. Data analysis and visualization were performed with Python 3.7, using the NumPy, SciPy, and Seaborn packages.

## Results

### Study Selection

The systematic search of Embase, PubMed, Scopus, and Web of Science databases yielded a total of 691 articles related to prostate anatomy segmentation using deep learning algorithms. After the removal of duplicates, title screening, abstract screening, and full-text reviews, a total of 48 articles were included. The flowchart in Figure 3 depicts the study selection process in detail.

### Characteristics of Selected Studies

Tables 1–3 summarize the characteristics of the studies selected for systematic review. The selected studies were published between 2016 and 2022. Most studies used a combination of private and public datasets (18 of 48); however, some studies used only private datasets (15 of 48) or only public datasets (15 of 48). In addition to varying types of datasets, more studies acquired their training data from multiple MRI vendors (30 of 48) as opposed to only one MRI vendor (18 of 48). For prostate anatomy segmentation, the entire prostate gland can be segmented or specific anatomic prostate zones, such as the transition zone or peripheral zone, can be segmented. In this systematic review, most studies performed whole-gland segmentation or a combination of whole-gland segmentation with peripheral zone and/or transition zone segmentation (43 of 48). The remaining studies investigated only peripheral zone and/or transition zone segmentation (five of 48). From the selected studies, the following are the median values for commonly reported deep learning parameters: epochs, 100 (27 of 48 studies reported); initial learning rate, $1.0 \times 10^{-4}$ (38 of 48 studies reported); and batch size, eight (31 of 48 studies reported).

### Deep Learning Algorithm Performance Characteristics

The first journal article describing prostate MRI segmentation using deep learning was published in 2016 (6). The authors of that article demonstrated a mean DSC of 0.87 for whole prostate gland segmentation. During the next 6 years, the highest-performing algorithm demonstrated a mean DSC of 0.98, equaling to an improvement of 0.11.

With use of a threshold DSC of 0.86 for expert performance (six readers and 80 prostate MRI examinations) of whole prostate gland segmentation from the literature (7), most published deep learning algorithms (39 of 42 [93%]) had a DSC at or above expert level (Fig 4). Of note, even though slightly higher human reference performances have been published in the literature, with a DSC of approximately 0.94 (8,9) and 40 prostate MRI examinations, we chose the study with the most segmentations as a reference standard (7).

The published studies on deep learning for prostate segmentation had various implementations of algorithm testing. Most studies (32 of 48 [67%]) performed external testing, wherein a

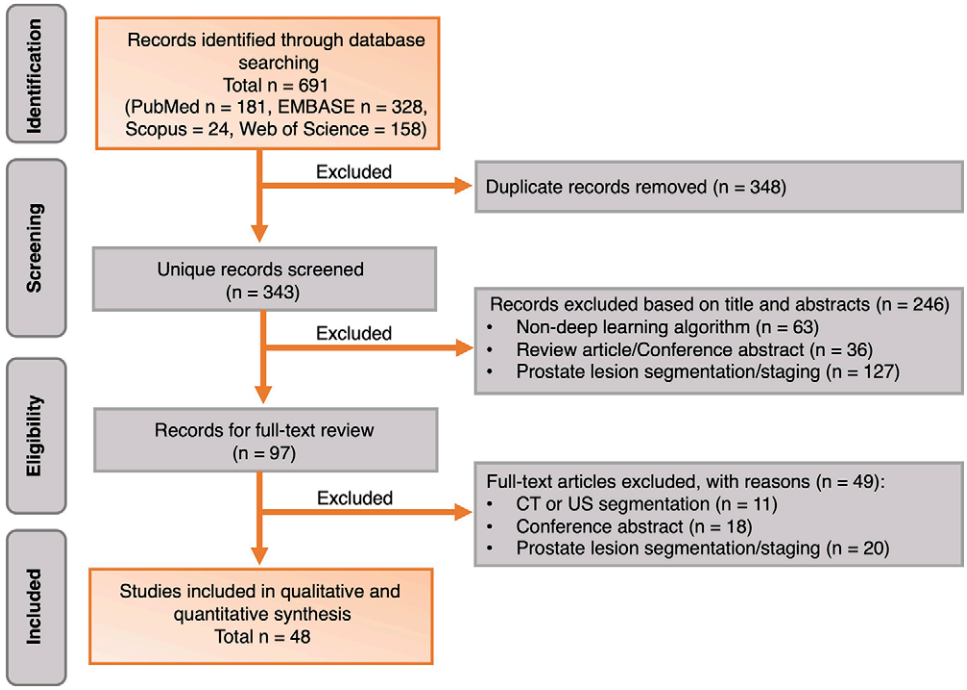


**Figure 3:** Flow diagram of study selection. PubMed, Embase, Scopus, and Web of Science databases were used to identify articles related to deep learning and prostate MRI segmentation. Articles found in multiple databases were removed, leaving only unique journal articles. Unique articles were filtered if they did not contain "deep learning" or "prostate MRI segmentation" in the title or abstract. The remaining articles were comprehensively reviewed. If the study, despite mentioning MRI and being included initially, was focused on CT or US and not MRI, it was removed. Likewise, if the entry was from a conference abstract, or if prostate lesions were segmented, the article was removed from further analysis.

**Table 1: Summary of 48 Included Studies**

| Variable | No. of Studies |
|---|---|
| Publication date | |
| 2016 | 1 |
| 2017 | 3 |
| 2018 | 5 |
| 2019 | 12 |
| 2020 | 11 |
| 2021 | 12 |
| 2022 | 4 |
| Vendor | |
| General Electric | 2 |
| Philips | 3 |
| Siemens | 10 |
| Mixed | 33 |
| Dataset availability | |
| Public | 15 |
| Private | 15 |
| Mixed | 18 |
| Zonal segmentation | |
| Peripheral and transition | 5 |
| Whole gland | 33 |
| Mixed | 10 |

select sample of data were held out from algorithm training, and algorithm performance was measured with the held out sample. The remaining studies (16 of 48 [33%]) used internal testing or did not explicitly describe using external testing methods.

The most common neural network architectures from the selected articles included U-Net (21 of 48), ResNet (four of 48), and DenseNet (three of 48). DSC scores did not statistically significantly differ between the most common neural network architectures based on ANOVA analysis.

