Deep Learning Improves Speed and Accuracy of Prostate Gland Segmentations on Magnetic Resonance Imaging for Targeted Biopsy

Simon John Christoph Soerensen, Richard E. Fan, Arun Seetharaman et al.

Study Need and Importance: Targeted biopsy improves prostate cancer diagnosis. Accurately outlining the prostate on magnetic resonance imaging (MRI) is critical for accurate biopsy, yet manual gland segmentation is tedious and time-consuming. To address this clinical problem, we sought to develop a deep learning model to rapidly and accurately segment the prostate on MRI and to implement it as part of routine magnetic resonance-ultrasonography (MR-US) fusion biopsy in the clinic.

What We Found: We trained a deep learning model, ProGNet, on 805 T2-weighted MRI scans and retrospectively tested it on 100 independent internal cases and 56 external cases. Segmentations from a urologic oncology expert were used as ground-truth labels. We compared ProGNet performance to two deep learning networks (U-Net and HED) and radiology technicians using the Dice similarity coefficient (DSC). We then prospectively implemented ProGNet as part of the fusion biopsy workflow for 11 patients. ProGNet (DSC [0.92±0.02]) significantly outperformed radiology technicians (DSC [0.89±0.05]) and 2 other deep learning models, U-Net (DSC =0.85±0.06) and HED (DSC=0.80±0.08). ProGNet took just 35 seconds per case (vs 10 minutes for radiology technicians) to yield a clinically usable segmentation file.

Limitations: We compared ProGNet model outputs to nonphysician trained radiology technicians (our institution’s workflow). The findings remain relevant to other institutions where physicians perform segmentations because the model provides quick and precise segmentations very similar to the urologic oncology expert. Also, the approach is not immediately accessible for widespread clinical use. Work is ongoing to integrate the model into routine clinical workflow for all cases at our institution and then expand elsewhere to improve speed and accuracy of prostate segmentation for targeted biopsy (see figure).

Interpretation for Patient Care: Our deep learning model segmented the prostate more accurately and 17 times faster than humans. With the successful implementation of our model into clinical practice, we reduced error in one critical step of the targeted biopsy process.
Deep Learning Improves Speed and Accuracy of Prostate Gland Segmentations on Magnetic Resonance Imaging for Targeted Biopsy

Simon John Christoph Soerensen, Richard E. Fan, Arun Seetharaman, Leo Chen, Wei Shao, Indrani Bhattacharya, Yong-hun Kim, Rewa Sood, Michael Borre, Benjamin I. Chung,* Katherine J. To'o, Mirabela Rusu†,‡ and Geoffrey A. Sonn§,†

From the Department of Urology (SJCS, REF, LC, BIC, GAS), Stanford University School of Medicine, Stanford, California, Department of Urology (SJCS, MB), Aarhus University Hospital, Aarhus, Denmark, Department of Electrical Engineering (AS, RS), Stanford University, Stanford, California, Department of Radiology (WS, IB, KJT, MR, GAS), Stanford University School of Medicine, Stanford, California, Department of Computer Science (YK), Stanford University, Stanford, California, and Veterans Affairs (KJT), Palo Alto Health Care System, Palo Alto, California

Purpose: Targeted biopsy improves prostate cancer diagnosis. Accurate prostate segmentation on magnetic resonance imaging (MRI) is critical for accurate biopsy. Manual gland segmentation is tedious and time-consuming. We sought to develop a deep learning model to rapidly and accurately segment the prostate on MRI and to implement it as part of routine magnetic resonance-ultrasound fusion biopsy in the clinic.

Materials and Methods: A total of 905 subjects underwent multiparametric MRI at 29 institutions, followed by magnetic resonance-ultrasound fusion biopsy at 1 institution. A urologic oncology expert segmented the prostate on axial T2-weighted MRI scans. We trained a deep learning model, ProGNet, on 805 cases. We retrospectively tested ProGNet on 100 independent internal and 56 external cases. We prospectively implemented ProGNet as part of the fusion biopsy procedure for 11 patients. We compared ProGNet performance to 2 deep learning networks (U-Net and holistically-nested edge detector) and radiology technicians. The Dice similarity coefficient (DSC) was used to measure overlap with expert segmentations. DSCs were comparing paired t-tests.

