

ARTIFICIAL INTELLIGENCE IN IMAGING



BIR World Partner Network



中华医学会放射学分会
Chinese Society of Radiology

INTRODUCTION

Artificial Intelligence in Imaging is a joint publication published in collaboration between the British Institute of Radiology (BIR) and the Chinese Society of Radiology (CSR).

Here you will find scholarly articles from the flagship journals of both societies (*BJR* and *CJAR*) on artificial intelligence and imaging. Topics include deep learning, automation in intensity modulated radiotherapy treatment planning, machine learning, automated tools for screening and the use of radiomics in predicting behaviour in various cancers.

The CSR and BIR are members of the BIR World Partner Network (WPN), a community of 19 international societies brought together by the BIR, with the shared goal of advancing knowledge and understanding of all their members to improve the scope and value of imaging and radiation oncology for patients across the world. This publication is an initiative arising from WPN meetings and discussions.

ABOUT THE BIR AND BJR

The BIR is an international membership organisation for everyone working in imaging, radiation oncology and the underlying sciences.

Our aims are to:

- Support the work of our members and their colleagues to achieve professional excellence
- Provide continuing professional development for our multidisciplinary community
- Publish cutting-edge research for our authors and readers across the world
- Influence and connect with the wider professional sector.

BJR is the international research journal of the British Institute of Radiology and the oldest radiological journal in the world.

ABOUT THE CSR AND CJAR

The largest radiology society in China, CSR is a non-profit national community dedicated to radiology research and clinical treatment, founded in 1937.

As a subordinate of the Chinese Medical Association (CMA), CSR's mission is to unite all radiologists and radiologic technologists nationwide, and actively promote the development of Chinese radiology.

Chinese Journal of Academic Radiology, (CJAR) the first English-language journal in the radiology field in China, provides a communication forum for all international scholars and radiologists in the area of clinical practice and translational research on imaging techniques including radiology, image-fused nuclear medicine and ultrasound.

CONTENTS LIST

Current applications and challenges of radiomics in urothelial cancer

Chinese Journal of Academic Radiology

Gumuyang Zhang, Lili Xu, Hao Sun, Zhengyu Jin

Will machine learning end the viability of radiology as a thriving medical specialty?

British Journal of Radiology

Stephen Chan, Eliot L Siegel

Radiomics in prostate cancer: basic concepts and current state-of-the-art

Chinese Journal of Academic Radiology

Shan Yao, Hanyu Jiang, Bin Song

Automation in intensity modulated radiotherapy treatment planning—a review of recent innovations

British Journal of Radiology

Mohammad Hussein, Ben J M Heijmen, Dirk Verellen, Andrew Nisbet

Detectability of pulmonary nodules by deep learning: results from a phantom study

Chinese Journal of Academic Radiology

Qiong Li, Qing-chu Li, Rui-ting Cao, Xiang Wang, Ru-tan Chen, Kai Liu, Li Fan, Yi Xiao, Zi-tian Zhang, Chi-Cheng Fu, Qiong Song, Weiping Liu, Qu Fang, Shi-yuan Liu

Population-based opportunistic osteoporosis screening: Validation of a fully automated CT tool for assessing longitudinal BMD changes

British Journal of Radiology

Perry J. Pickhardt, Scott J. Lee, Jiamin Liu, Jianhua Yao, Nathan Lay, Peter M Graffy, Ronald M Summers

Radiomic features based on MRI for prediction of lymphovascular invasion in rectal cancer

Chinese Journal of Academic Radiology

Yu Fu, Xiangchun Liu, Qi Yang, Jianqing Sun, Yunming Xie, Yiyang Zhang, Huimao Zhang

The challenge of clinical adoption—the insurmountable obstacle that will stop machine learning?

BJR | Open

Jonathan Taylor, John Fenner



Current applications and challenges of radiomics in urothelial cancer

Gumuyang Zhang¹ · Lili Xu¹ · Hao Sun¹ · Zhengyu Jin¹

Received: 24 June 2019 / Revised: 14 October 2019 / Accepted: 6 November 2019
© Springer Nature Singapore Pte Ltd. 2019

Abstract

New discoveries and technologies have begun to change paradigms of urothelial cancer therapy in recent years. One of the novel techniques which emerged in the imaging community is radiomics, which refers to the high-throughput extraction of quantitative image features from medical images. Radiomics, being noninvasive and easy to perform, has shown great potential in oncology by providing valuable information about tumor type, aggressiveness, progression, response to treatment and prognosis and enabling us to gain insights into the true utility of personalized medicine in the management of cancer in the near future. With rapid development in this area, radiomics has already been applied in urothelial cancer to predict pathological grade, clinical stage, lymph node metastasis and treatment response demonstrating promising results. In this review, we highlight advances in clinical applications of radiomics in urothelial cancer, discuss about the challenges and implications of radiomics for radiologists and suggest the future directions that we could move toward in order to fully realize the potentials of radiomics to improve personalized management of patients with urothelial cancer.

Keywords Urothelial cancer · Radiomics · Imaging biomarker · Precision medicine

Introduction

Urothelial cancer is a common malignancy worldwide. More than 90% of the cases are bladder cancer while upper tract urothelial cancer (UTUC) is a rare subset [1]. Bladder cancer is the sixth most common cancer in the USA with an estimation of 81,000 new cases and 17,000 deaths each year, and the standardized mortality rate varies from 2 to 10/100,000 per year in men and 0.5 to 4/100,000 per year in women [2, 3]. According to the National Central Cancer Registry of China, the new cases and deaths for bladder cancer are about 80,500 cases and 32,900 cases, respectively, with an upward trend in recent years [4]. Over the past three decades, few signs of progress have been made in the treatment for urothelial cancer. For patients with nonmuscle-invasive bladder cancer, around one-third of patients experience recurrences

or progression despite receiving standard treatment. The 5-year survival of patients with muscle-invasive bladder cancer is less than 50%, and the median overall survival of patients with advanced or metastatic bladder cancer is about 15 months [5, 6]. The deadlock of urothelial cancer treatment has been broken with significant advances in our understanding of underlying tumor biology and immunology in recent years. The advent of immune checkpoint inhibitors (ICIs) including anti-programmed cell death 1 (PD-1) and anti-programmed cell-death ligand 1 (PD-L1) antibodies has revolutionized the treatment for many advanced solid tumors including urothelial cancer [7]. In spite of dramatic improvements in clinical outcomes in certain patients, there are still quite a few important unmet clinical needs in urothelial cancer management, for example, how to identify patients most likely to benefit from ICIs and how to predict treatment response for individuals [8]. It is necessary to develop new tools that have the potential to tackle those difficulties in tailoring treatment for each patient with urothelial cancer, especially in the era of precision medicine.

A wide range of “-omic” technologies, such as genomics and proteomics, have been investigated in the field of oncology to improve current biomarkers for the diagnosis and therapy of tumors including urothelial cancer. The term “radiomics” has been introduced several years ago,

✉ Hao Sun
sunhao_robert@126.com

✉ Zhengyu Jin
jinzy@pumch.cn

¹ Department of Radiology, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, No. 1 Shuaifuyuan, Wangfujing Street, Dongcheng District, Beijing 100730, People's Republic of China

and it has become a novel research field with rapid development. Radiomics refers to analysis and translation of medical images into mineable and measurable high-dimensional data producing quantitative features in relation to prediction targets such as gene expression and clinical outcomes [9, 10]. Radiomics generally involves five major steps: data selection, medical imaging, feature extraction, exploratory analysis and modeling. Figure 1 generally describes the process. The whole process mainly relies on computer algorithms rather than human visual assessment, and this advantage of quantitative analysis allows radiomics to reveal information related to cellular and molecular properties of the tissue that may not be perceived by human naked eyes [11]. As tumors are extremely heterogeneous, a biopsy of a limited tissue sample is unlikely to represent the entire tumor. Unlike biopsy, radiomics can

examine a tumor as a whole and evaluate intratumoral heterogeneity at the same time [12]. These unique abilities of radiomics allow it to be applied as a promising biomarker to noninvasively indicate biological processes, pathological changes or responses to therapeutic intervention.

Up to now, many published papers have demonstrated the huge potential of radiomics to improve clinical diagnosis, disease monitoring and outcome predictions in various solid tumors including urothelial cancer. In this review, we briefly summarize the current clinical applications of radiomics in urothelial cancer along with the latest development in the field. Then, we discuss the challenges we met in this field and the implications for radiologists. At last, we offer our perspectives of future research directions of radiomics in urothelial cancer.