Of the studies that conducted external testing, approximately 88% (28 of 32) performed whole-gland segmentation, with a mean DSC of 0.904 ± 0.029 (SD). Of the studies that conducted internal testing, 94% (15 of 16) performed whole-gland segmentation, with a mean DSC of 0.914 ± 0.039. There was no evidence of a difference in the reported DSC between internally and externally tested studies ($P$ = .35) used for whole-gland segmentation.

Aside from various implementations of algorithm testing, prostate MRI segmentation articles had varying sample sizes. From the selected articles, sample sizes ranged from 40 to 550 MRI examinations. From the selected studies that performed whole-gland segmentation (39 of 42), Spearman correlation was used to determine whether there was a relationship between algorithm performance (DSC) and sample size. The correlation coefficient was 0.247, which indicates a weak positive correlation between larger sample sizes and improved prostate segmentation performance.

In addition to sample size, other parameters that varied between studies included MRI sequences and endorectal coil use.

**Table 2: Characteristics of Selected Articles for MRI Prostate Segmentation Using Deep Learning and External Testing**

| Author (Reference) | Year | Data Source | Total Sample Size (*n*) | CNN Architecture | MRI Vendor | Endorectal Coil | MRI Sequence | Prostate Zones | Mean DSC | Published Code |
|---|---|---|---|---|---|---|---|---|---|---|
| Astono et al (19) | 2020 | Mixed | 91 | U-Net | GE, Siemens | Yes | T2W | WG | 0.89 | Private |
| Bardis et al (20) | 2021 | Private | 242 | U-Net | Philips, Siemens | No | T2W | PZ, TZ, WG | 0.91 (TZ), 0.77 (PZ), 0.94 (WG) | Private |
| Clark et al (23) | 2017 | Mixed | 141 | U-Net | GE, Philips, Siemens | Yes | DWI, T2W | WG | 0.89 | Private |
| Cuocolo et al (25) | 2021 | Mixed | 204 | U-Net, ENet, ERFNet | Siemens | Yes | DCE, DWI, T1W, T2W | PZ, TZ, WG | 0.71 (PZ), 0.87 (TZ), 0.91 (WG) | Private |
| da Silva et al (26) | 2020 | Mixed | 91 | FCN | GE, Siemens | Yes | T2W | WG | 0.89 | Private |
| Geng et al (27) | 2019 | Public | 110 | ResNet | GE, Philips, Siemens | Yes | T2W | WG | 0.95 | Private |
| Ghavami et al (28) | 2019 | Private | 232 | U-Net, VNet, ResNet | Siemens | No | T2W | WG | 0.89 | Private |
| Guo et al (6) | 2016 | Private | 66 | Autoencoder | GE, Philips | Yes | T2W | WG | 0.87 | Private |
| Hassanzadeh et al (29) | 2019 | Public | 50 | DenseNet, U-Net | GE, Siemens | Yes | T2W | WG | 0.87 | Private |
| Jensen et al (30) | 2019 | Private | 40 | U-Net | GE, Siemens | No | T2W | PZ, TZ | 0.69 (PZ), 0.81 (TZ) | Private |
| Jia et al (32) | 2018 | Public | 304 | LeNet, ReLU | GE, Siemens | Yes | DCE, DWI, PDW, T2W | WG | 0.91 | Private |
| Jia et al (33) | 2022 | Public | 180 | ResNet | GE, Philips, Siemens | Yes | T2W | WG | 0.93 | Private |
| Jin et al (34) | 2021 | Mixed | 206 | V-Net | GE, Siemens | Yes | T2W | WG | 0.96 | Private |
| Lee et al (37) | 2020 | Private | 330 | V-Net | Siemens | No | T2W | TZ, WG | 0.76 (TZ), 0.87 (WG) | Private |
| Liu et al (42) | 2019 | Mixed | 313 | ResNet | Siemens | No | DCE, DWI, PDW, T2W | PZ, TZ | 0.74 (PZ), 0.79 (TZ) | Public |
| Meyer et al (43) | 2021 | Mixed | 108 | U-Net | Philips, Siemens | No | DCE, DWI, PDW, T2W | WG | 0.93 | Public |
| Nai et al (44) | 2020 | Public | 160 | DenseNet | Siemens | No | ADC, DWI, T2W | PZ, TZ, WG | 0.71 (PZ), 0.86 (TZ), 0.89 (WG) | Private |
| Rouvière et al (45) | 2022 | Mixed | 207 | U-Net | GE, Philips, Siemens | Yes | T2W | WG | 0.96 | Private |

***(Table 2 continues)***

**Table 2 (*continued*): Characteristics of Selected Articles for MRI Prostate Segmentation Using Deep Learning and External Testing**

| Author (Reference) | Year | Data Source | Total Sample Size (*n*) | CNN Architecture | MRI Vendor | Endorectal Coil | MRI Sequence | Prostate Zones | Mean DSC | Published Code |
|---|---|---|---|---|---|---|---|---|---|---|
| Rundo et al (46) | 2019 | Mixed | 80 | U-Net | Philips, Siemens | Yes | T2W | PZ, TZ | 0.91 (PZ), 0.94 (TZ) | Private |
| Sanford et al (47) | 2020 | Private | 406 | U-Net | GE, Philips, Siemens | Yes | T2W | WG | 0.92 | Private |
| Salvaggio et al (48) | 2022 | Private | 103 | ENet, U-Net | Philips | No | DWI, T1W, T1CE, T2W | WG | 0.90 | Private |
| Salvi et al (49) | 2022 | Private | 60 | U-Net | GE | Yes | T2W | WG | 0.89 | Public |
| Sunoqrot et al (50) | 2021 | Private | 244 | U-Net | Siemens | No | T2W | PZ, TZ, WG | 0.82 (PZ), 0.92 (TZ), 0.94 (WG) | Private |
| Tian et el (52) | 2020 | Mixed | 112 | ResNet | GE, Siemens | Yes | T2W | WG | 0.94 | Private |
| Tian et al (53) | 2018 | Mixed | 140 | FCN | GE, Siemens | Yes | T2W | WG | 0.85 | Private |
| To et al (54) | 2018 | Mixed | 280 | DenseNet | GE, Philips, Siemens | Yes | ADC, T2W | WG | 0.95 | Private |
| Ushinsky et al (55) | 2021 | Private | 299 | U-Net | Philips, Siemens | No | T2W | WG | 0.91 | Private |
| Wang et al (56) | 2021 | Mixed | 270 | U-Net, FCN | Siemens | Yes | T2W | WG | 0.89 | Public |
| Yan et al (58) | 2019 | Public | 80 | DSNet, FCN | GE, Siemens | Yes | T2W | WG | 0.91 | Private |
| Zabihollahy et al (59) | 2019 | Private | 225 | U-Net | GE | No | ADC, T2W | PZ, TZ, WG | 0.87 (PZ), 0.94 (TZ), 0.93 (WG) | Private |
| Zhu et al (61) | 2020 | Mixed | 40 | U-Net, DenseNet | GE, Philips, Siemens | Yes | T2W | WG | 0.91 | Private |
| Zhu et al (62) | 2019 | Private | 163 | U-Net | Philips | No | DWI, T2W | PZ, WG | 0.79 (PZ), 0.93 (WG) | Private |

