Abdominal Radiology

Published online: 5 May 2018

Abdom Radiol (2019) 44:1985–1989
https://doi.org/10.1007/s00261-018-1613-1

Are we at a crossroads or a plateau? Radiomics and machine learning in abdominal oncology imaging

Ronald M. Summers

Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences Department, National Institutes of Health Clinical Center, Bldg. 10 Room 1C224D, MSC 1182, Bethesda, MD 20892-1182, USA

Abstract

Advances in radiomics and machine learning have driven a technology boom in the automated analysis of radiology images. For the past several years, expectations have been nearly boundless for these new technologies to revolutionize radiology image analysis and interpretation. In this editorial, I compare the expectations with the realities with particular attention to applications in abdominal oncology imaging. I explore whether these technologies will leave us at a crossroads to an exciting future or to a sustained plateau and disillusionment.

Key words: Machine learning—Radiomics—Computer-aided detection—Image segmentation

Recent advances in computer science have led to dramatic improvements in automated radiology image analyses [1, 2]. Whereas in the past it could take several years to design, build, and perfect a computer system to do computer-aided diagnosis for a particular type of radiology image, using the latest software tools, it is now possible to develop such systems in a matter of months. In addition, the level of expertise required to design such a system has dropped dramatically. Interested researchers can download machine learning authoring software that makes the process efficient and relatively painless. The underlying machine learning technologies continue to improve and the latest advances are generally widely and publicly available.

How has all of this affected radiomics and machine learning of abdominal oncology imaging? First, it means that many new diseases and therapies can be studied using machine learning and radiomics [3, 4]. Second, it means that the priorities have shifted to a demand for well-annotated images rather than a need for sophisticated mathematics and software. Third, it means that existing labeled data sets can be repurposed to develop systems to perform automated diagnosis for various radiology image analysis tasks.

Fourth, it means that holistic techniques that can analyze the whole scan for a multitude of abnormalities are becoming available [5]. This last point is an interesting one. In the past, combining CAD systems was unattractive because high aggregate false positive rates from the systems would lead to inefficiency and a perceived lack of clinical relevance. The latest systems permit an arbitrary number of diagnostic predictions to be made for each image.

However, with all of the remarkable advances that have occurred in just the past few years, there are worrisome signs that progress is reaching a plateau. Many radiology researchers are using the same radiomics and machine learning technologies to address different clinical problems [6]. Radiomics texture analyses have become commonplace and typically use the same set of texture features. Such analyses frequently report positive results but reproducibility by independent groups is lacking [7]. The improvements in performance as measured by sensitivity and specificity have stalled for many applications. False positive rates for automated detection are generally still too high, exceeding one false positive per patient exam. Segmentation accuracy using machine learning struggles to reach Dice similarity coefficients in the mid 80% range for more challenging organs and lesions [8, 9].

Current data indicate that in general even the best machine learning systems do not yet perform at the level of a radiologist. However, if clinical image analysis

This work is dedicated to the memory of my former colleague, Andrew Dwyer, MD.