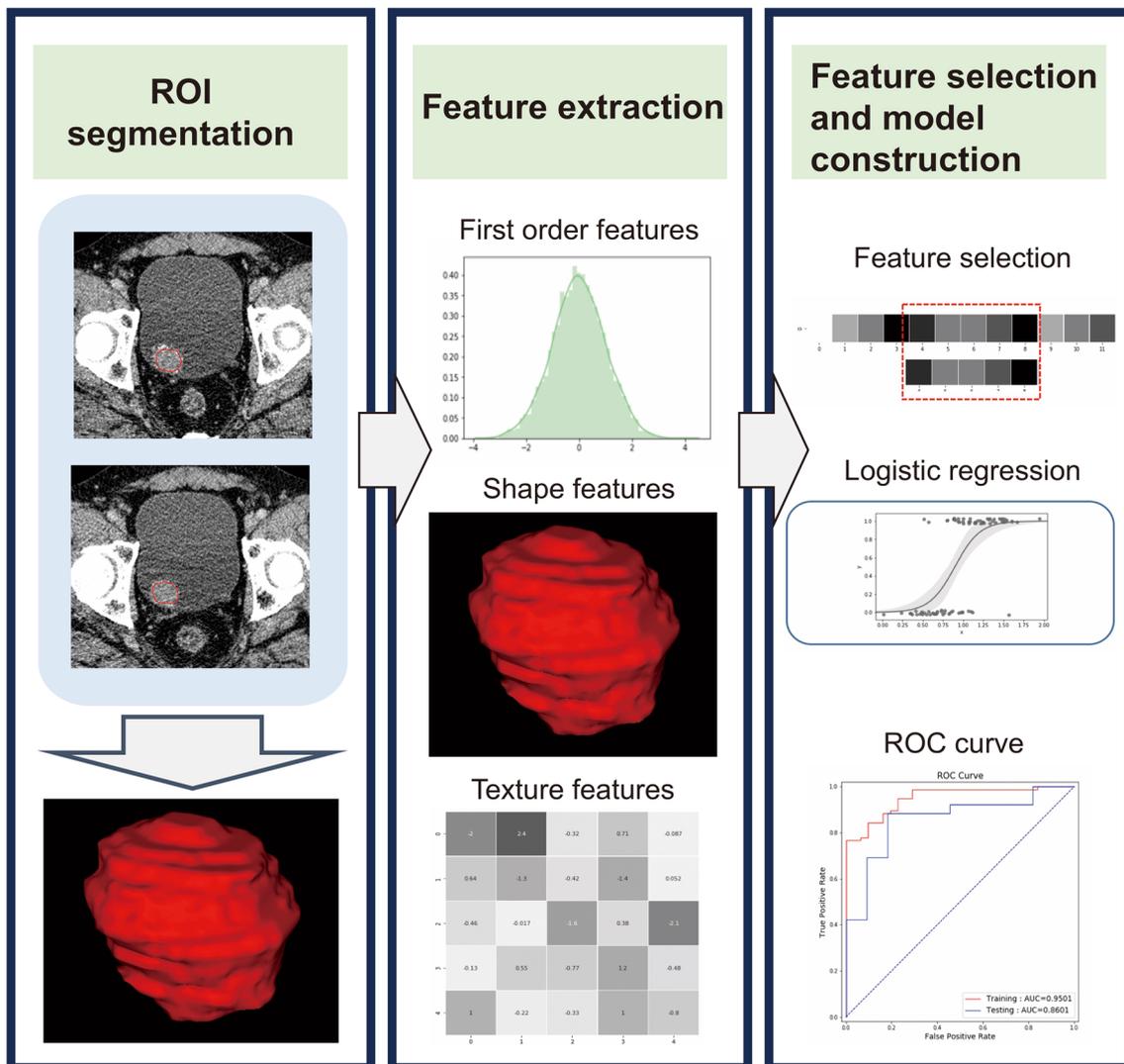


Fig. 1 Radiomics in urothelial cancer. Representative images of a bladder cancer, with region of interest (ROI) segmentation shown in red followed by feature extraction, feature selection and model construction

Radiomics applications in urothelial cancer

The past few years have witnessed considerable scientific advances in applications of artificial intelligence in human malignant neoplasms. Different technical methods including texture analysis, machine learning, deep learning and radiomics have been employed in studies regarding urothelial cancer. We would like to focus on radiomics studies in this review but we will also mention related studies using other technical methods as well, so as to give a comprehensive review of achievements in this field in the last several years. In the following paragraphs, we are going to group these studies by the targeted clinical question and analyze in detail the data related to each topic.

Evaluation of pathological grade

Pathological grade of urothelial cancer has important implications for prognosis and treatment selection. Low-grade urothelial cancer has a lower rate of recurrence and stage progression, and it could be treated with less invasive techniques [13]. A few studies have explored the feasibility of texture analysis, machine learning and radiomics to distinguish between low- and high-grade urothelial cancer. A study by our group in 105 patients with urothelial carcinoma found that low-grade tumors demonstrated lower texture features of mean, entropy and mean of positive pixels (MPP) quantified from CT images and MPP could differentiate low- from high-grade tumors with an area under the curve (AUC) of 0.779 [14]. Mammen et al. performed a similar texture analysis of CT scans of 48 patients with UTUC and found entropy was also greater in high-grade tumors with an AUC of 0.83 [15]. In a study by Zhang et al. in 61 patients with bladder cancer, 102 texture features were extracted from diffusion-weighted images (DWI) and apparent diffusion coefficient (ADC) maps, 47 of which were found to be significantly different between low- and high-grade tumors [16]. By the method of support vector machine with recursive feature elimination (SVM-RFE), 22 features were selected to build the classifier and reached an AUC of 0.861. Wang et al. conducted a similar but more comprehensive radiomics analysis of MRI images for preoperative evaluation of pathological grade [17]. They examined T2-weighted (T2W), DWI and ADC maps of 70 patients and validated them in a cohort of 30 patients with bladder cancer. Multimodal features were combined to construct radiomics models which achieved an AUC around 0.92 in both training and validation cohorts. The above studies revealed that texture features extracted from CT or MRI images could reflect the difference between low- and high-grade urothelial

carcinoma. A radiomics approach has the potential to act as a noninvasive tool to assist in preoperative grading of urothelial cancer. But the sample size of these studies is small, and as UTUC is quite rare compared to bladder cancer, only our group and Mammen et al. have included UTUC [14, 15]. More data should be gathered to further evaluate the ability of radiomics for predicting pathological grade, especially in UTUC.

Evaluation of clinical T stage

Accurate local staging of bladder cancer is key to realize optimal management of an individual patient. Patients with nonmuscle-invasive bladder cancer (NMIBC, stage $\leq T1$) are mostly treated with bladder-sparing methods such as transurethral resection of a bladder tumor (TURBT) and intravesical therapies, whereas patients with muscle-invasive bladder cancer (MIBC, stage $\geq T2$) are treated with cystectomy, radiation therapy or chemotherapy and usually have a poor prognosis [13, 18]. Thus, discrimination between NMIBC and MIBC has great value in guiding therapeutic choices. In a study by Garapati et al., morphological and texture features were extracted from CT images of 84 bladder cancer lesions and a linear discriminant analysis, a neural network, a SVM and a random forest classifier were used to combine the features to stratify the stage of bladder cancer into two groups: $\geq T2$ and $< T2$ groups [19]. The classification accuracies of the four classifiers were similar with AUCs around 0.9. Several MRI-based radiomics studies have also reported promising results of accurate stratification for stages of bladder cancer. Xu et al. extracted a total of 1104 radiomic features from T2W and DW images of 44 patients with bladder cancer and selected 19 features to build an optimal discriminative model by the method of SVM-RFE and synthetic minority oversampling technique (SMOTE) to differentiate between NMIBC and MIBC or stage $\geq T2$ and $< T2$ [20]. An AUC of 0.9857 was reached by the SVM-RFE + SMOTE classifier, which outperformed the diagnostic accuracy by experts. In a study by Tong et al., T2W images of 65 patients with bladder cancer were used to quantify intensity and texture parameters to classify patients into $\geq T2$ or $< T2$ [21]. Nine optimal features were selected from a total of 15,834 features and demonstrated an AUC of 0.813 for the differentiation purpose. Lim et al. performed a slightly different study from the previous ones [22]. They evaluated whether T2W and ADC texture features of bladder cancer and extravesical fat as well in 36 patients could be used to predict MIBC ($\geq T2$) and extravesical disease ($\geq T3$) after TURBT. Results show that greater entropy of bladder cancers and extravesical fat was found in category $\geq T3$ than in category $\leq T2$ and in category $\geq T2$ than in category T1 tumors with AUCs in

the range of 0.74–0.85. These studies indicated that radiomics could help with local staging of bladder cancer and has the potential to improve current patient management. The main limitation of this study is the very small sample size, and none of the above studies has a patient population over 100. It is no doubt that these preliminary studies proved the concept of applying radiomics to evaluate the local stage of bladder cancer, but where the study results are still valid or radiomics could really promote precise local staging of bladder cancer needs further investigation and validation in larger cohorts.