Note.—"Data Source" is considered public if the training data used in the journal article are publicly available such as the SPIE-AAPM-NCI PROSTATEx Challenge (PROSTATEx), National Cancer Institute Cancer International Society for Biomedical Imaging 2013 Prostate Segmentation Challenge (NCI-ISIBI 2013), Prostate MR Image Segmentation Challenge (PROMISE12), The Cancer Imaging Archive (TCIA), or Initiative for Collaborative Computer Vision Benchmarking (I2CVB). Datasets labeled "private" were curated by the journal authors and were not made publicly available. Articles using private and public datasets are categorized as "mixed." The architecture for each deep learning algorithm is as follows: Dense Convolutional Neural Network (DenseNet), Dense VNet (DVNet), U-Net (Net), VNet (VNet), Holisitic Neural Network (HNN), Fully Convolutional Network (FCN), Efficient Neural Network (ENet), Efficient Residual Factorized Convolutional Network (ERFNet), Residual Networks (ResNet), AutoEncoder, LeNet (LeNet), ReLU (Rectified Linear Unit), Domain Adversarial Neural Network (DANN). Authors can publish their deep learning algorithms through various hosting websites (eg, GitHub). If an author published the code associated with journal article, it is designated "public" under the "Published Code" column. If code for the deep learning algorithm was not published, then it is labeled "private." ADC = apparent diffusion coefficient, CNN = convolutional neural network, DCE = dynamic contrast-enhanced, DSC = Dice similarity coefficient, DWI = diffusion-weighted imaging, GE = General Electric, PDW = proton density–weighted, PZ = peripheral zone of the prostate, T1CE = T1-weighted contrast-enhanced, T1W = T1-weighted, T2W = T2-weighted, TZ = transition zone of the prostate, WG = whole prostate gland.

**Table 3: Characteristics of Selected Articles for MRI Prostate Segmentation Using Deep Learning and Internal Testing**

| Author (Reference) | Year | Data Source | Total Sample Size (*n*) | CNN Architecture | MRI Vendor | Endorectal Coil | MRI Sequence | Prostate Zones | Mean DSC | Published Code |
|---|---|---|---|---|---|---|---|---|---|---|
| Agarwal et al (18) | 2021 | Public | 50 | DVNet | GE, Siemens | Yes | T2W | WG | 0.87 | Private |
| Cheng et al (21) | 2019 | Public | 80 | HNN | Siemens | Yes | T2W | TZ, WG | 0.90 (TZ), 0.92 (WG) | Private |
| Cheng et al (22) | 2017 | Public | 250 | FCN, HNN | Philips | Yes | T2W | WG | 0.90 | Private |
| Comelli et al (24) | 2021 | Private | 85 | U-Net, ENet, ERFNet | Philips | No | T2W | WG | 0.91 | Private |
| Jia et al (31) | 2020 | Public | 110 | ResNet | GE, Philips, Siemens | Yes | T2W | WG | 0.90 | Private |
| Karimi et al (35) | 2018 | Mixed | 75 | DenseNet | GE, Siemens | Yes | T2W | WG | 0.88 | Private |
| Khan et al (36) | 2020 | Mixed | 51 | FCN, U-Net | Siemens | Yes | T2W | PZ, TZ | 0.79 (PZ), 0.93 (TZ) | Private |
| Liu et al (38) | 2021 | Private | 335 | DANN | Siemens | Yes | T2W | WG | 0.93 | Private |
| Liu et al (39) | 2020 | Public | 50 | Autoencoder | GE, Siemens | Yes | T2W | WG | 0.89 | Private |
| Liu et al (40) | 2020 | Public | 304 | ResNet | Siemens | No | DCE, DWI, PDW, T2W | PZ, TZ | 0.79 (PZ), 0.89 (TZ) | Private |
| Liu et al (41) | 2020 | Public | 79 | U-Net | Philips, Siemens | Yes | ADC, DWI, T2W | WG | 0.92 | Private |
| Tao et al (51) | 2021 | Public | 50 | U-Net | GE, Siemens | Yes | T2W | WG | 0.92 | Private |
| Wang et al (57) | 2019 | Mixed | 90 | FCN | GE, Siemens | Yes | T2W | WG | 0.85 | Private |
| Yan et al (12) | 2021 | Public | 270 | U-Net, FCN | GE, Siemens | Yes | T2W | WG | 0.98 | Private |
| Zavala-Romero et al (60) | 2020 | Mixed | 550 | U-Net | GE, Siemens | No | DCE, DWI, T2W | PZ, WG | 0.82 (PZ), 0.91 (WG) | Private |
| Zhu et al (63) | 2018 | Private | 80 | U-Net | Philips | Yes | T2W | WG | 0.94 | Private |