Results: ProGNet (DSC = 0.92) outperformed U-Net (DSC = 0.85, p < 0.0001), holistically-nested edge detector (DSC = 0.80, p < 0.0001), and radiology technicians (DSC = 0.89, p < 0.0001) in the retrospective internal test set. In the prospective cohort, ProGNet (DSC = 0.93) outperformed radiology technicians (DSC = 0.90, p < 0.0001). ProGNet took just 35 seconds per case (vs 10 minutes for radiology technicians) to yield a clinically utilizable segmentation file.

Conclusions: This is the first study to employ a deep learning model for prostate gland segmentation for targeted biopsy in routine urological clinical practice, while reporting results and releasing the code online. Prospective and retrospective evaluations revealed increased speed and accuracy.

Key Words: deep learning, magnetic resonance imaging, imaging-guided biopsy, ultrasonography

Abbreviations and Acronyms

2D = 2-dimensional
3D = 3-dimensional
DSC = Dice similarity coefficient
HED = holistically-nested edge detector
MRI = magnetic resonance imaging
MR-US = magnetic resonance-ultrasound

Accepted for publication February 28, 2021. This work was supported by Stanford University (Departments of Radiology and Urology) and by the generous philanthropic support of donors to the Urologic Cancer Innovation Laboratory at Stanford University.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

* Financial and/or other relationship with Intuitive Surgical and Ethicon.
† Equal study contribution.
‡ Financial and/or other relationship with GE Healthcare, Philips Healthcare, and National Institutes of Health.
§ Correspondence: Department of Urology, Stanford University School of Medicine, 300 Pasteur Dr. 5287, Stanford, California 94305 (telephone: 650-733-5586; FAX: 650-498-5346; email: gsonn@stanford.edu)
weighted MRI (T2-MRI). Providing prostate segmentations on T2-MRI is both tedious and time-consuming. Clinical implementation of an automated method to accurately segment the prostate on T2-MRI will save substantial time for urologists and radiologists while potentially improving biopsy accuracy.

Recent advancements in deep learning have enabled deep neural networks to rapidly perform medical imaging analysis tasks. Achieving generalizable results requires large amounts of training data from multiple institutions. Different methods have been proposed to automate prostate gland segmentation but have often used small data sets (usually 40–250 cases). did not use volumetric context from adjacent T2-MRI slices to make predictions. failed to evaluate on external cohorts solely used single-institution training sets did not release code for comparison or did not publish model accuracy.

Deep learning for medical applications has rarely—and never for the essential prostate segmentation task—been integrated into clinical practice, while reporting results and releasing the code online. Our objective was to develop a deep learning model, ProGNet, to segment the prostate rapidly and accurately on T2-MRI prior to magnetic resonance-ultrasound (MR-US) fusion targeted biopsy. To promote clinical utilization, we aimed to integrate the deep learning model into our clinical workflow as part of fusion biopsy and share our code online.

**MATERIALS AND METHODS**

**Patient Selection**

A total of 916 men underwent multiparametric MRI at 29 academic or private practice institutions in the U.S. in 2013–2019, followed by fusion targeted biopsy at Stanford University. Consent for data collection prior to biopsy was obtained under IRB-approved protocols (IRB No. IRB-57842), and the data registry was Health Insurance Portability and Accountability Act (HIPAA) compliant. Subjects included for real-time biopsy in the prospective cohort consented as part of an additional IRB-protocol that enabled the use of ProGNet in their clinical care.