Prediction of lymph node (LN) metastasis

LN metastasis in patients with bladder cancer indicates a poorer prognosis; thus, the accurate prediction of LN metastasis in patients with bladder cancer assists in treatment decision making. Routine CT and MRI identify positive LN metastasis according to the size of LN, but the efficacy is quite low with a sensitivity of 31–45%, which indicates a certain proportion of patients being understaged [23–25]. A study by Wu et al. in 118 patients with bladder cancer found 150 radiomic features in each patient's arterial-phase CT images, among which nine LN status-related features were used to build the radiomic signature for LN metastasis and achieved favorable prediction efficacy [26]. The radiomics nomogram incorporating the radiomics signature and CT-reported LN status also demonstrated good calibration and discrimination in the training set (AUC 0.9262) and the validation set (AUC 0.8986). The same research group conducted another study in 103 patients with bladder cancer in the purpose of developing and validating an MRI-based radiomics signature for the individual preoperative prediction of LN metastasis [25]. A total of 718 radiomic features were extracted from T2W images, and nine features were selected to construct the radiomic signature which showed a favorable outcome in the training set with an AUC of 0.9005 and in the validation set with an AUC of 0.8447. The radiomics signature and the MRI-reported LN status constituted the nomogram, and it demonstrated good calibration and discrimination in the training (AUC 0.9118) and validation (AUC 0.8902) sets. These two studies proved the promising value of radiomics in the prediction of LN metastasis in bladder cancer. Despite the relatively satisfactory performance of CT- and MRI-based radiomics, a shared limitation of the two studies is that they lack external validation. Multicenter validation with larger cohorts is required to confirm the ability of radiomics to accurately predict LN metastasis. Genetic markers have been shown to be predictive of LN metastasis, the addition of genetic markers to the nomogram might further improve the accuracy of radiomics to predict LN metastasis and further studies may work on this issue.

Prediction of recurrence risk

A prominent characteristic of bladder cancer is its high recurrent rate, which could reach up to 61% for patients with nonmuscle-invasive bladder cancer in the first 2 years (TFTY) after TURBT [27]. Preoperative prediction of recurrence risk is critical for prognostication and individualized follow-up regimens for patients. Xu et al. developed and validated a nomogram combining MRI-based radiomics and clinical predictors for predicting the TFTY recurrence risk [28]. Of the 1872 features extracted from T2W, DW, ADC and dynamic contrast-enhanced images, the 32 features with the highest AUC were selected for calculating Rad-Score. The nomogram developed by Rad-Score and clinical predictor of muscle-invasive status produced a good performance in the training (accuracy 88%, AUC 0.915) and validation cohorts (accuracy 80.95%, AUC 0.838). This preliminary study demonstrates the ability of radiomics together with clinical factors to address the important clinical issue of recurrence risk prediction for bladder cancer. So far, we have not found any study investigating the potential of CT-based radiomics for prediction of recurrence risk for bladder cancer. As CT plays an important role in preoperative evaluation and postoperative follow-up in patients with bladder cancer, it is worth exploring the value of CT-based radiomics in this clinical issue as well.

Treatment response assessment

Neoadjuvant chemotherapy before cystectomy has been shown to improve survival but only 30% of the patients have complete treatment response; a reliable prediction of the efficacy of neoadjuvant chemotherapy is beneficial for patients with bladder cancer [29]. In a study by Cha et al., they explored the feasibility of three CT-based radiomics models employing different design principles to distinguish between patients with and without complete chemotherapy responses [30]. The three models included a model using deep-learning convolution neural network (DL-CNN), a model using radiomic features extracted from segmented lesions and a model using radiomic features extracted from pre- and post-treatment paired regions of interest. All the three models produced comparable AUCs compared to two expert radiologists ranging from 0.69 to 0.77. It is obvious that the accuracy is not satisfactory in terms of AUCs but this study is the first to indicate the potential of using DL-CNN and radiomics methods to assess treatment response of chemotherapy for patients with bladder cancer. The small sample size of this study (82 patients in the training set and 42 patients in the test set) could be a major factor that impacted the performance of prediction models. More radiomics studies with larger cohorts targeting the prediction

of treatment responses for urothelial cancer should be conducted in the future.

Challenges and implications for radiologists

The above studies provide improved insight into the utility of radiomics in the management of urothelial cancer. These studies demonstrate the capability of radiomics to assist more precise characterization and stratification of patients with urothelial cancer. As it is impossible to biopsy each and every lesion, radiomics offers a noninvasive and economic approach to reveal the tumor heterogeneity in different individuals, different lesions and even within the same lesion. By using radiomics as biomarkers, we may begin to appreciate the complexity of tumor biology and tailor treatment for each patient with urothelial cancer.

There is no doubt that radiomics could facilitate the process of clinical decision making, but up to now, radiomics for urothelial cancer remains in research and not in clinical use. There are quite a lot of challenges ahead of us for applying radiomics in daily practice to improve patient care. The workflow of radiomics includes data selection, medical imaging, feature extraction, exploratory analysis and modeling and implementation of radiomics is rather a complicated process. One of the major challenges lies in the optimal collection and integration of multiple data sources that can produce accurate and robust predictions. Currently, the field of radiomics lacks standardized evaluation criteria and reporting guidelines; the clinical utility of those published prediction models still needs to be further evaluated for their performance [31]. To promote the development and acceptance of radiomics, Lambin et al. have proposed the radiomics quality score (RQS) to evaluate the quality of radiomic studies [11]. The RQS evaluates each necessary step in a radiomic analysis, both rewards and penalizes the methodology and analyses of a study. Investigators should be encouraged to follow the rigorous evaluation criteria and reporting guidelines to avoid overly optimistic claims about robustness and generalizability.

High reproducibility and replicability are essential for the widespread acceptance of radiomics-based models or decision support systems in clinical practice. As radiomics studies comprise multiple steps, and each could be affected by a wide range of factors, details of these subprocesses should be disclosed by researchers; otherwise, reproducibility and replicability in radiomics would not be possible. Large-scale data sharing is imperative for the validation and generalization of radiomics; thus, disclosure of imaging protocols, analyzed scans, segmentations, details of feature extraction and modeling methodology should be provided as supplementary material in future publications.

Future directions

Medical imaging is evolving from being a diagnostic tool to becoming a vital part in the era of personalized medicine. It is of great importance to promoting precision medicine, especially in countries like China which has a large population but limited government investment in health care and low average expense per patient need. Application and generalization of novel techniques could help to provide the optimal treatment for patients while avoiding unnecessary cost, relieving the heavy economic burden of diseases for both individuals and society. Radiomics, with the advantages of being noninvasive and economical, is worthy of further investigation and application.

With the advances in radiomics, it has made it possible to correlate clinically feasible quantitative imaging with tissue pathophysiology. Radiogenomics highlighting the link between radiomic features and gene expression patterns allows the acceleration of their incorporation into personalized medicine approaches [32]. Over the years, there have been many studies investigating the application of gene expression signatures for prediction of tumor characteristics and outcomes of urothelial cancer, including stage, risk of recurrences, the progression of nonmuscle-invasive and muscle-invasive bladder cancer and survival [9, 31, 33–37]. But up to now, studies focusing on identifying the association between specific imaging traits and gene profile of urothelial cancer have not been reported yet. The research in radiogenomics of urothelial cancer is still at the initial stage and remains to be further explored. Standardized gene assay and radiomics workflow would enable radiogenomics biomarkers to meaningfully improve diagnosis, prognosis and prediction of response to treatment of urothelial cancer.

The past few years have been an exciting time for the field of urothelial cancer. With the introduction of ICIs such as atezolizumab and pembrolizumab, significant advances have been made in the treatment for urothelial cancer. ICIs have been proved to be effective with safe and tolerable side effects in a subset of patients with urothelial cancer, but the majority have primary disease progression [38, 39]. It is crucial to identify patients who are most likely to benefit from ICIs. Certain pathological markers assessed by immunohistochemistry have shown the potential to predict the treatment response of ICIs [40]. Noninvasive imaging biomarkers for optimal patient selection are still under investigation. Promising results have been reported recently that the radiomic signature of tumor-infiltrating CD8 cells could be useful in inferring clinical outcomes for patients with cancer treated with immunotherapy. But only eight patients with urothelial cancer were included in the study [41]. The full potential

of radiomics as a biomarker for immunotherapy in patients with urothelial cancer should be investigated and validated in a larger cohort and ideally in prospective randomized trials.

Conclusions

Urothelial cancer is among the most prevalent cancers worldwide and, as one of the most heterogeneous cancers known, needs a personalized approach for diagnosis and treatment. The development of radiomics to obtain quantitative features from imaging traits has shown the potential to aid diagnosis, guide therapy and monitor treatment response of urothelial cancer. To an extent, the applicability of radiomics in clinical practice depends on standardized data collection, evaluation criteria and reporting guidelines, and large-scale data sharing is fundamental for the full potential that radiomics represents. The research in radiogenomics of urothelial cancer and radiomics as a biomarker for immunotherapy has just started and needs further investigation. It is promising that radiomics-based decision support system for precision diagnosis and treatment for urothelial cancer will improve the quality of patient care in the near future.