Note.—"Data Source" is considered public if the training data used in the journal article are publicly available such as the SPIE-AAPM-NCI PROSTATEx Challenge (PROSTATEx), National Cancer Institute Cancer International Society for Biomedical Imaging 2013 Prostate Segmentation Challenge (NCI-ISIBI 2013), Prostate MR Image Segmentation Challenge (PROMISE12), The Cancer Imaging Archive (TCIA), or Initiative for Collaborative Computer Vision Benchmarking (I2CVB). Datasets labeled "private" were curated by the journal authors and were not made publicly available. Articles using private and public datasets are categorized as "mixed." The architecture for each deep learning algorithm is as follows: Dense Convolutional Neural Network (DenseNet), Dense VNet (DVNet), U-Net (Net), VNet (VNet), Holisitic Neural Network (HNN), Fully Convolutional Network (FCN), Efficient Neural Network (ENet), Efficient Residual Factorized Convolutional Network (ERFNet), Residual Networks (ResNet), AutoEncoder, LeNet (LeNet), ReLU (Rectified Linear Unit), Domain Adversarial Neural Network (DANN). Authors can publish their deep learning algorithms through various hosting websites (eg, GitHub). If an author published the code associated with journal article, it is designated "public" under the "Published Code" column. If code for the deep learning algorithm was not published, then it is labeled "private." ADC = apparent diffusion coefficient, CNN = convolutional neural network, DCE = dynamic contrast-enhanced, DSC = Dice similarity coefficient, DWI = diffusion-weighted imaging, GE = General Electric, PDW = proton density–weighted, PZ = peripheral zone of the prostate, T1CE = T1-weighted contrast-enhanced, T1W = T1-weighted, T2W = T2-weighted, TZ = transition zone of the prostate, WG = whole prostate gland.

Most studies used T2-weighted imaging to segment the prostate gland (32 of 48). The remaining studies used a combination of T2-weighted, proton density–weighted, dynamic contrast-enhanced imaging, diffusion-weighted imaging, apparent diffusion coefficient, T1-weighted, and T1-weighted contrast-enhanced (11 of 33) imaging. There was no evidence of a difference in DSC between algorithms that were trained strictly on T2-weighted images (0.911 ± 0.035) and those trained on images with multiple MRI sequences (0.912 ± 0.018) ($P$ = .95). Most studies included MRI data with endorectal coils (32 of 48 studies) and fewer used MRI with surface coils (16 of 48 studies). There was no evidence for difference in DSC between

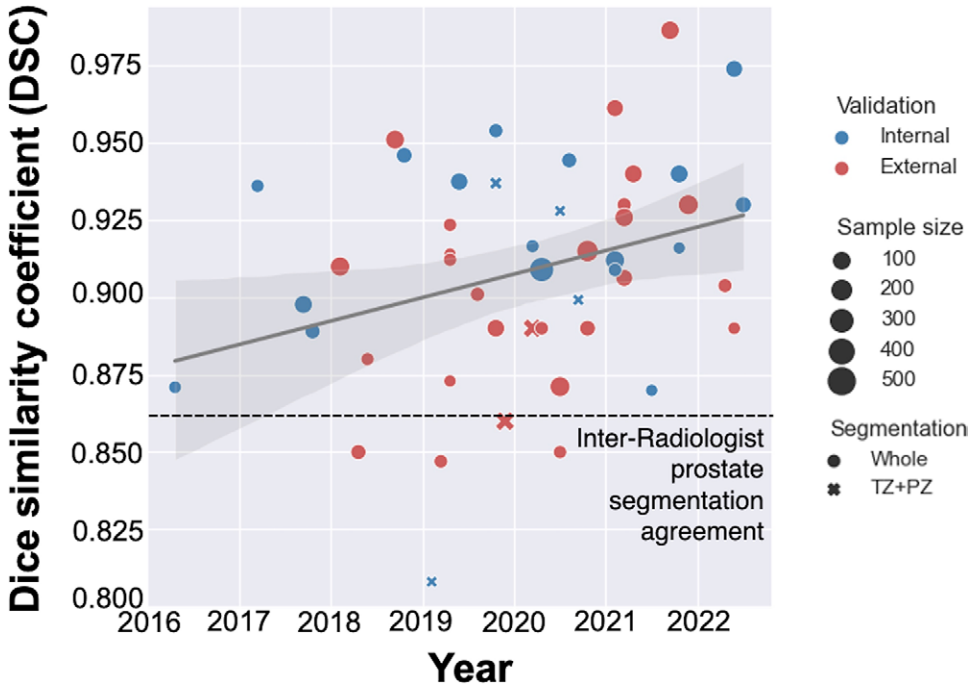


**Figure 4:** Scatterplot shows the mean Dice similarity coefficient (DSC) from each published prostate MRI segmentation deep learning algorithm plotted over time. The size of each data point corresponds to the number of individuals in the deep learning algorithm training set. The color of each data point corresponds to whether a model was externally tested or internally tested. With external testing, a test dataset is held out from training. For internal testing, a model is tested with a subset of data from the original training data. The reported DSC between expert radiologist segmentations is approximately 0.860 (7) and is plotted as a dashed line on the scatterplot. Segmentation was categorized as whole gland if the deep learning algorithm segmented the entire prostate or transitional zone and peripheral zone (TZ+PZ) if only the zones of the prostate were segmented. The gray line depicts a linear fit to illustrate improvement over time.

algorithm performance with endorectal coil (0.910 ± 0.035) versus without endorectal coil (0.914 ± 0.021) ($P$ = .75).

### DSC by Vendor

From the selected studies, prostate MRI scans were acquired from one or multiple major MRI vendors (ie, General Electric, Philips, Siemens). For selected studies that used only one major vendor, the mean DSC of each MRI vendor are as follows: General Electric (three of 48 studies), 0.92 ± 0.03; Philips (four of 48 studies), 0.92 ± 0.02; and Siemens (six of 48 studies), 0.91 ± 0.03. The remaining studies used a combination of two major MRI vendors or all three vendors (35 of 48 studies), which resulted in overall greater variability of reported accuracy, as depicted in Figure 5. ANOVA indicated no evidence of a difference in DSC between MRI vendors ($P$ = .53).

### DSC by Anatomic Zone

Deep learning algorithms can be applied to segment the entire prostate gland (whole-gland segmentation) or to segment different zones (peripheral and/or transition zone). From the selected studies, a few authors investigated segmentation of the different prostate zones (15 of 48). Most articles that segmented by prostate zone segmented the transitional zone or peripheral zone (12 of 15); however, a few segmented the central gland and peripheral zone (three of 15). Because the older term "central gland" has been superseded by "transitional zone" as a result of possible confusion with the "central zone," central gland segmentation is henceforth referred to as "transitional zone segmentation." No studies investigated anterior fibromuscular stroma segmentation. The mean DSC for peripheral zone segmentation was 0.787 ± 0.059, whereas for transition zone segmentation, the mean DSC was slightly higher at 0.874 ± 0.052. Both peripheral and transition zone segmentation were lower than mean DSC for whole-gland segmentation (0.905 ± 0.039) (Fig 6).