**Magnetic Resonance Imaging**

We collected axial T2-MRI for all men in the study. Of the men in the study 85% underwent multiparametric MRI at Stanford University (vs 15% elsewhere) on GE (GE Healthcare, Waukesha, Wisconsin, 88%), Siemens (Siemens Healthcare, Erlangen, Germany, 10%), or Philips (Philips Healthcare, Amsterdam, Netherlands, 2%) scanners. Scans were performed at 1.5 Tesla (2%) or 3 Tesla (98%) using multichannel external body array coils. Most scans included both 2D and 3D T2 sequences. Protocol features relevant to 2D T2-MRI can be found in table 1, as that was the sequence we used for training and testing the deep learning segmentation model.

**Table 1. Data summary of 2D T2-MRI in internal training and test sets**

| Internal Data Set—MRI Characteristics | Total No. | No. data set composition: |
|--------------------------------------|----------|--------------------------|
| Total No.                            | 916      | Training 805             |
|                                      |          | Retrospective testing 100|
|                                      |          | Prospective testing 11   |
| % Institution:                       |          | Stanford University 85   |
|                                      |          | 28 other institutions 15 |
| % Scanner:                           |          | GE 93                    |
|                                      |          | Siemens 4                |
|                                      |          | Philips 3                |
| Weighting                            |          | T2                       |
| Direction                            |          | Axial                    |
| % MRI sequence:                      |          | Spin-echo 99.7           |
|                                      |          | Research mode 0.2         |
|                                      |          | Echo planar spin echo 0.1|
| % Magnetic field strength:           |          | 3T 98                    |
|                                      |          | 1.5T 2                   |
| Median slice thickness in mm (IQR)   |          | 3.6 (3.55–4.2)           |
| In-plane resolution in mm (range)    |          | 0.27×0.27–0.94×0.94      |
| Median No. slices (IQR)              |          | 25 (25–30)               |
| Matrix size in pixels×pixels (range) |          | 256×256–640×640          |
| Median echo time (IQR)               |          | 126 (122–128)            |

**Classical Pre-Fusion Biopsy Procedure**

Fusion biopsy was performed at Stanford University using the Artemis device (Eigen, Grass Valley, California). Following our institutional protocol, 7 trained radiology technicians, with a mean experience of 9 years, segmented the prostate on axial T2-MRI using ProFuse software (Eigen, Grass Valley, California). Body MRI radiologists and fellows provided feedback to help improve segmentations. Immediately prior to biopsy, a urologic oncology expert (GAS) with 7 years of experience with MR-US fusion targeted biopsy refined the gland segmentations in ProFuse.

**Data Sets**

We randomly split T2-MRI from the 905 subjects who underwent MR-US fusion biopsy at Stanford University into a training set (805) and an independent internal retrospective test set (100). Eleven additional cases were evaluated prospectively. Segmentations from a urologic oncology expert (GAS) were used as ground-truth labels for training and testing. To obtain more diverse testing data, we included T2-MRI acquired on Siemens scanners at 1.5 or 3 Tesla from 2 publicly available data sets, PROMISE1222 (26) and NCI-ISBI23 (30). Both data sets included expert segmentations of the prostate.

**Deep Learning Pre-Processing.** All axial T2-MRI were automatically cropped to a 256×256 matrix, as this invariably included the entire prostate and is the input utilized by our model. All individual scans had the same pixel resolution right-to-left and anterior-posterior. A histogram-based intensity standardization method was automatically applied to normalize pixel intensities, which vary in T2-MRI from...
various institutions.\textsuperscript{24,25} The training set was then augmented by flipping the T2-MRI scans left-to-right.\textsuperscript{26}

**ProGNet Architecture.** Our deep learning model, ProGNet, is a novel convolutional neural network for prostate segmentation on T2-MRI based on the U-Net architecture (fig. 1).\textsuperscript{27} ProGNet integrates information from 3 consecutive T2-MRI slices and predicts segmentations on the middle slice, thereby learning the “2.5D” volumetric continuity of the prostate on MRI. This approach of considering adjacent slices together, rather than in isolation, is much more analogous to how experts interpret images in the clinical setting.