Acknowledgements This study was funded by the Fundamental Research Funds for the Central Universities (Grant No. 3332018022); Beijing Municipal Natural Science Foundation (Grant No. 7192176); Basic Scientific Research Program of Chinese Academy of Medical Sciences (Grant Nos. 2019PT320008 and 2018PT32003); and National Natural Science Foundation of China (Grant No. 81901742).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Miyazaki J, Nishiyama H. Epidemiology of urothelial carcinoma. *Int J Urol*. 2017;24(10):730–4. <https://doi.org/10.1111/iju.13376>.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30. <https://doi.org/10.3322/caac.21442>.
- Mahdavi N, Ghoncheh M, Pakzad R, Momenimovahed Z, Salehiniya H. Epidemiology, incidence and mortality of bladder cancer and their relationship with the development index in the world. *Asian Pac J Cancer Prev*. 2016;17(1):381–6.
- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115–32. <https://doi.org/10.3322/caac.21338>.
- Advanced Bladder Cancer Meta-analysis C. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. *Eur Urol*. 2005;48(2):202–5. <https://doi.org/10.1016/j.eururo.2005.04.006>.
- Koshkin VS, Grivas P. Emerging role of immunotherapy in advanced urothelial carcinoma. *Curr Oncol Rep*. 2018;20(6):48. <https://doi.org/10.1007/s11912-018-0693-y>.
- Felsenstein KM, Theodorescu D. Precision medicine for urothelial bladder cancer: update on tumour genomics and immunotherapy. *Nat Rev Urol*. 2018;15(2):92–111. <https://doi.org/10.1038/nrurol.2017.179>.
- Theodora Katsila ML, Patrinos GP, Aristotelis B, Dimitrios K. The new age of -omics in urothelial cancer—re-wording its diagnosis and treatment. *EBioMedicine*. 2018;28:43–50. <https://doi.org/10.1016/j.ebiom.2018.01.044>.
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. <https://doi.org/10.1038/ncomms5006>.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–6. <https://doi.org/10.1016/j.ejca.2011.11.036>.
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749–62. <https://doi.org/10.1038/nrcli.nonc.2017.141>.
- Neri E, Del Re M, Paiar F, Erba P, Cocuzza P, Regge D, et al. Radiomics and liquid biopsy in oncology: the holons of systems medicine. *Insights Imaging*. 2018;9(6):915–24. <https://doi.org/10.1007/s13244-018-0657-7>.
- Spies PE, Agarwal N, Bangs R, Boorjian SA, Buyyounouski MK, Clark PE, et al. Bladder cancer, version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw*. 2017;15(10):1240–67. <https://doi.org/10.6004/jnccn.2017.0156>.
- Zhang G-M-Y, Sun H, Shi B, Jin Z-Y, Xue H-D. Quantitative CT texture analysis for evaluating histologic grade of urothelial carcinoma. *Abdom Radiol*. 2016;42(2):561–8. <https://doi.org/10.1007/s00261-016-0897-2>.
- Mammen S, Krishna S, Quon M, Shabana WM, Hakim SW, Flood TA, et al. Diagnostic accuracy of qualitative and quantitative computed tomography analysis for diagnosis of pathological grade and stage in upper tract urothelial cell carcinoma. *J Comput Assist Tomogr*. 2018;42(2):204–10. <https://doi.org/10.1097/RCT.0000000000000664>.
- Zhang X, Xu X, Tian Q, Li B, Wu Y, Yang Z, et al. Radiomics assessment of bladder cancer grade using texture features from diffusion-weighted imaging. *J Magn Reson Imaging*. 2017;46(5):1281–8. <https://doi.org/10.1002/jmri.25669>.
- Wang H, Hu D, Yao H, Chen M, Li S, Chen H, et al. Radiomics analysis of multiparametric MRI for the preoperative evaluation of pathological grade in bladder cancer tumors. *Eur Radiol*. 2019. <https://doi.org/10.1007/s00330-019-06222-8>.
- Nishiyama H. Asia consensus statement on NCCN clinical practice guideline for bladder cancer. *Jpn J Clin Oncol*. 2018;48(1):3–6. <https://doi.org/10.1093/jjco/hyx130>.
- Garapati SS, Hadjiiski L, Cha KH, Chan H-P, Caoili EM, Cohan RH, et al. Urinary bladder cancer staging in CT urography using machine learning. *Med Phys*. 2017;44(11):5814–23. <https://doi.org/10.1002/mp.12510>.
- Xu X, Zhang X, Tian Q, Wang H, Cui LB, Li S, et al. Quantitative identification of nonmuscle-invasive and muscle-invasive bladder carcinomas: a multiparametric MRI radiomics analysis. *J Magn Reson Imaging*. 2019;49(5):1489–98. <https://doi.org/10.1002/jmri.26327>.
- Tong Y, Udupa JK, Wang C, Chen J, Venigalla S, Guzzo TJ, et al. Radiomics-guided therapy for bladder cancer: using an optimal biomarker approach to determine extent of bladder cancer invasion

- from t2-weighted magnetic resonance images. *Adv Radiat Oncol*. 2018;3(3):331–8. <https://doi.org/10.1016/j.adro.2018.04.011>.
22. Lim CS, Tirumani S, van der Pol CB, Alessandrino F, Sonpavde GP, Silverman SG, et al. Use of quantitative T2-weighted and apparent diffusion coefficient texture features of bladder cancer and extravesical fat for local tumor staging after transurethral resection. *Am J Roentgenol*. 2019;212(5):1060–9. <https://doi.org/10.2214/ajr.18.20718>.
 23. Abol-Enein H, Tilki D, Mosbah A, El-Baz M, Shokeir A, Nabeeh A, et al. Does the extent of lymphadenectomy in radical cystectomy for bladder cancer influence disease-free survival? A prospective single-center study. *Eur Urol*. 2011;60(3):572–7. <https://doi.org/10.1016/j.eururo.2011.05.062>.
 24. Zehnder P, Studer UE, Skinner EC, Dorin RP, Cai J, Roth B, et al. Super extended versus extended pelvic lymph node dissection in patients undergoing radical cystectomy for bladder cancer: a comparative study. *J Urol*. 2011;186(4):1261–8. <https://doi.org/10.1016/j.juro.2011.06.004>.
 25. Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer. *EBioMedicine*. 2018;34:76–84. <https://doi.org/10.1016/j.ebiom.2018.07.029>.
 26. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A Radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. *Clin Cancer Res*. 2017;23(22):6904–11. <https://doi.org/10.1158/1078-0432.Ccr-17-1510>.
 27. Soukup V, Capoun O, Cohen D, Hernandez V, Burger M, Comperat E, et al. Risk stratification tools and prognostic models in non-muscle-invasive bladder cancer: a Critical Assessment from the European Association of Urology Non-muscle-invasive Bladder Cancer Guidelines Panel. *Eur Urol Focus*. 2018. <https://doi.org/10.1016/j.euf.2018.11.005>.
 28. Xu X, Wang H, Du P, Zhang F, Li S, Zhang Z, et al. A predictive nomogram for individualized recurrence stratification of bladder cancer using multiparametric MRI and clinical risk factors. *J Magn Reson Imaging*. 2019. <https://doi.org/10.1002/jmri.26749>.
 29. Meeks JJ, Bellmunt J, Bochner BH, Clarke NW, Daneshmand S, Galsky MD, et al. A systematic review of neoadjuvant and adjuvant chemotherapy for muscle-invasive bladder cancer. *Eur Urol*. 2012;62(3):523–33. <https://doi.org/10.1016/j.eururo.2012.05.048>.
 30. Cha KH, Hadjiiski L, Chan H-P, Weizer AZ, Alva A, Cohan RH, et al. Bladder cancer treatment response assessment in CT using radiomics with deep-learning. *Sci Rep*. 2017. <https://doi.org/10.1038/s41598-017-09315-w>.
 31. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm? *J Clin Oncol*. 2011;29(22):2951–2. <https://doi.org/10.1200/JCO.2011.36.1329>.
 32. Rutman AM, Kuo MD. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. *Eur J Radiol*. 2009;70(2):232–41. <https://doi.org/10.1016/j.ejrad.2009.01.050>.
 33. Blaveri E, Brewer JL, Roydasgupta R, Fridlyand J, DeVries S, Koppie T, et al. Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clin Cancer Res*. 2005;11(19 Pt 1):7012–22. <https://doi.org/10.1158/1078-0432.CCR-05-0177>.
 34. Blaveri E, Simko JP, Korkola JE, Brewer JL, Baehner F, Mehta K, et al. Bladder cancer outcome and subtype classification by gene expression. *Clin Cancer Res*. 2005;11(11):4044–55. <https://doi.org/10.1158/1078-0432.CCR-04-2409>.
 35. Birkhahn M, Mitra AP, Williams AJ, Lam G, Ye W, Datar RH, et al. Predicting recurrence and progression of noninvasive papillary bladder cancer at initial presentation based on quantitative gene expression profiles. *Eur Urol*. 2010;57(1):12–20. <https://doi.org/10.1016/j.eururo.2009.09.013>.
 36. Catto JWF, Abbod MF, Wild PJ, Linkens DA, Pilarsky C, Rehman I, et al. The application of artificial intelligence to microarray data: identification of a novel gene signature to identify bladder cancer progression. *Eur Urol*. 2010;57(3):398–406. <https://doi.org/10.1016/j.eururo.2009.10.029>.
 37. Smith SC, Baras AS, Dancik G, Ru YB, Ding KF, Moskaluk CA, et al. A 20-gene model for molecular nodal staging of bladder cancer: development and prospective assessment. *Lancet Oncol*. 2011;12(2):137–43. [https://doi.org/10.1016/S1470-2045\(10\)70296-5](https://doi.org/10.1016/S1470-2045(10)70296-5).
 38. Powles T, Smith K, Stenzl A, Bedke J. Immune checkpoint inhibition in metastatic urothelial cancer. *Eur Urol*. 2017;72(4):477–81. <https://doi.org/10.1016/j.eururo.2017.03.047>.
 39. Rijnnders M, de Wit R, Boormans JL, Lolkema MPJ, van der Veldt AAM. Systematic review of immune checkpoint inhibition in urological cancers. *Eur Urol*. 2017;72(3):411–23. <https://doi.org/10.1016/j.eururo.2017.06.012>.
 40. Buder-Bakhaya K, Hassel JC. Biomarkers for Clinical Benefit of Immune Checkpoint Inhibitor Treatment-A Review From the Melanoma Perspective and Beyond. *Front Immunol*. 2018;9:1474. <https://doi.org/10.3389/fimmu.2018.01474>.
 41. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol*. 2018;19(9):1180–91. [https://doi.org/10.1016/s1470-2045\(18\)30413-3](https://doi.org/10.1016/s1470-2045(18)30413-3).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cite this article as:

Chan S, Siegel EL. Will machine learning end the viability of radiology as a thriving medical specialty?. *Br J Radiol* 2019; **91**: 20180416.