## Discussion

In this systematic review, we investigated the performance of deep learning algorithms on prostate gland MRI segmentation. We found that approximately 93% of the deep learning algorithms performed this task at or above the level of expert radiologists (threshold DSC, 0.86), despite wide variations in sample size, underlying data source (MRI vendors), and validation methods.

Multiple methods have been used to improve prostate anatomy segmentation, including different data preparation pipelines, altering deep learning network architecture, number of neural network layers, or optimizing training parameters, all of which can influence final performance (10). The first publication on prostate gland MRI segmentation using deep learning implemented an auto-encoder architecture, one of the original architectures in the field (11). Since then, approximately 40% (19 of 48) of the published algorithms for prostate anatomy segmentation have implemented a novel deep learning architecture or added a supplemental feature to a pre-existing neural network

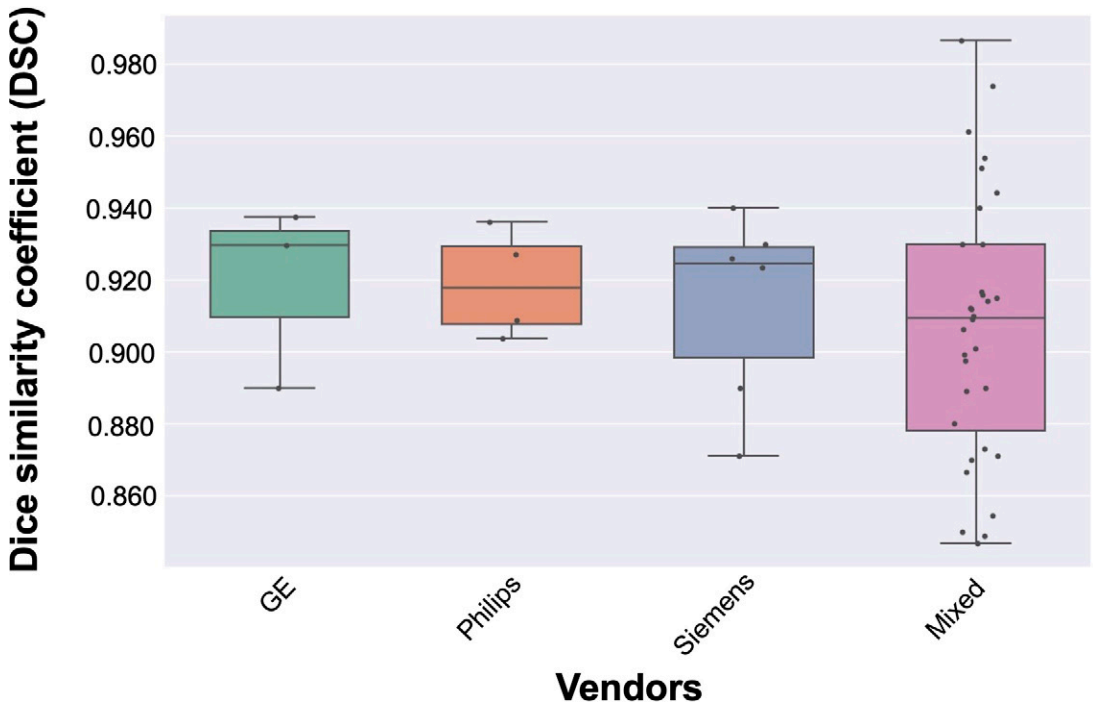


**Figure 5:** Box plots of Dice similarity coefficients (DSCs) for different MRI vendors. Each data point represents the mean DSC from one of the included studies. The data are categorized according to the MRI vendor used to collect the training and test data for each respective study, overlaid by a combined boxplot (middle line represents the median, box margins represent the IQR, and whiskers represent 10th–90th quantile). GE = General Electric.

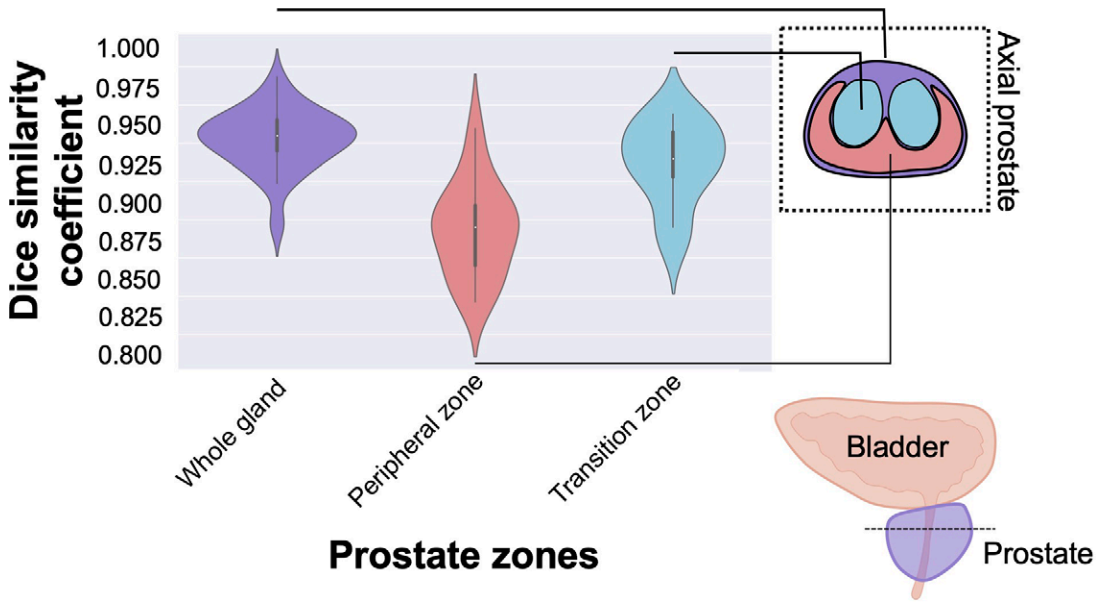


**Figure 6:** Box-and-violin plot of mean reported Dice similarity coefficients (DSCs) of included studies performing whole prostate gland segmentation, peripheral zone segmentation, or transitional zone segmentation. Accuracy of the whole gland was best overall, highlighting that the gland is relatively well delineated. The mean DSC for reference expert radiologist segmentation is 0.859 ± 0.054 for the whole gland and 0.738 ± 0.144 for the transition zone (7).

architecture (eg, generative adversarial networks or active shape models) to improve performance. Implementing different deep learning network architectures or add-ons to existing deep learning algorithms increased the DSC from 0.85 up to a maximum of 0.98 over the course of 6 years, equaling to an annual increase of 0.007 DSC (6,12).