Correspondence to: Ronald M. Summers; *email:* rms@nih.gov

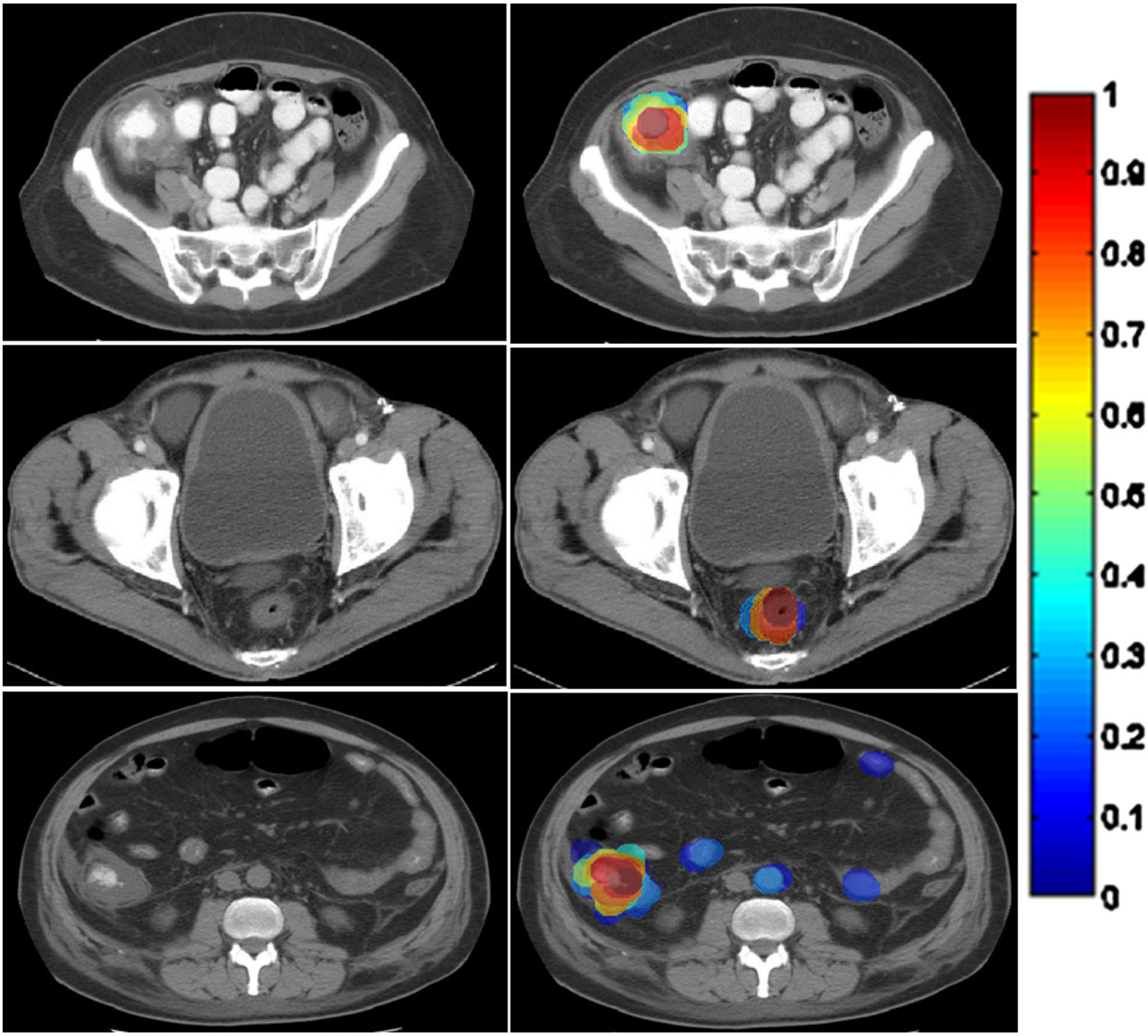

Fig. 1. Example of colitis detection on abdominopelvic CT. The color code indicates the colitis score, with higher scores indicating a greater likelihood of the presence of colitis. Reproduced from Ref. [18].