REVIEW ARTICLE

Will machine learning end the viability of radiology as a thriving medical specialty?

¹STEPHEN CHAN, MD, MBA, MPH and ²ELIOT L SIEGEL, MD

¹Department of Radiology, Harlem Hospital Center and Columbia University, New York, NY, USA

²Department of Diagnostic Radiology and Nuclear Medicine, VA Maryland Health Care System, Baltimore, MD, USA

Address correspondence to: Dr Stephen Chan
E-mail: sc56md@gmail.com

ABSTRACT

There have been tremendous advances in artificial intelligence (AI) and machine learning (ML) within the past decade, especially in the application of deep learning to various challenges. These include advanced competitive games (such as Chess and Go), self-driving cars, speech recognition, and intelligent personal assistants. Rapid advances in computer vision for recognition of objects in pictures have led some individuals, including computer science experts and health care system experts in machine learning, to make predictions that ML algorithms will soon lead to the replacement of the radiologist. However, there are complex technological, regulatory, and medicolegal obstacles facing the implementation of machine learning in radiology that will definitely preclude replacement of the radiologist by these algorithms within the next two decades and beyond. While not a comprehensive review of machine learning, this article is intended to highlight specific features of machine learning which face significant technological and health care systems challenges. Rather than replacing radiologists, machine learning will provide quantitative tools that will increase the value of diagnostic imaging as a biomarker, increase image quality with decreased acquisition times, and improve workflow, communication, and patient safety. In the foreseeable future, we predict that today's generation of radiologists will be replaced not by ML algorithms, but by a new breed of data science-savvy radiologists who have embraced and harnessed the incredible potential that machine learning has to advance our ability to care for our patients. In this way, radiology will remain a viable medical specialty for years to come.

INTRODUCTION

Recent articles in the medical and lay press have underscored the tremendous progress made in “artificial intelligence”, and raised the prospect that computers, using machine learning (ML) algorithms, will soon replace radiologists.^{1,2} As recently as 2016, Geoffrey Hinton - founder of the branch of machine learning known as “deep learning” - was quite emphatic in stating this perspective, recently stating, “I think that if you work as a radiologist you are like Wile E. Coyote in the cartoon. You're already over the edge of the cliff, but you haven't yet looked down. There's no ground underneath. It's just completely obvious that in five years deep learning is going to do better than radiologists. It might be ten years.”² At that time, Hinton clearly indicated that machine learning would be a disruptive technology for radiologists. As described in Christiansen's seminal work,³ there are three essential characteristics of a disruptive technology, each of which is satisfied or could potentially be satisfied by machine learning [Table 1]. Since machine learning appears to fulfill these three essential characteristics, one could conclude that machine learning represents a disruptive technology. However, more recent work by

Christiansen et al suggests that there are two other criteria defining a disruptive technology.⁴ These include: (1) the presence of only a low-end foothold or a new market foothold in the industry; and (2) the unknowing or deliberate ignorance of the new technology by the incumbent leaders in the industry. As indicated later in this article, we believe that neither of these latter two criteria is met by machine learning in radiology. With respect to machine learning, it now appears that leading radiology organizations have begun to adopt strategies for handling this potentially disruptive technology.⁵

In 2016, Chockley (a medical student) and Emanuel (an internal medicine physician and “Obamacare” architect) identified three threats to the future practice of diagnostic imaging, with machine learning singled out as the “ultimate threat”.⁶ They made the following two assertions: (1) “machine learning will become a powerful force in radiology in the next 5 to 10 years, not in multiple decades”; and (2) “indeed, in a few years there may [be] no specialty called radiology”.⁶ If they meant that the computer will largely replace the radiologist in 5 to 10 years (as implied in

Table 1. Three fundamental characteristics of a disruptive technology (as related to machine learning)

Key characteristics of disruptive technology	Is this true of machine learning and why?
The overall performance level offered by early versions of the disruptive technology is far inferior to the current technology.	Partially true. As of 2018, there is no version of an ML algorithm whose performance can match the accuracy and breadth of a human radiologist.
The customers currently served by the incumbent industry leaders often provide little (or even negative) feedback about the value of the new technology.	True. No one in the current generation of clinicians is requesting that radiology interpretations be provided solely by ML systems.
The customers who benefit most from the emergence of a new technology with inferior performance characteristics are often different from the ones currently served by the market leaders.	Probably true. Initial customers for ML systems have not yet been identified, although they may include clinicians (or hospitals) from developing nations, research subjects from population health studies, or large corporations with preventive health imaging needs.

ML, machine learning;

their work), then we completely disagree and believe that this is an ill-informed prediction borne out of a lack of domain knowledge of radiology. Their view reflects a fundamental misunderstanding of the nature of the work performed by radiologists, as well as a lack of appreciation of exactly how difficult it will be for machine learning to replace the wide variety of imaging interpretation and patient care tasks inherent in the practice of radiology. We note that the specter of a future in which radiologists are no longer needed to provide image interpretation services has seriously alarmed forward-thinking medical students, radiology residents, and fellows, impelling some to ask if they should quit or avoid radiology residency because of the risk of not getting a job after residency.^{7,8} Indeed, that fear could potentially damage the radiology profession by discouraging talented medical students from choosing radiology as their future career. We seek to allay such fear by careful examination of the recent developments in machine learning, and by detailed evaluation of the kind of technological development necessary to render the broad range of radiological diagnosis. Specifically, there are two fundamental sources of misunderstanding that lead many individuals to conclude that radiologists can be easily replaced by machine learning.

Misunderstanding #1: Machine learning can easily absorb and process the wide variation of information and ambiguity inherent in interpretation of medical images.

Remarkable achievements have been made in machine learning such as the impressive computer vision performance on identification of objects in everyday pictures from the Stanford ImageNet challenge⁹ and the victory of Google's AlphaGo over the 2016 human champion of Go.¹⁰ Computer scientists cite these accomplishments to assert that unsupervised machine learning will soon be rendering medical imaging findings and diagnoses. However, board games such as "Go" focus on a very "narrow" artificial intelligence task where a winning vs losing status can be assessed, whereas medical imaging is associated with far greater

amount of ambiguity, and a larger variety of features, classifications, and outputs. It is also likely that thousands of "narrow" algorithms based on separate large, well-annotated databases will be required for a computer to begin to compete with a radiologist for comprehensive diagnostic assessment of even a single modality covering a single anatomical region of the body.

Advances in self-reinforcement learning have led to substantial further improvements in "AlphaGo" resulting in "AlphaGoZero" which utilizes an approach in which the computer is provided with the basic rules of the game and learns by playing itself large numbers of games rather than learning by analyzing the play of human experts.¹¹ Although possible in games with simple defined rules such as Go or chess, analogous self-reinforcement learning is not so easily attainable in radiology, given the lack of a simple set of rules of the "radiology game" to allow this sort of self-play. Barring an unforeseen major technological breakthrough, it is likely that human annotation and guidance will likely be necessary at multiple stages in the development of machine learning in medical imaging, augmented by increases in computing power and conceptual advances in artificial intelligence. This pattern is exemplified by the technological development of the Google Translate app in which significant conceptual advances in ML-based language translation needed to be made by computer scientists, who then were able to render the sequential and contextual information inherent in languages far more amenable to deep neural networks.¹

Misunderstanding #2: Computer-aided detection and computer-aided diagnosis is an immediate technological precursor to ML algorithms.

Chockley and Emanuel cite the current performance of computer-aided detection (CADe) and computer-aided diagnosis (CADx) in various areas of radiology - including the field of mammography—as evidence of success stories, with machine learning "working as well as or better than the experienced radiologist". Indeed, many papers and presentations describing CAD systems in mammography have claimed a performance level in lesion detection similar to that of an experienced radiologist.¹²⁻¹⁴ Based on that research, CAD was approved by the Food and Drug Administration (FDA) for use with mammography and has been widely introduced into radiology practices across the U.S. as an adjunctive technology for mammography.¹⁵ However, in spite of its widespread use for the past decade, it has not been shown to improve detection rates in academic settings, and it is unclear whether or not CAD improves the detection rate of invasive breast carcinoma in community practice.¹⁶ In addition, the use of CAD can be detrimental if its limitations are not understood.¹⁷ While review of mammographic images with adjunctive CAD would likely be considered the de facto standard of care in community mammography practice,¹⁸ we note that CAD systems have not replaced the practicing radiologist. In practice, survey data suggests that more than half (~62%) of radiologists have never or rarely changed their report as a result of CAD findings in mammography, and about a third of radiologists never or rarely use the findings generated by CAD.¹⁹ There has been an initial demonstration of a machine learning tool to help separate

high-risk breast lesions that are more likely to become cancerous from those that are at lower risk.²⁰ However, we know of no CAD program in clinical use that continually receives feedback about its diagnostic performance – a task that is essential to learning from experience. Finally, we are not aware of any mammography CAD/machine learning software program that formally compares a prior mammogram to a current mammogram, just as a human reader would do. Yet, comparison with prior imaging studies remains a fundamental diagnostic task in mammography and radiology, especially in assessment of interval change.