Deep learning is notoriously "data hungry"; however, it is not clear exactly how many samples are needed to train an algorithm for prostate segmentation. In the present analysis, we found multiple deep learning algorithms that segmented prostate anatomy at a clinically acceptable level (DSC > 0.85) with comparably few training data (<50 MRI examinations).

Of note, there have been specific efforts to produce robust segmentations from even smaller sample sizes (13), but such methods have yet to find widespread adoption. In addition, we observed that larger training data only marginally improved deep learning algorithm performance. One reason for this may be that the baseline performance at small samples is already excellent (ie, at or above human level), and any further performance increase follows the law of diminishing returns. Another factor may be that there is less variability in small datasets, so increasing the dataset size, particularly in single-center studies, may slightly degrade performance at the unmeasured benefit of increased robustness.

Finally, there was no evidence of a statistically significant difference in mean DSC between MRI vendors (General Electric, Philips, or Siemens); however, studies that included a combination of MRI vendors had greater variability of DSC than did studies that used data from a single MRI vendor. This variability in performance can be explained by a multitude of known factors because different MRI vendors often have proprietary hardware designs and software algorithms for image acquisition and reconstruction. These differences can lead to variations in image contrast, spatial resolution, and signal-to-noise ratio, which can subsequently affect the performance of deep learning algorithms. Moreover, variations in the magnetic field, often due to differences in hardware design, can lead to artifacts or differences in image intensity.

Previous studies have shown intervendor variability for MRI signal measurements (14,15), although the magnitude of those signal differences may not be clinically relevant (16). Several strategies may be used to mitigate these effects, such as harmonization techniques during preprocessing or incorporating regularization during training to avoid overfitting on data from a single vendor. Nevertheless, our results corroborate that the additional variability introduced by training on multivendor data may slightly degrade segmentation performance. However, it stands to reason that this comes at the benefit of making future segmentations more robust.

Since the advent of deep learning, algorithms have demonstrated remarkable proficiency in segmenting prostate anatomy, often surpassing the accuracy of board-certified diagnostic radiologists. Over the years, substantial resources have been dedicated to refining prostate anatomy segmentation. This has been evident from the numerous online competitions aimed at enhancing performance metrics and the continuous evolution of deep learning network architectures to achieve even marginal improvements in the DSC. However, because these metrics already surpass human experts, it raises the question of the real-world clinical value of pursuing further marginal gains in these arbitrary segmentation metrics.

Rather than solely chasing higher performance metrics, our findings advocate for a shift in research priorities to include the following:

1. Robustness and uncertainty estimation: Future research should delve deeper into ensuring the robustness of segmentation algorithms, especially in challenging scenarios. Techniques for uncertainty estimation (17) can provide clinicians with a measure of confidence in the algorithm's output, which is crucial for clinical decision-making.

2. Prospective evaluations in diverse settings: It is imperative to prospectively evaluate the performance of these algorithms in real-world clinical settings. This includes assessing their robustness across diverse patient populations, variations in MRI vendors and protocols, and even in scenarios with suboptimal image quality. Such evaluations can offer insights into the algorithm's generalizability and its potential pitfalls.

3. Clinical integration and patient outcomes: Emphasis should also be placed on the tangible benefits of automated prostate anatomy segmentation in the clinical workflow. This involves assessing not just the algorithm's accuracy but also its effect on patient outcomes, workflow efficiency, and overall clinical utility. Ideally, the application of such segmentations (eg, in radiation therapy planning or focal therapy application) should be evaluated in a randomized prospective fashion.

Prostate MRI deep learning algorithms hold substantial promise in augmenting and improving real-world prostate cancer management. Automated methods for prostate gland segmentation can increase the efficiency and accuracy of MRI/US fusion–targeted biopsies. Although targeted prostate biopsy introduces the advantage of coregistering MRI and US images, the approach can still be prone to operator error or bias in MRI gland and tumor segmentation (1–3). The use of MRI for computer-assisted surgery continues to be established, but there may also be an important role for deep learning–based prostate anatomy segmentation with regard to intraoperative nerve sparing or functional outcome prediction (4).

To facilitate future research focusing less on narrowly optimizing arbitrary similarity metrics, it may be useful to establish an independent service to evaluate algorithms on truly novel and unseen test data. Such a service could require preregistering the study or algorithm (eg, connecting it with an arXiv ID and/or code repository) and would only allow testing of any given algorithm once. This could be tied to a small payment that would first cover operating costs and second discourage multiple iterative tests ("training on the test set").

This systematic review of prostate gland segmentation had some limitations. The number of selected studies was relatively small for segmentation of individual prostate zones ($n$ = 15). In addition, there was inherent heterogeneity of those data, considering that some studies included private data in their analysis; therefore, segmentation may not be standardized across all studies. Moreover, only articles indexed in the databases Embase, PubMed, Scopus, and Web of Science were included. There may be additional relevant articles published (eg, on arXiv or in more technically oriented publications), which these databases may not capture.

In conclusion, we found that most deep learning algorithms performed at or above the level of human experts for the task of prostate segmentation when measured using DSC. Future work should focus on segmentation robustness and investigate whether further accuracy gains actually translate to improvements in patient outcomes.