Unlike existing methods,\textsuperscript{17,19,20} ProGNet automatically refines predicted segmentations to ensure spatial and volumetric continuity using robust post-processing steps. First, predictions that are not connected to the prostate are removed. Second, a Gaussian filter (sigma = 1) smoothen segmentation borders. Third, the most apical predictions are removed if they are \(<15\) mm in diameter (a sign of the model segmenting into the membranous urethra or penis).

**Deep Learning Experiments.** We compared ProGNet prostate segmentation performance to 2 common deep learning networks: the U-Net and the holistically-nested edge detector (HED).\textsuperscript{10,27} All models were trained for 150 epochs using an NVIDIA V100 graphics card and the TensorFlow 2.0 deep learning framework. We trained and tested the U-Net and HED on the same internal retrospective cases as the ProGNet model.

**Clinical Implementation**

We prospectively used ProGNet for 11 consecutive targeted biopsy cases to demonstrate our approach’s clinical utility. The expert urologist (GAS) modified the ProGNet segmentations prior to biopsy in a real-world setting as part of the usual standard of care. The ProGNet code can be downloaded at \url{http://med.stanford.edu/ucl/GlandSegmentation.html}.

The ProGNet code is easily run by users without coding experience on as many MRI cases as desired without any manual processing. It outputs T2-DICOM (Digital Imaging and Communications in Medicine) folders with both the T2-MRI and a segmentation file that users load into the biopsy software.

**Statistical Analysis**

We compared ProGNet and radiology technicians’ performances in the prospective and retrospective cohorts by comparing segmentation overlap with the expert using the Dice similarity coefficient (DSC). The DSC is widely used to evaluate overlap in segmentation tasks, and its value ranges from 0 to 1; 1 indicates perfect overlap between segmentations, while 0 indicates no overlap. We compared our model’s performance in the internal test sets to 2 deep learning networks, the U-Net and HED. In each test set, DSCs for radiology technicians, U-Net, & HED were compared to DSCs for ProGNet using Bonferroni-corrected paired t-tests. In an attempt to determine how gland segmentation accuracy may impact the location of the target, we also applied the Hausdorff distance metric to compare ProGNet and radiology technician segmentation errors. We defined a 2-sided \( p < 0.05 \) as the threshold for statistical significance. Results were expressed as mean ± standard deviation. We calculated speed of ProGNet (time spent opening & running the automatic ProGNet code) and radiology technicians (time spent segmenting in the ProFuse software) in the retrospective internal test set.
RESULTS

Retrospective Internal Test Set
In the retrospective multisite internal test set, ProGNet (mean DSC = 0.92 ± 0.02) outperformed the U-Net (mean DSC = 0.85 ± 0.06, \( p < 0.0001 \)) and HED (mean DSC = 0.80 ± 0.08, \( p < 0.0001 \)) deep learning models. ProGNet exceeded the segmentation performance of experienced radiology technicians (mean DSC = 0.92 ± 0.02 vs DSC = 0.89 ± 0.05, \( p < 0.0001 \); table 2 and fig. 2). Comparing gland segmentation error, the ProGNet model reduced the mean Hausdorff distance by 2.8 mm compared to the radiology technicians.

ProGNet also delivered the highest level of precision in segmentation as defined by a narrow range in DSC (fig. 3) and the proportion of cases with a DSC ≥ 0.90. The DSC was ≥ 0.90 in 88% of ProGNet cases, compared to 27% for U-Net, 8% for HED, and 61% for radiology technicians.

In a sensitivity analysis, we split the retrospective internal test set into scans obtained at Stanford University (88) vs elsewhere (12) and observed that ProGNet outperformed U-Net, HED, and radiology technicians both on scans obtained at Stanford and elsewhere (table 3).