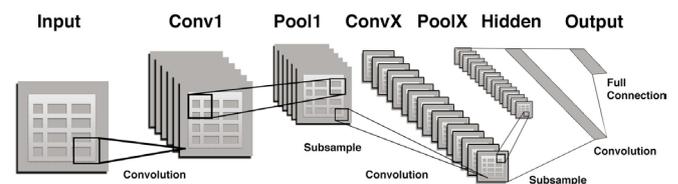
What is machine learning?

The term “machine learning” encompasses a variety of advanced iterative statistical methods used to discover patterns in data and, although inherently non-linear, is based heavily on linear algebra data structures. It can be utilized to help to improve prediction performance, dynamically adjusting the output model when the data change over time. Historically, there have been two very broad groupings of artificial intelligence applied to cognitive problems in everyday human work issues. The first is expert systems, in which software programs are constructed to mimic human performance, based upon rules that were derived from “experts” by the programmer. An example of this was the medical diagnostic program “Internist I”, which was designed to capture the expertise of the chairman of internal medicine at the University of Pittsburgh School of Medicine, Dr Jack Myers.²¹ The second is machine learning, in which the most recent advances in computer vision and speech recognition have come from a form of machine learning known as “deep learning”, which uses a technique known as convolutional neural networks, and is based on a set of algorithms that attempt to model high-level abstractions in data. Neural networks use a variety of approaches loosely based on what are referred to as interconnected cells, analogous to the interneurons of the human nervous system.

Convolutional neural networks (CNNs) are a special type of neural network that is optimized for image pattern recognition.²² Unlike other types of artificial neural networks, the majority of nodes (neurons) in a CNN are only connected to a subset of other nodes, particularly those in closer proximity in an image which enhances their ability to recognize local features of an image. In brief, a CNN consists of multiple layers between the input and output layers (Figure 1) The main building blocks are the convolution layer which can be thought of as a series of adjustable image filters that can emphasize or de-emphasize certain aspects of an image such as borders, colors, noise, and texture. Each of the multiple convolution layers within a CNN is followed by a pooling layer which serves to reduce the number of parameters. For example, the commonly used technique referred to as “max pooling” simply chooses the maximum pixel value within each small portion of an image and assigns all pixels to that value. The final layer of a CNN is a fully connected one similar to other types of artificial neural networks.

The term “artificial intelligence (AI)” is currently commonly utilized in medical imaging in both the lay and scientific literature to refer to machine learning in general and CNNs specifically. Although the architecture of both CAD programs and

Figure 1. A portion of the input (image) is passed to each successive pair of convolutional/pooling layers (filter/parameter reducers) with several convolution and pooling layers added before an output (prediction) is made. Initial layers tend to represent general features such as edges and colors and later layers represent features of increasing complexity, such as corners, and then textures, followed by even complex features such as a snout or whiskers, and finally entire objects such as a dog or cat. Finally, there is a “fully connected layer” that flattens input from other layers transforming them into a decision on whether the output belongs to a certain class (e.g.dog vs cat). Errors in classification in a training set are then “back-propagated” to modify and/or update the filters so that overall errors are minimized.



ML-based algorithms are designed by humans, the essential discriminatory functions of the AI algorithms emerge directly from the data, and, unlike CAD, do not require humans to identify and compute the critical features.²³ This emergence of algorithms from the data is what prompted Wired magazine to suggest that machine learning may represent “the end of code”.²⁴ Specifically, the algorithms to predict such things as the presence of an intracranial bleed, or malignancy in a prostate MRI study will emerge directly from the “learned”, iteratively adjusted values of the nodes in a CNN. Those values themselves represent the trained model and the “training” continues with the introduction of each annotated dataset. Although inputs to CNNs are not always raw images (and may be segmented and co-registered prior to classification), the many steps such as feature extraction, segmentation, registration, and statistical analysis utilized by the previous generation of so-called CAD (computer-aided detection or computer-aided diagnosis) software are not required. Both AI and CAD techniques can be utilized to develop medical imaging software, but AI algorithms typically require much more annotated data but subsequently take much less time to develop using fewer steps. Most developers previously utilizing these more human, understandable basic steps have made the transition to the use of CNNs. Creation of a machine learning model can be performed much faster (e.g. 5 to 6 days rather than 5 to 6 months or years) than traditional computer-aided detection and diagnosis (CAD_d/CAD_x).

The nature of learning in machine learning can be confused with that associated with humans. Machine learning has been defined as “algorithm-driven learning by computers such that the machine’s performance improves with greater experience” and indicated that it “involves the construction of algorithms that learn from data and make predictions on the basis of those data.”²⁶ Although this definition may imply a type of self-reinforcement learning “with greater experience”, in actuality, these algorithms in diagnostic imaging have improved largely by the addition of annotated data based on human review or patient

outcome information, rather than repetitive application of work-in-progress. This is a fundamental difference from the use of machine learning in strategy games such as Chess and Go in which there are well-defined parameters of success. The current regulatory constraint on progressive learning of these algorithms is that clearance of these products has been based on a well-defined training set and a test set (in order to establish performance against ground truth). Therefore, from a regulatory perspective, it is unclear whether or not the FDA would allow the continued modification or improvement of an ML system by incorporation of additional local patient data from a given clinical radiology practice. As is true of other related statistical techniques such as linear regression, additional “processed” data are required to enhance the model rather than simply “learning from experience” *per se*. Therefore, computers remain far less efficient than humans at learning and generalizing concepts from a relatively small dataset. This has been mitigated somewhat by techniques such as transfer learning which takes patterns learned from a related task as a starting point effectively kickstarting the training (also known as “one-shot learning”) that takes advantage of existing knowledge to train using just a single or very few examples.²⁵

Will an ML system soon be capable of replicating the work of a radiologist?

Major technological developments in machine learning have been made over the last few years, including advances in deep learning algorithms, further advances in graphics processing units speed and memory, and the exponential growth of corporate investment. However, there are several independent factors which suggest that successful replacement of the radiologist’s work is likely to be substantially more difficult than is currently envisioned by some non-radiologist health care experts and computer science futurists. These potential challenges are based upon unique aspects of the radiological image, the visual processing capability of the radiologist, and the role of the radiologist in maximizing and maintaining clinical relevance in image interpretation. In particular, the job of the radiologist is not simply to detect findings related to various imaging studies, but to determine “what is wrong with this picture” and help determine the future course of action in the diagnostic evaluation and therapeutic decision-making. Determining “what is wrong with this picture” is a much harder task that extends far beyond the capabilities of the current generation of computer vision systems. Contextualization of the imaging information in diagnostic evaluation and therapeutic decision-making may be an even more difficult task to replicate.

A major challenge for ML algorithms is the greater technical complexity of the radiological image as compared to those images typically used in object-recognition tests for computer “visualization”. In addition to the differing imaging modalities, this complexity includes a wide variety of manifestations of normal and pathological findings, multiple sequential images in a cross sectional/volumetric dataset, with much higher complexity of data and raw number of pixels/voxels in medical images. It also includes a high level of ambiguity and difficulty in annotation that is not inherent in the ImageNet challenges that have used

common objects such as dogs, cats, bikes, cars, etc. Another major technical challenge is the development of a “reasonable” detection rate of abnormalities without an excessive rate of false-positive findings as compared with human performance. For more than 20 years, CADE and CADx programs, such as those used to detect lung nodules or breast masses, have been fraught with the issue of frequent false-positive findings (*i.e.* low specificity) and we suggest that this problem may also be an intrinsic problem for deep learning algorithms.^{22,26} This problem is further complicated by: (1) the multiple classes of imaging abnormalities detected on diagnostic imaging studies; (2) the time and expense associated with the collection of large annotated datasets (such as ChestX-ray14 and the Cancer Imaging Archive) required for deep learning, of which a fair number are available in the public domain,^{27,28} but many more are needed²⁹; (3) the difficulties associated with ensuring sufficient, detailed image annotation; and (4) the rapid changes in imaging technology (*e.g.* 2D to 3D mammography tomosynthesis) that makes a multiyear annotation effect obsolete due to major technological improvements in imaging modalities. All of these challenges must be addressed before machine learning can replicate the work of a radiologist.

What are the technical details underlying challenges in object recognition and identification of abnormalities on diagnostic imaging studies?