**Author contributions:** Guarantors of integrity of entire study, **M.K.F., A.S.B.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **M.K.F., A.B., S.W., A.S.B.**; clinical studies, **A.B.**; experimental studies, **A.B.**; statistical analysis, **M.K.F., A.B., S.W., A.S.B.**; and manuscript editing, **M.K.F., A.B., S.W., H.A.V., E.K., A.S.B.**

**Data sharing:** Data generated or analyzed during the study are available from the corresponding author by request.

**Disclosures of conflicts of interest: M.K.F.** The Department of Radiology (New York-Presbyterian Weill Cornell Medical Center) receives funding from the Peter Michael Foundation. **A.B.** Consulting fees from Johnson & Johnson; support for travel from Johnson & Johnson. **S.W.** No relevant relationships **H.A.V.** No relevant relationships. **H.H.** Funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748 (payments made to Memorial Sloan Kettering Cancer Center); External Advisory Board of the Sidney Kimmel Comprehensive Cancer Center (SKCCC) at Johns Hopkins (no compensation received); International Advisory Board of the University of Vienna, Austria (no compensation received); Scientific Committee of the DKFZ (German Cancer Research Center), Germany (no compensation received); Board of Directors of Ion Beam Applications (IBA) (receives compensation); Board of Trustees of the German Cancer Research Center (DKFZ) (no compensation received); Board of Directors of Paige (no compensation received). **E.K.** No relevant relationships. **A.S.B.** No relevant relationships.

## References


1. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. J Urol 2013;190(2):419–426.
2. Ghafoor S, Becker AS, Woo S, et al. Comparison of PI-RADS versions 2.0 and 2.1 for MRI-based calculation of the prostate volume. Acad Radiol 2021;28(11):1548–1556.
3. Barrett T, Lawrence EM, Priest AN, et al. Repeatability of diffusion-weighted MRI of the prostate using whole lesion ADC values, skew and histogram analysis. Eur J Radiol 2019;110:22–29.
4. Liechti MR, Muehlematter UJ, Schneider AF, et al. Manual prostate cancer segmentation in MRI: interreader agreement and volumetric correlation with transperineal template core needle biopsy. Eur Radiol 2020;30(9): 4806–4815.
5. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264–269, W64.
6. Guo Y, Gao Y, Shen D. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. IEEE Trans Med Imaging 2016;35(4):1077–1089.
7. Becker AS, Chaitanya K, Schawkat K, et al. Variability of manual segmentation of the prostate in axial T2-weighted MRI: a multi-reader study. Eur J Radiol 2019;121:108716
8. Montagne S, Hamzaoui D, Allera A, et al. Challenge of prostate MRI segmentation on T2-weighted images: inter-observer variability and impact of prostate morphology. Insights Imaging 2021;12(1):71.
9. Molière S, Hamzaoui D, Granger B, et al. Reference standard for the evaluation of automatic segmentation algorithms: quantification of inter observer variability of manual delineation of prostate contour on MRI. Diagn Interv Imaging 2024;105(2):65–73.
10. Rodrigues NM, Silva S, Vanneschi L, Papanikolaou N. A comparative study of automated deep learning segmentation models for prostate MRI. Cancers (Basel) 2023;15(5):1467.
11. Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. IEEE Access 2019;7:53040–53065.
12. Yan L, Liu D, Xiang Q, et al. PSP net-based automatic segmentation network model for prostate magnetic resonance imaging. Comput Methods Programs Biomed 2021;207:106211.
13. Chaitanya K, Karani N, Baumgartner CF, et al. Semi-supervised task-driven data augmentation for medical image segmentation. Med Image Anal 2021;68:101934.
14. Keenan KE, Gimbutas Z, Dienstfrey A, et al. Multi-site, multi-platform comparison of MRI T1 measurement using the system phantom. PLoS One 2021;16(6):e0252966.
15. Lee Y, Callaghan MF, Acosta-Cabronero J, Lutti A, Nagy Z. Establishing intra- and inter-vendor reproducibility of $T_1$ relaxation time measurements with 3T MRI. Magn Reson Med 2019;81(1):454–465.
16. Donati OF, Chong D, Nanz D, et al. Diffusion-weighted MR imaging of upper abdominal organs: field strength and intervendor variability of apparent diffusion coefficients. Radiology 2014;270(2):454–463.
17. Abdar M, Pourpanah F, Hussain S, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Inf Fusion 2021;76:243–297.
18. Agarwal A, Mishra A, Basavarajaiah M, Sharma P, Tanwar S. Dilated volumetric network: an enhanced fully convolutional network for volumetric prostate segmentation from magnetic resonance imaging. Pattern Recognit Image Anal 2021;31(2):228–239.
19. Astono IP, Welsh JS, Chalup S, Greer P. Optimisation of 2D U-net model components for automatic prostate segmentation on MRI. Appl Sci (Basel) 2020;10(7):2601.
20. Bardis M, Houshyar R, Chantaduly C, et al. Segmentation of the prostate transition zone and peripheral zone on MR images with deep learning. Radiol Imaging Cancer 2021;3(3):e200024.
21. Cheng R, Lay N, Roth HR, et al. Fully automated prostate whole gland and central gland segmentation on MRI using holistically nested networks with short connections. J Med Imaging (Bellingham) 2019;6(2):024007.
22. Cheng R, Roth HR, Lay N, et al. Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks. J Med Imaging (Bellingham) 2017;4(4):041302.
23. Clark T, Zhang J, Baig S, Wong A, Haider MA, Khalvati F. Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. J Med Imaging (Bellingham) 2017;4(4):041307.
24. Comelli A, Dahiya N, Stefano A, et al. Deep learning-based methods for prostate segmentation in magnetic resonance imaging. Appl Sci (Basel) 2021;11(2):782.
25. Cuocolo R, Comelli A, Stefano A, et al. Deep learning whole-gland and zonal prostate segmentation on a public MRI dataset. J Magn Reson Imaging 2021;54(2):452–459.
26. da Silva GLF, Diniz PS, Ferreira JL, et al. Superpixel-based deep convolutional neural networks and active contour model for automatic prostate segmentation on 3D MRI scans. Med Biol Eng Comput 2020;58(9):1947–1964.
27. Geng L, Wang J, Xiao Z, Tong J, Zhang F, Wu J. Encoder-decoder with dense dilated spatial pyramid pooling for prostate MR images segmentation. Comput Assist Surg (Abingdon) 2019;24(sup2):13–19.
28. Ghavami N, Hu Y, Gibson E, et al. Automatic segmentation of prostate MRI using convolutional neural networks: Investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration. Med Image Anal 2019;58:101558.
29. Hassanzadeh T, Hamey LGC, Ho-Shon K. Convolutional neural networks for prostate magnetic resonance image segmentation. IEEE Access 2019;7:36748–36760.
30. Jensen C, Sørensen KS, Jørgensen CK, et al. Prostate zonal segmentation in 1.5T and 3T T2W MRI using a convolutional neural network. J Med Imaging (Bellingham) 2019;6(1):014501.
31. Jia H, Xia Y, Song Y, et al. 3D APA-Net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images. IEEE Trans Med Imaging 2020;39(2):447–457.
32. Jia H, Xia Y, Song Y, Cai W, Fulham M, Feng DD. Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging. Neurocomputing 2018;275: 1358–1369.
33. Jia H, Cai W, Huang H, Xia Y. Learning multi-scale synergic discriminative features for prostate image segmentation. Pattern Recognit 2022;126:108556.
34. Jin Y, Yang G, Fang Y, et al. 3D PBV-Net: an automated prostate MRI data segmentation method. Comput Biol Med 2021;128:104160.
35. Karimi D, Samei G, Kesch C, Nir G, Salcudean SE. Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. Int J CARS 2018;13(8):1211–1219.
36. Khan Z, Yahya N, Alsaih K, Ali SSA, Meriaudeau F. Evaluation of deep neural networks for semantic segmentation of prostate in T2W MRI. Sensors (Basel) 2020;20(11):3183.
37. Lee DK, Sung DJ, Kim CS, et al. Three-dimensional convolutional neural network for prostate MRI segmentation and comparison of prostate volume