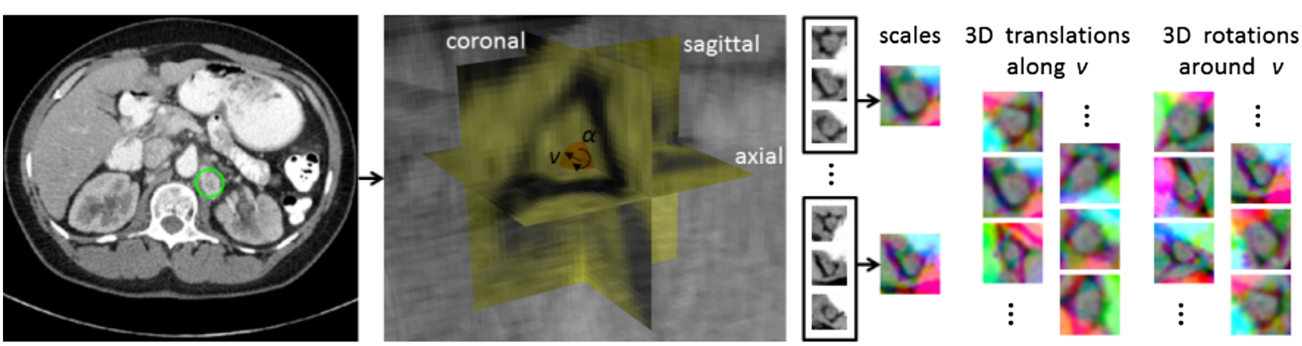

Fig. 2. Example of an automated deep learning-based lymphadenopathy detector for abdominopelvic CT. The example shows how three orthogonal planes are taken through a left paraaortic lymph node candidate. The three planes are scaled and stretched to augment the dataset for training the deep learning system. Reproduced from [28].