First, it is true that computers with deep learning algorithms have approached human levels of performance in object recognition – as demonstrated in the Stanford ImageNet Large Scale Visual Recognition Competition (ILSVRC).⁹ However, object recognition is a necessary but not sufficient prerequisite to performing this task on medical imaging studies. The set of validation images used in the ILSVRC are characterized by lower resolution, fewer classes and instances of objects per image, and larger objects, as compared to those features on the typical medical image used for diagnostic purposes. Stated another way, the task of object recognition on medical images is far more difficult because the objects (*i.e.* imaging findings) are more numerous, more varied, and far more complex than those on the standard test images for the ILSVRC. The issues of greater resolution, increased frequency of objects per unit space, and wider variety of object shapes and characteristics on medical images together pose a far greater challenge for computer-based object recognition than those posed by simple recognition of discrete objects. Medical evaluation of imaging findings typically requires analysis of multiple features, requiring several levels of analysis beyond object detection and classification (extending beyond the classic visual task of discriminating “dog vs cat”). Unless this learning algorithm can be trained with hundreds or thousands of additional algorithms to distinguish varying features of a recognized object, it will not yield any useful information about such questions. In the medical imaging realm, many kinds of imaging pathology require detailed analysis of a combination of features, likely requiring a greater degree of testing and validation, as well as an ensemble of multiple narrow algorithms. However, we recognize that focused applications of deep learning to specific medical imaging problems have already been devised and evaluated, especially in the fields of cardiothoracic imaging and breast imaging.^{30,31}

In order for an ML system to replicate fully the multifactorial nature of the radiologist's assessment of an image (for example, a chest radiograph), it will likely need to be trained not by a single large dataset (containing many disparate types of radiographic abnormalities), but by the presentation of multiple datasets that specifically reinforce the learning associated with each class of imaging abnormalities (such as cardiac, mediastinal, pulmonary, and osseous) as well as additional datasets with various important subclasses of imaging abnormalities (for example, congenital heart disease). The final aggregate of the multiple datasets for chest radiographic images will need to be extremely large and extensively annotated, in order to ensure that the computer's experience matches both the depth and breadth of the radiologist's knowledge. Of course, a less ambitious training approach could be devised to ascertain whether a radiograph is normal or abnormal for triage purposes, but this approach would not replicate the bandwidth and detailed accuracy of expert performance.

Another major problem is the establishment of a gold standard. For example, within a large dataset of chest radiographs in patients suspected to have tuberculosis, there may be variability among several clinical radiologists in image interpretation. In clinical practice, one individual radiologist may want to not miss a case of tuberculosis due to its high clinical impact and thus would annotate cases as positive with subtle/non-specific findings of TB, while another radiologist may not want to overcall tuberculosis and may instead look for the more classical signs specific to the disease. Thus, when creating a predictive machine learning model, does one attempt to create different radiologist "personas" (e.g. high sensitivity vs high specificity profiles), or predict what a specific radiologist will report, or somehow create a middle-of-the-road report or "consensus" report? Alternatively, does the annotation of the final outcome of an imaging study get labeled as the actual sputum lab result or the actual clinical outcome? If so, then cases that are obviously normal or obviously strongly suggestive of TB will be labeled differently due to the clinical outcome. (In a recent academic study on this topic, the combination of sputum results, original radiologist interpretations, and confirmation by a single overreading, expert radiologist was required for inclusion into the pulmonary TB database.)³⁰ Finally, is the task to predict how a specific radiologist performs or how an "average" radiologist performs in interpretation of a radiograph or in prediction of the clinical outcome? If the goal is to predict clinical outcome, then issues such as prevalence of disease in a particular population may weight too heavily on the performance of the system. All of these questions raise important, clinically relevant issues that have not yet been resolved.

In machine learning, the computer's greatest strength - its abilities to process data endlessly and to repeat the same steps without tiring - could also represent a type of Achilles' heel. This problem is due to the issue of overfitting - defined as the functioning of a learning model (or prediction model) that fits so well with its training dataset to the extent that it models the statistical noise, fluctuations, biases, and errors inherent in the dataset, negatively impacting the performance on new data (i.e. diagnostic imaging studies not previously presented). This is more likely to occur in medical imaging than in other computer vision applications due to the relatively large number of categories of normal and abnormal findings and limited

numbers of annotated training sets. More succinctly, Domingos indicates that overfitting has occurred "when your learner outputs a classifier that is 100% accurate on the training data but only 50% accurate on test data, when in fact it could have output one that is 75% accurate on both"²² While the notion of accuracy in machine learning was relatively simple in the reported studies of object recognition, we note that radiology has a rich scientific history of measurement of diagnostic accuracy, including the development of receiver-operating characteristic (ROC) analysis.³²⁻³⁴

Classifier performance is central to making informed decisions about machine learning, and yet the typical use of a single measure of diagnostic accuracy, while simple, is inadequate for technical evaluation. Publications of medical machine learning studies are much more informative and rigorous when they utilize ROC analysis because its measures of sensitivity and specificity are not dependent on prevalence of disease (as is true of accuracy). In addition, the measure of diagnostic accuracy is typically derived from use of a single arbitrary threshold, whereas ROC analysis demonstrates the performance using all known threshold values. However, since the prevalence of a disease does affect the performance of any diagnostic classifier, it would also be helpful to know the prevalence of the disease in the test population, so that the false-positive and false-negative rates could be determined. Precision, which roughly translates as the likelihood that a positive test means that the disease or finding is truly present (otherwise known as the positive predictive value), can demonstrate the relative strength or weakness in a classifier for findings or diseases that are low prevalence.^{26,35}

The problem of overfitting in medical imaging is also magnified by the wide variety of "odd" shapes of normal structures, and the myriads of anatomical variants related to extra or missing anatomical structures (such as accessory ossicles or congenitally absent or hypoplastic structures). This problem is made most evident by considering the problems faced by a radiology researcher who is collecting and classifying the many types of anatomical structures and abnormalities that are found on chest radiography. That researcher would have to obtain images and related data for the computer to demonstrate abnormalities of the heart, mediastinum, lungs, bones, pleura, and various other structures. Distinguishing anatomical variants from pathological entities has been an important function of the practicing radiologist, with a whole atlas devoted to helping them avoid making a false-positive diagnosis.³⁶ In other scientific fields, such as the field of genomics, there has been recognition of the unacceptably high "false-positive" rate associated with various kinds of "wild-type" variations that mimic findings associated with genetic mutations associated with cancer.³⁷ In one study of ML algorithms devoted to this problem, they characterized the types of false-positive errors into six different groups and suggested that "feature-based analysis of 'negative' or wild-type positions can be helpful to guide future developments in software".³⁷ This is akin to the problem with anatomical variants in diagnostic radiology.

Because the deep learning approach is highly complex, and because no method has been developed that allows a given algorithm to "explain" its reasoning, technology experts are generally

not able to understand fully the reasons for the algorithm's conclusions, and not able to predict the occurrence and frequency of failure or error in performance of the algorithm.³⁸ Therefore, validation and regulatory approval could take more time due to the "black box" nature of machine learning approaches. Fortunately, major advances have been made in recent years in illuminating the contents of the CNN black box.³⁹ One such advance, saliency maps, was originally proposed in 1998, and is based on the "feature-integration theory" of human visual attention.⁴⁰ In 2013, two image visualization techniques for visualization inside deep convolutional networks were demonstrated, one of which involved saliency maps.⁴¹ For a given output category value (*e.g.* a type of interstitial lung disease), saliency maps display the pixels of the image (*e.g.* CT of the thorax) that were most important for image classification. More recently, other more sophisticated techniques have been developed that organize non-human interpretable convolution layers into an explanatory and potentially interactive graph or image that can be used to speed up the learning process and identify inaccuracies or important areas of an image ignored by a CNN allowing refinement of the model and improving performance.^{39,42}

In contrast, CAD algorithms have been developed over several decades, many of which are focused on specific clinical imaging problems, and therefore have relatively narrow imaging applications. Examples of these applications include: (1) fracture detection, bone age determination, and bone mineral density quantitation in orthopedic radiology; (2) brain hemorrhage detection, multiple sclerosis detection and quantitation, and regional brain segmentation and volumetry in neuroradiology; and (3) coronary and/or carotid artery stenosis evaluation, and cardiac function assessment in cardiovascular radiology. In order for an ML system to replicate the performance of a radiologist, it would have to incorporate large portfolios of narrow ML algorithms, each of which has been devised to answer a specific clinical question. The use of combinations of algorithms to solve a single narrow machine learning problem or problems has been referred to as ensemble methods in machine learning and has been successful in winning machine learning competitions on classification of complex datasets. Yet the integration and orchestration of such a wide and varied array of learning algorithms - possibly from several different developers - into a single clinical system would likely require substantial amounts of time and effort in validation and testing (according to the "no free lunch" theorem of ensemble learning),⁴³ not to mention the potential regulatory challenges. In the field of artificial intelligence, the "holy grail" is to devise a form of "general artificial intelligence", which could replicate average human intelligence. General artificial intelligence, as opposed to a collection of narrow artificial intelligences, could help overcome this technological hurdle. Unfortunately, the majority of computer scientists do not believe that generalized artificial intelligence will emerge in the next 20 years, if ever. However, there are other ways that narrow artificial intelligence can help to improve the radiology work process, aside from diagnostic interpretation. There is a wide range of opportunities to increase operational efficiency, improve the radiology workflow, and provide decision support to clinicians and radiologists.