measurements by use of artificial neural network and ellipsoid formula. AJR Am J Roentgenol 2020;214(6):1229–1238.
38. Liu Y, Miao Q, Surawech C, et al. Deep learning enables prostate MRI segmentation: a large cohort evaluation with inter-rater variability analysis. Front Oncol 2021;11:801876.
39. Liu Q, Fu M, Jiang H, Gong X. Densely dilated spatial pooling convolutional network using benign loss functions for imbalanced volumetric prostate segmentation. Curr Bioinform 2020;15(7):788–799.
40. Liu Y, Yang G, Hosseiny M, et al. Exploring uncertainty measures in Bayesian deep attentive neural networks for prostate zonal segmentation. IEEE Access 2020;8:151817–151828.
41. Liu Q, Dou Q, Yu L, Heng PA. MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Trans Med Imaging 2020;39(9):2713–2724.
42. Liu Y, Sung K, Yang G, et al. Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention. IEEE Access 2019;7:163626–163632.
43. Meyer A, Chlebus G, Rak M, et al. Anisotropic 3D multi-stream CNN for accurate prostate segmentation from multi-planar MRI. Comput Methods Programs Biomed 2021;200:105821.
44. Nai YH, Teo BW, Tan NL, et al. Evaluation of multimodal algorithms for the segmentation of multiparametric MRI prostate images. Comput Math Methods Med 2020;2020:8861035.
45. Rouvière O, Moldovan PC, Vlachomitrou A, et al. Combined model-based and deep learning-based automated 3D zonal segmentation of the prostate on T2-weighted MR images: clinical evaluation. Eur Radiol 2022;32(5):3248–3259.
46. Rundo L, Han C, Nagano Y, et al. USE-Net: incorporating squeeze-and-excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. Neurocomputing 2019;365:31–43.
47. Sanford TH, Zhang L, Harmon SA, et al. Data augmentation and transfer learning to improve generalizability of an automated prostate segmentation model. AJR Am J Roentgenol 2020;215(6):1403–1410.
48. Salvaggio G, Comelli A, Portoghese M, et al. deep learning network for segmentation of the prostate gland with median lobe enlargement in T2-weighted MR images: comparison with manual segmentation method. Curr Probl Diagn Radiol 2022;51(3):328–333.
49. Salvi M, De Santi B, Pop B, et al. Integration of deep learning and active shape models for more accurate prostate segmentation in 3D MR images. J Imaging 2022;8(5):133.
50. Sunoqrot MRS, Selnæs KM, Sandsmark E, et al. The reproducibility of deep learning-based segmentation of the prostate gland and zones on T2-weighted MR images. Diagnostics (Basel) 2021;11(9):1690.
51. Tao L, Ma L, Xie M, Liu X, Tian Z, Fei B. Automatic segmentation of the prostate on MR images based on anatomy and deep learning. Proc SPIE Int Soc Opt Eng 2021;11598:115981N.
52. Tian Z, Li X, Zheng Y, et al. Graph-convolutional-network-based interactive prostate segmentation in MR images. Med Phys 2020;47(9):4164–4176.
53. Tian Z, Liu L, Zhang Z, Fei B. PSNet: prostate segmentation on MRI based on a convolutional neural network. J Med Imaging (Bellingham) 2018;5(2):021208.
54. To MNN, Vu DQ, Turkbey B, Choyke PL, Kwak JT. Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging. Int J CARS 2018;13(11):1687–1696.
55. Ushinsky A, Bardis M, Glavis-Bloom J, et al. A 3D-2D hybrid U-Net convolutional neural network approach to prostate organ segmentation of multiparametric MRI. AJR Am J Roentgenol 2021;216(1):111–116.
56. Wang W, Wang G, Wu X, et al. Automatic segmentation of prostate magnetic resonance imaging using generative adversarial networks. Clin Imaging 2021;70:1–9.
57. Wang B, Lei Y, Tian S, et al. Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation. Med Phys 2019;46(4):1707–1718.
58. Yan K, Wang X, Kim J, Khadra M, Fulham M, Feng D. A propagation-DNN: deep combination learning of multi-level features for MR prostate segmentation. Comput Methods Programs Biomed 2019;170:11–21.
59. Zabihollahy F, Schieda N, Krishna Jeyaraj S, Ukwatta E. Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. Med Phys 2019;46(7):3078–3090.
60. Zavala-Romero O, Breto AL, Xu IR, et al. Segmentation of prostate and prostate zones using deep learning: a multi-MRI vendor analysis. Strahlenther Onkol 2020;196(10):932–942.
61. Zhu Q, Du B, Yan P. Boundary-weighted domain adaptive neural network for prostate MR image segmentation. IEEE Trans Med Imaging 2020;39(3):753–763.
62. Zhu Y, Wei R, Gao G, et al. Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. J Magn Reson Imaging 2019;49(4):1149–1156.
63. Zhu Q, Du B, Turkbey B, Choyke P, Yan P. Exploiting interslice correlation for MRI prostate image segmentation, from recursive neural networks aspect. Complexity 2018;2018:4185279.