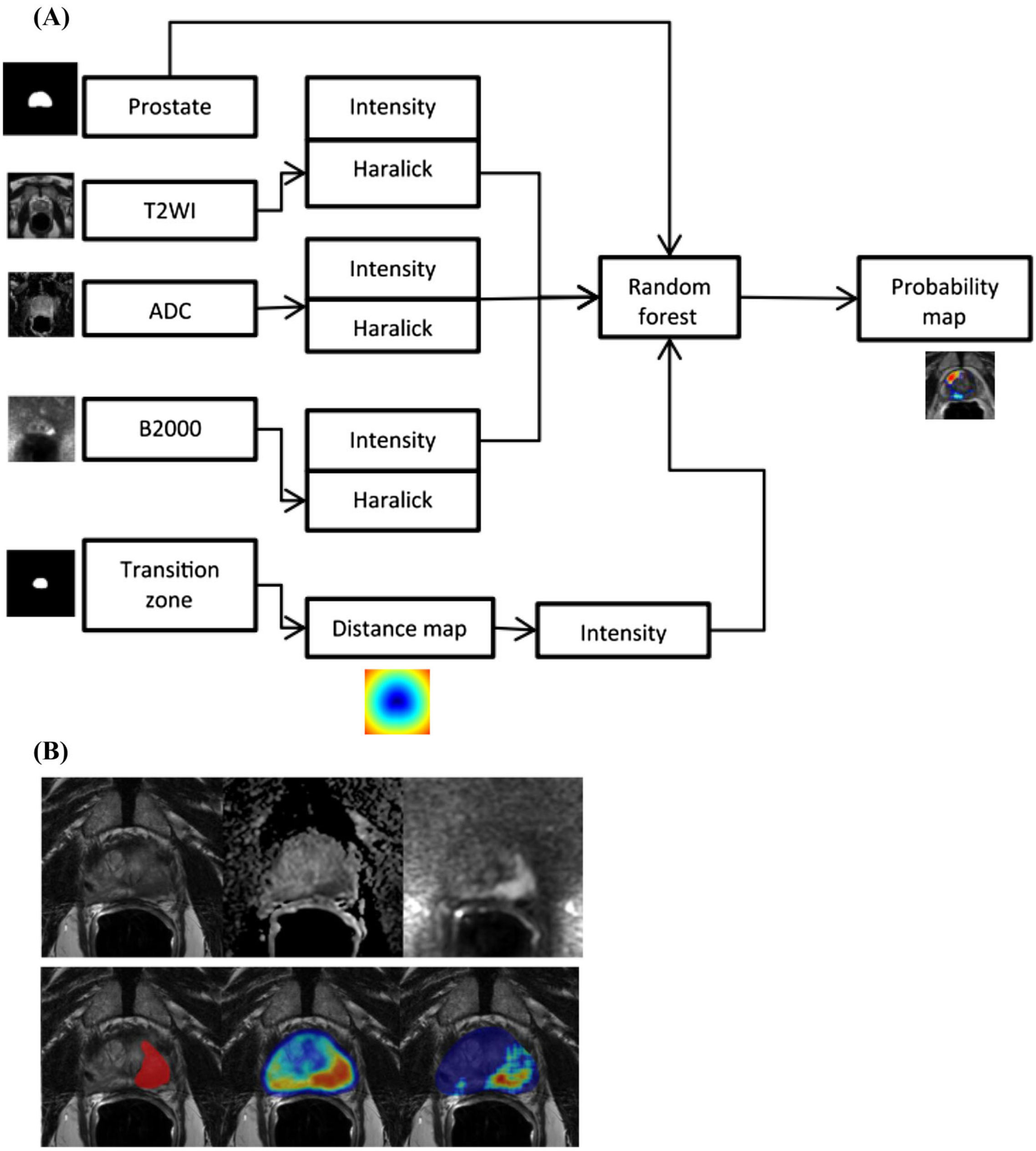

Fig. 3. Example machine learning-based system to detect cancer on multiparametric prostate MRI. **A** Flowchart of method showing how T2WI, ADC map and high b-value image are analyzed to yield malignancy probability map. **B** 74-year-old patient with a cancerous lesion in the left apical midperipheral zone (Gleason 9). T2WI, ADC map and high b-value image, annotation, probability map, feature-based CAD output. The probability map from the random forest machine learning-based detector identifies the cancer better than does the feature-based CAD output. Reproduced from [36].

problems are sufficiently narrowly defined, it is possible to achieve physician-level performance for some tasks such as skin lesion classification and assessment of retinal fundus photographs for diabetic retinopathy [10, 11]. Similar narrowly focused radiology tools have reached radiologist level performance, such as tuberculosis detection on chest radiographs and bone age estimation on pediatric hand radiographs [12, 13]. There are still few examples of physician-level performance in radiomics and machine learning for abdominal oncology imaging.

Can we expect machine learning performance to continue to improve? While the level of hype about deep learning has led various commenters to claim that general artificial intelligence is on the horizon, the data now support a relatively slow or gentle slope of improvement in radiology automated image analyses. With particular attention to abdominal oncology, there are many challenges that still need to be addressed.

The use of radiomics and machine learning to predict patient outcomes is still in its infancy. A number of works have investigated specific tumors and therapies. In some cases, radiomics and machine learning analyses can predict which patients have greater or lower likelihood of long-term survival [14, 15]. There are many tumors and therapies and combinations thereof, and only a tiny handful of these have been investigated. It would be ideal if such data were made widely available upon collection so that investigators who have already devised survival prediction software could apply their technology to the latest clinical data and therefore accelerate clinical decision-making.

The use of radiomics and machine learning for disease detection is at a somewhat more advanced stage. For abdominal oncology, many tools have been developed for identifying lesions in the liver, kidney, bladder, and bone [4]. Aside from CT colonography for colorectal cancer screening, little work has been done on automated bowel oncology lesion detection. Early work has been done on localization of small and large bowel on CT [16, 17]. A recent paper has shown how colitis may be accurately detected, a common complication found in patients undergoing certain oncology therapies [18] (Fig. 1).

Lesion segmentation is another area amenable to radiomics and machine learning. Automated segmentation has been applied widely to abdominal organ segmentation as a prerequisite to lesion detection and segmentation. Liver lesion segmentation is one of the most well studied of the software tools [19–22]. Some work has been done in bone lesion segmentation [23–25]. Early work has been done on detection of subcutaneous metastases in melanoma [26]. Bladder lesion segmentation has been conducted as part of a project to evaluate automated detection and therapeutic response [27]. Deep learning has improved abdominopelvic lymphadenopathy detection and segmentation [28–30] (Fig. 2). Public datasets are available for researchers interested in developing lymph node detectors [31]. Little work has been done at automatically detecting peritoneal, muscular, gynecologic, and prostatic lesions on CT [32]. Prostate MRI lesion detection and segmentation have been relatively better studied, with lesion segmentation being in a more nascent stage of development [33–36] (Fig. 3).