External Test Sets
Given most T2-MRI scans in our training and test sets came from 1 institution and were acquired on GE scanners, to further evaluate generalizability, we assessed ProGNet performance on 2 publicly available data sets consisting solely of Siemens scans. ProGNet achieved a mean DSC of 0.87 ± 0.05 on MRI scans from the PROMISE12 data set (26, fig. 4). In the NCI-ISBI data set (30), ProGNet achieved a mean DSC of 0.89 ± 0.05. As shown in table 2, ProGNet’s performance on external data is consistent with results acquired on the internal data and outperforms both HED and U-Net.

Segmentation Time
After a single 20-hour training session, it took ProGNet approximately 35 seconds to segment each case in the 100-case retrospective internal test set (~1 hour in total). Conversely, radiology technicians averaged 10 minutes per case (~17 hours in total). This does not account for the additional time involved in adjusting the segmentations by the expert urologist (range: 3–7 minutes per case).

PROSPECTIVE EVALUATION
To demonstrate this approach’s feasibility in clinical practice, we successfully integrated ProGNet into our clinical workflow. ProGNet (mean DSC = 0.93 ± 0.03) significantly outperformed radiology technicians (mean DSC = 0.90 ± 0.03, \( p < 0.0001 \)) in the 11-case prospective fusion biopsy test set.

DISCUSSION
In this study, we developed a robust deep learning model, ProGNet, to automatically segment the prostate on T2-MRI and clinically implemented it as part of real-time fusion targeted biopsy in a prospective cohort. Targeted biopsy involves multiple potential sources of error, such as MRI and ultrasound segmentation, MRI lesion segmentation, MR-US alignment, and patient motion during biopsy. The primary goals of utilizing a deep learning model to segment the prostate are to improve accuracy and speed, and to reduce error in 1 critical step of the biopsy process.

Our study has 4 key findings. First, ProGNet performed significantly better than trained radiology technicians and 2 state-of-the-art prostate segmentation networks in multiple independent testing cohorts. Importantly, ProGNet had far fewer poorly performing outlier cases (1 in 8 cases with DSC < 0.90) than radiology technicians (1 in 3 cases). Having fewer poorly performing cases translates into less time spent by a urologist refining the segmentation prior to biopsy.

Second, the speed of segmentation was approximately 17 times faster for ProGNet than radiology technicians; ProGNet saved ~16 hours of segmentation time in the 100-case test set alone. This does not even account for the additional time the expert urologist spends adjusting inaccurate segmentations before biopsy.

Third, ProGNet performed better than or equal to other prostate segmentation models.10–14,16–20 The generalizability of ProGNet results from the large

### Table 2. Deep learning and radiology technician prostate MRI segmentation performances (mean DSC ± SD) in internal and external test sets

|                      | Prospective Internal Test Set | Retrospective Internal Test Set | PROMISE12 External Test Set | NCI-ISBI External Test Set |
|----------------------|-------------------------------|--------------------------------|-----------------------------|---------------------------|
| No. cases            | 11                            | 100                            | 26                          | 30                        |
| ProGNet              | 0.93 (± 0.03)                 | 0.92 (± 0.02)                  | 0.87 (± 0.05)               | 0.89 (± 0.05)             |
| U-Net                | 0.83 (± 0.01)                 | 0.85 (± 0.06)                  | 0.85 (± 0.08)               | 0.86 (± 0.07)             |
| HED                  | 0.78 (± 0.10)                 | 0.80 (± 0.08)                  | 0.78 (± 0.13)               | 0.80 (± 0.11)             |
| Radiology technicians| 0.90 (± 0.04)                 | 0.89 (± 0.05)                  |                             |                           |

DSCs of U-Net and HED deep learning models, as well as radiology technicians, were compared with DSCs of ProGNet in each test set using Bonferroni-corrected paired t-tests. All tests showed statistical significance (\( p < 0.0001 \)). Bolded entries represent highest mean dice score in each test set.
training (805) and testing (167) cohorts. Prior publications typically included only 40–250 cases. ProGNet performed well on internal and external cohorts comprised of scans from GE, Siemens, and Philips acquired at multiple institutions with different magnet strengths. It is important to note that lack of access to code prevented us from directly comparing prior methods to ProGNet in our independent test sets. Instead, we compared ProGNet to the U-Net and HED deep learning models commonly used for prostate gland segmentation and trained those models ourselves.10,11