How can we move forward? At this stage, multi-institutional data collections with labels provided by multiple trained observers are highly desirable. There are still challenges with getting institutions to make clinical images available. There are also considerable challenges getting sufficient expert annotations although some early progress is being made [37]. Funding bodies have been somewhat reluctant to fund projects primarily focused on creating a data corpus. However given the relative ease of performing the analyses once data are in hand, focus may need to shift to creating such collections. Approaches to improve the efficiency and lower the cost of annotation (e.g., crowdsourcing) would be highly desirable research directions.

Most large hospitals have PACS systems loaded with images from oncology patients. Images and accompanying reports can be collected and aggregated with those from other institutions. Text mining of the reports can be combined with radiomics and machine learning of radiology images to train software how to detect abnormalities in test images. Lesion measurements, particularly those done for RECIST purposes, can be mined from the PACS and used to train systems to perform lesion detection [5]. Such data are already collected and can be easily mined to produce disease-detecting systems. Whether such easily accumulated data can be used to produce high-performing systems is still an open research problem.

Whether progress in the application of radiomics and machine learning to abdominal oncology imaging is at a crossroads or a plateau will depend to a great extent on how researchers surmount these challenges. If community backing can be obtained for large-scale multi-institutional data collection and labeling, then we are at a crossroads leading to the potential for amazing progress. However, if such collections cannot be obtained, then we may remain stuck at a broad plateau for the foreseeable future. It is in the best interests of both our patients and our profession that we work together to ensure that progress continues unabated.

Compliance with ethical standards

Funding This study was funded by the Intramural Research Program of the National Institutes of Health, Clinical Center (Grant Number 1Z01 CL040004).

Conflicts of interest The author has pending and/or awarded patents for automated image analyses, and received royalty income from iCAD, Zebra Medical, Imbio, ScanMed and Koninklijke Philips. His lab received research support from Ping An Technology Company Ltd. and NVIDIA.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent N/A.

References

1. Greenspan H, van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 35(5):1153–1159
2. Litjens G, Kooi T, Bejnordi BE, et al. (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88