Fourth, to our knowledge, this is the first study to clinically implement a deep learning model to segment the prostate on MRI prior to fusion biopsy in a live setting, while reporting results and releasing the code online. Commercial vendors such as Philips DynaCAD automate segmentation for clinical use, but this is only available to those who purchase that software. It is unclear how well DynaCAD performs due to its use of proprietary software and lack of description of its performance using metrics such as the Dice score.21 We have also released our code publicly so that researchers,
companies, or clinicians can easily test or implement our model. Finally, we put great effort into enabling our model outputs to be implemented using Eigen’s ProFuse software; we envision future integration with other targeted biopsy vendors.

Our study has 5 noteworthy limitations. First, while ProGNet statistically significantly outperformed 2 deep learning models and radiology technicians using the Dice score metric, it is unclear if this translates into clinically significantly better targeting of suspicious lesions. Our analysis indicates that use of ProGNet rather than technicians translates into a mean 2.8 mm reduction in error, which may be important in targeting smaller lesions. Second, only 1 experienced urologist (GAS) provided the clinical reference standard using the ProFuse software. While the software does not produce 100% accurate segmentations due to automatic smoothing of the borders, the urologist meticulously corrected each case prior to biopsy as accurately as the software allowed. The model learned to be very accurate due to the extensive training data set, even when it was not provided with perfect segmentations. Using the data available from urologist segmentations during targeted biopsy as ground truth was a pragmatic decision given the difficulty of getting an additional expert to segment almost 1,000 cases. Our methods considered the urologist as the gold standard, which prevented us from determining if the ProGNet segmentations were more accurate than the urologist’s. Third, rather than comparing model outputs to urologists or radiologists, we compared them to nonphysician trained radiology technicians (the workflow at our institution). The findings remain relevant to other institutions where physicians perform segmentations because of the much greater speed of the ProGNet model and the similarity between the urologic oncology expert’s segmentations and the ProGNet model. Fourth, our data set did not include cases with an endorectal coil, and most of our scans in the training set were performed at 1 institution on scans from 1 manufacturer (GE). However, we found that the deep learning model still performed well on MRIs acquired outside our institution on different scanners. Fifth, our current MRI segmentation approach optimizes only 1 step of the targeted biopsy process. Work is ongoing to automate and optimize other steps in the biopsy process.

Notwithstanding these limitations, our study describes the development and external validation of a deep learning prostate segmentation model whose average accuracy and speed exceed radiology technicians. Furthermore, we demonstrate clinical
utilization of the model in a prospective clinical setting. In the future, we expect to expand our model’s use within our institution and elsewhere to improve the speed and accuracy of prostate segmentations for targeted biopsy.

CONCLUSIONS

Despite the enormous potential of deep learning to perform image analysis tasks, clinical implementation has been minimal to date. To our knowledge, deep learning has not been used clinically for the important and time-consuming prostate segmentation task, while having the code released online. We developed a deep learning model to segment the prostate gland on T2-MRI and proved that it outperformed common deep learning networks as well as trained radiology technicians. The model saved almost 16 hours in segmentation time in a 100-patient test set alone. Most importantly, we successfully integrated it with biopsy software to allow
clinical use in a urological clinic in a proof-of-principle fashion.

ACKNOWLEDGMENTS
The authors thank Rajesh Venkataraman for help converting the segmentation files into a Digital Imaging and Communications in Medicine (DICOM) format that can be read by the ProFuse software (Eigen, Grass Valley, California). The authors also acknowledge the efforts of Rhea Liang and Chris LeCastillo of the 3D and Quantitative Imaging Laboratory at Stanford University.