3. Kontos D, Summers RM, Giger M (2017) Special section guest editorial: radiomics and deep learning. J Med Imaging (Bellingham) 4(4):041301
4. Summers RM (2016) Progress in fully automated abdominal CT interpretation. AJR Am J Roentgenol 207(1):67–79
5. Yan K, Wang X, Lu L, et al (2018) Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. http://arxiv.org/abs/1711.10535. (Accepted by IEEE CVPR, 2018)
6. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted interventional (MICCAI), Part III, LNCS 9351, pp. 234–241.
7. Summers RM (2016) Texture analysis in radiology: does the emperor have no clothes? Abdom Radiol (NY) 42(2): 342–345
8. Roth HR, Lu L, Farag A, et al. (2015) DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: MICCAI 2015, Part I, LNCS 9349. Springer, Cham, pp. 556–564
9. Zhou Y, Xie L, Shen W, et al. (2017) *A fixed-point model for pancreas segmentation in abdominal CT scans*. Cham: Springer, pp 693–701
10. Gulshan V, Peng L, Coram M, et al. (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410
11. Esteva A, Kuprel B, Novoa RA, et al. (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
12. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 284(2):574–582
13. Larson DB, Chen MC, Lungren MP, et al. (2018) Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology 287(1):313–322
14. Smith AD, Gray MR, del Campo SM, et al. (2015) Predicting overall survival in patients with metastatic melanoma on antiangiogenic therapy and RECIST stable disease on initial posttherapy images using CT texture analysis. AJR Am J Roentgenol 205(3):W283–W293
15. Lovinfosse P, Polus M, Van Daele D, et al. (2018) FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. Eur J Nucl Med Mol Imaging 45(3):365–375
16. Zhang W, Liu J, Yao J, et al. (2013) Mesenteric vasculature-guided small bowel segmentation on 3-D CT. IEEE Trans Med Imaging 32(11):2006–2021
17. Cherry KM, Peplinski B, Kim L, et al. (2015) Sequential Monte Carlo tracking of the marginal artery by multiple cue fusion and random forest regression. Med Image Anal 19(1):164–175
18. Liu JM, Wang D, Lu L, et al. (2017) Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks. Med Phys 44(9):4630–4642
19. Ben-Cohen A, Klang E, Diamant I, et al. (2015) Automated method for detection and segmentation of liver metastatic lesions in follow-up CT examinations. J Med Imaging (Bellingham) 2(3):034502
20. Linguraru MG, Richbourg WJ, Liu J, et al. (2012) Tumor burden analysis on computed tomography by automated liver and tumor segmentation. IEEE Trans Med Imaging 31(10):1965–1976
21. Hoogi A, Beaulieu CF, Cunha GM, et al. (2017) Adaptive local window for level set segmentation of CT and MRI liver lesions. Med Image Anal 37:46–55
22. Hame Y, Pollari M (2012) Semi-automatic liver tumor segmentation with hidden Markov measure field model and non-parametric distribution estimation. Med Image Anal 16(1):140–149
23. O'Connor SD, Yao J, Summers RM (2007) Lytic metastases in thoracolumbar spine: computer-aided detection at CT–preliminary study. Radiology 242(3):811–816
24. Burns JE, Yao J, Wiese TS, et al. (2013) Automated detection of sclerotic metastases in the thoracolumbar spine at CT. Radiology 268(1):69–78
25. Yao J, Burns JE, Sanoria V, Summers RM (2017) Mixed spine metastasis detection through positron emission tomography/computed tomography synthesis and multiclassifier. J Med Imaging (Bellingham) 4(2):024504
26. Solomon J, Mavinkurve S, Cox D, Summers RM (2004) Computer-assisted detection of subcutaneous melanomas: feasibility assessment. Acad Radiol 11(6):678–685
27. Cha K, Hadjiiski L, Chan HP, et al. (2015) Detection of urinary bladder mass in CT urography with SPAN. Med Phys 42(7):4271–4284
28. Roth HR, Lu L, Seff A, et al. (2014) A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. Med Image Comput Comput Assist Interv 17(Pt 1):520–527
29. Seff A, Lu L, Barbu A, Roth H, Shin H-C, Summers RM (2015) Leveraging mid-level semantic boundary cues for automated lymph node detection. In: Medical image computing and computer-assisted intervention (MICCAI). Springer, Berlin, pp. 53–61
30. Nogues I, Lu L, Wang X, et al. (2016) Automatic lymph node cluster segmentation using holistically-nested neural networks and structured optimization in CT images. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W (eds) MICCAI, Part II, LNCS vol 9901, pp 388–397
31. Roth HR, Lu L, Seff A, et al. (2015) CT lymph nodes. https://doi.org/10.7937/K9/TCIA.2015.AQIIDCNM
32. Liu JF, Wang SJ, Linguraru MG, Yao JH, Summers RM (2014) Tumor sensitive matching flow: a variational method to detecting and segmenting perihepatic and perisplenic ovarian cancer metastases on contrast-enhanced abdominal CT. Med Image Anal 18(5):725–739
33. Clark T, Zhang J, Baig S, et al. (2017) Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. J Med Imaging (Bellingham) 4(4):041307
34. Tian Z, Liu L, Zhang Z, Fei B (2018) PSNet: prostate segmentation on MRI based on a convolutional neural network. J Med Imaging (Bellingham) 5(2):021208
35. Cheng R, Roth HR, Lay N, et al. (2017) Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks. J Med Imaging (Bellingham) 4(4):041302
36. Lay N, Tsehay Y, Greer MD, et al. (2017) Detection of prostate cancer in multiparametric MRI using random forest with instance weighting. J Med Imaging 4(2):024506
37. Lugo-Fagundo C, Vogelstein B, Yuille A, Fishman EK (2018) Deep learning in radiology: now the real work begins. J Am Coll Radiol 15(2):364–367