REFERENCES
1. Liu W, Patil D, Howard DH et al: Adoption of prebiopsy magnetic resonance imaging for men undergoing prostate biopsy in the United States. Urology 2018; 117: 57.
2. Kasivisvanathan V, Rannikko AS, Borghi M et al: MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med 2018; 378: 1767.
3. Rouvière O, Puech P, Renard-Penna R et al: Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naïve patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. Lancet Oncol 2019; 20: 100.
4. van der Leest M, Cornel E, IsraëI B et al: Head-to-head comparison of transrectal ultrasound-guided prostate biopsy versus multiparametric prostate resonance imaging with subsequent magnetic resonance-guided biopsy in biopsy-naïve men with elevated prostate-specific antigen: a large prospective multicenter clinical study. Eur Urol 2019; 75: 570.
5. Ahdoot M, Wilbur AR, Reese SE et al: MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. N Engl J Med 2020; 382: 917.
6. Sonn GA, Margolis DJ and Marks LS: Target detection: magnetic resonance imaging-ultrasound fusion-guided prostate biopsy. Urol Oncol Semin Original Invest 2014; 32: 903.
7. Ravi D, Wong C, Deligianni F et al: Deep learning for health informatics. IEEE J Biomed Health Inform 2017; 21: 4.
8. AllBadawy EA, Saha A and Mazurowski MA: Deep learning for segmentation of brain tumors: impact of cross-institutional training and testing. Med Phys 2018; 45: 1150.
9. Zech JR, Badgeley MA, Liu M et al: Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med 2018; 15: e1002683.
10. 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 Imag 2019; 6: 1.
11. Zabiollahly F, Schiela N, Krishna Jayaraj S et al: Prostate segmenteation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. Med Phys 2019; 46: 3078.
12. 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: 447.
13. Liu Q, Dou Q, Yu L et al: MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Trans Med Imaging 2020; 39: 2713.
14. 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: 1707.
15. 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 Imag 2019; 6: 1.
16. Tian Z, Liu L, Zhang Z et al: PSNet: prostate segmentation on MRI based on a convolutional neural network. J Med Imag 2018; 5: 1.
17. Tian Z, Li X, Zheng Y et al: Graph-convolutional-network-based interactive prostate segmentation in MR images. Med Phys 2020; 47; 4164.
18. Turkbey B, Fotin SV, Huang RJ et al: Fully automated prostate segmentation on MRI: comparison with manual segmentation methods and specimen volumes. Am J Roentgenology 2013; 201: W720.
19. Lee DK, Sung DJ, Kim CS et al: Threedimensional 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: 1223.
20. 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: 1403.
21. Bezique A, Moriaty A, Farrell C et al: Determination of prostate volume. Acad Radiol 2018; 25: 1582.
22. Litjens G, Toth R, van de Ven W et al: Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Med Image Anal 2014; 18: 359.
23. Bloch N, Madabhushi A, Huisman H et al: NCI-ISBI 2013 challenge: automated segmentation of prostate structures. Cancer Imaging Archive 2015; 370.
24. NyuL LG and Udupa JK: On standardizing the MR image intensity scale. Magn Reson Med 1999; 42: 1072.
25. Reinhold JC, Dewey BE, Carass A et al: Evaluating the impact of intensity normalization on MR image synthesis. In: Medical Imaging 2019: Image Processing. Vol 10949. Bellingham, Washington: International Society for Optics and Photonics 2019; p 10949H1.
26. Wang J and Perez L: The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Networks Vis Recognit 2017; 11.
27. Ronneberger O, Fischer P and Brox T et al: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Lecture Notes in Computer Science. Edited by N Navab, J Hornegger, WM Wells et al: Cham, Switzerland: Springer International Publishing 2015; p 234.

EDITORIAL COMMENT
The application of artificial intelligence to health care promises to address many inefficiencies, from imaging diagnosis to treatment optimization and even scheduling and billing, but likely no field is better poised to exploit its advantages than radiology.