Is it likely that the job of the practicing radiologist is going to be completely displaced by artificial intelligence in the near future?

Acemoglu and Autor devised a $3 \times 2 \times 2$ matrix model by which "work" can be classified, according to whether it is based upon (1) low, medium, or high skills; (2) cognitive or manual labor; and (3) routine or non-routine tasks.⁴⁴ Based upon their analysis, they found two interesting results relevant to a radiologist. First, the rapid diffusion of new technologies which substitute capital for labor - such as computerization - resulted in decreased demand for work based upon routine tasks. This effect was present whether the work is cognitive or manual, but was predominantly found among workers with medium-skill levels. Interestingly, the types of workers found to be more resistant to job displacement included financial analysts (a non-routine, cognitive job) and hairdressers (a non-routine, manual job). (We do note that the asset management industry is devoting substantial economic resources - even more than that devoted to radiology - to incorporate artificial intelligence into financial analysis.⁴⁵) With respect to routine interpretation tasks performed by the practicing radiologist, it is likely that an ML system will soon perform some of the routine image interpretation tasks (for example, lung nodule screening or pre-operative chest radiography). However, many of radiologists' highly skilled work tasks, especially in complex image pattern recognition, will be more difficult to replicate over at least the next two decades, and therefore will require more time for adequate dataset generation and training, validation, and performance testing. This suggests that those radiologists who have acquired higher levels of skills (such as higher degrees of subspecialization, or greater experience in narrow, focused areas of clinical imaging) would be even more resistant to job displacement. Second, "technical change that makes highly skilled workers uniformly more productive" results in a lowering of the threshold for task difficulty that separates the medium-skill worker and the high-skill worker.⁴⁴ Therefore, in the face of potential displacement of radiologists from some image interpretation tasks, many radiologists will increasingly spend a higher percentage of time on other valuable radiology-based tasks. These radiology-based tasks include those listed in the ACR 3.0 Initiative, such as: consultation with referring physicians; timely oversight of ongoing complex imaging studies; direct patient contact including discussion about test results; verification of adherence to national imaging guidelines for proper test ordering; participation and data collection for radiology quality initiatives; and timely review of radiology-based patient outcomes.⁴⁶ The potential shift in the proportion of imaging interpretation activities in the daily work of the radiologist is also in keeping with the findings of the 2017 McKinsey report on the effects of automation on employment and productivity. While over half of all occupations have at least 30% work activities that could be automated, no more than 5% of all occupations could be entirely automated; this indicates that far more jobs will change than will be eliminated by automation.⁴⁷ In particular, the report states "high-skill workers who work closely with technology will likely be in strong demand, and may be able to take advantage of new opportunities for independent work." For radiologists, this could potentially include renewed focus on the entire spectrum of patient care in imaging. It is likely that new kinds of jobs for

radiologists will arise as a result of machine learning, similar to the way online retail activities led to both a decreased need for marketers and sales staff, and a tremendously increased demand for data scientists able to perform the data-mining activities needed to assess consumer wants and satisfaction.

For some expert radiologists, particularly those situated along the frontiers of their radiological subspecialties, there is also the possibility of being involved with a higher proportion of non-routine clinical work, including the interpretation of more complex imaging technologies that are found to be much more difficult to encode into an ML system. It is far less likely that sufficiently large datasets could be generated to provide neural networks the “experience” to answer questions about less common clinico-pathological entities, or to deal with non-routine clinical issues that often arise in medical practice. Therefore, there will remain an important role for the expert radiologist who can deal with the non-routine clinical work. This viewpoint is expressed by two experts in information systems and economics: “While computer reasoning from predefined rules and inferences from existing examples can address a large share of cases, human diagnosticians will still be valuable even after Dr Watson finishes its medical training because of the idiosyncrasies and special cases that inevitably arise. Just as it is much harder to create a 100 percent self-driving car than one that merely drives in normal conditions on a highway, creating a machine-based system for covering all possible medical cases is radically more difficult than building one for the most common situations”.⁴⁸

Finally, there is significant uncertainty as to whether or not certification by governmental regulatory agencies would initially allow these systems to operate autonomously, as opposed to requiring oversight by human radiologists. Similar to the steps established for CAD in mammography almost two decades ago, we believe that ML systems will, for the foreseeable future, be approved only for adjunctive use with radiologist oversight, over which time it could become the norm for machines and humans to work together in imaging study interpretation. At first, this may manifest as “worklist triage” in which cases suspected to be more likely to be abnormal by an ML algorithm will be prioritized for human interpretation.

Obtaining regulatory (FDA) clearance will continue to be an arduous process during the initial introductory phase of ML systems into the clinical care environment, because of all the intricate details involved in validation and approval of a plethora of ML systems. The FDA will likely need greater time, resources, and expertise to evaluate a completely different kind of imaging-based technology, and to understand the ramifications of a system wherein the underlying work processes—the learning algorithms themselves—are relatively opaque (*i.e.* a “black box”). Even after FDA approval, user acceptance of ML systems could be adversely affected if systemic errors or deviations are detected that cannot be explained. This suggests that post-market surveillance could become a more important feature with these systems.

The FDA’s “Clinical and Patient Decision Support Software draft guidance” issued in December 2017 exempts software that

provides decision support that merely makes it easier to perform simple calculations or retrieval of accessible data. However, deep learning applications are thought to be “black box” and thus must be FDA regulated.⁴⁹ In response to these challenges, the FDA has recently begun to make significant strides toward making the clearance process less onerous. One FDA guidance draft document, “Expansion of the Abbreviated 510(k) Program: Demonstrating Substantial Equivalence through Performance Criteria”⁵⁰ makes 510(k) clearance easier by allowing manufacturers to establish “substantial equivalence” functionally using performance metrics rather than requiring direct comparison testing and the same technology.⁵¹ A few companies have managed to obtain FDA clearance for their deep learning-based algorithms related to diagnostic imaging and diagnostic testing. Arterys (San Francisco, CA) was the first company to receive clearance by the FDA for a deep learning application (for a suite of oncology software for automated segmentation of solid tumors on liver CT and MRI scans, and lung CT scans), thereby setting a precedent for other applications using CNNs.⁵² Also, as of August 2018, the FDA has recently approved clinical decision support software for alerting providers of a potential stroke in patients,⁵³ an algorithm for detection of wrist fractures,⁵⁴ and an AI-based device to detect certain diabetes-related eye problems.⁵⁵

What is the likely pathway of incorporating machine learning into radiology practice?

Even if the use of machine learning technology throughout society continues to increase exponentially, it is not at all clear that ML algorithms in a relatively well-defined field such as medical imaging will necessarily experience such astronomical growth. Advances in computational speed may only guarantee that the same answer—including the wrong answer—could be provided 1000 times faster, unless there are new techniques or new insights that emerge with approaches to deeper neural networks or future approaches such as Bayesian deep learning networks. Currently, machine learning for various image recognition algorithms requires presentation of many well-annotated imaging studies by human researchers, who then periodically test each algorithm for reliability and accuracy. Large imaging datasets will need to be developed and shared across institutions and radiology practices; this is an activity that requires work and trust to overcome technological, institutional, and regulatory barriers. The longstanding requirements of the medical field for high levels of diagnostic accuracy (as measured by sensitivity and specificity) and precision in differential diagnosis, will likely serve both as important benchmarks by which to judge the usefulness of these computer-aided diagnostic algorithms, and as essential “brakes” to the otherwise headlong rush to introduce labor-saving technology to reduce costs. The incorporation of these machine learning programs into the medical arena will likely be more gradual than in other sectors such as industrial, financial, chemistry, astronomy, etc., with a reasonable likelihood of a monotonic increase in the rate of progress over the years. Our healthcare system is a complex adaptive system and change in portions of that system—such as in the radiology industry—is typically characterized by “punctuated equilibrium”—*i.e.* relatively long periods of incremental change, interrupted by relatively short bursts of intense change.⁵⁶ Thus, the one caveat that

we make to our prediction of gradual incorporation of machine learning would be the advent of an earth-shattering technological innovation in generalized artificial intelligence—such as the invention of “the master algorithm,” which is that universal learning algorithm that can be applied to disparate fields of knowledge, and yet still make robust, accurate predictions, when supplied with sufficient, appropriate data.⁵⁷ Only in that case would we suggest that machine learning has become a “10X force”—a change in the business force so large that it exceeds the usual competitive influences by an order of magnitude.⁵⁸ This “sea-change” would then motivate radiologists to prepare for an upcoming “strategic inflection point”—that point in time when the old ways of doing business and competing in the marketplace are no longer favored, and a new strategic paradigm takes over.⁵⁸ However, the history of science indicates that the timing of such an invention cannot be predicted in advance and likely will not occur any time soon.

In order to support their cognitive processes, current practicing radiologists have already learned to incorporate various kinds of technology, including quantitative analysis, three-dimensional imaging display tools, collaborative tools for consultation, and digital imaging resources. Future AI tools hold the promise of further expanding the work that radiologists can do, including in the realms of precision (personalized) medicine and population management. Rather than replacing radiologists, future AI tools could advance the kind of work that radiologists perform; this would be in line with the classic IBM Pollyanna Principle: “Machines should work; humans should think.”^{59,60} At the 2016 meeting of the Radiological Society of North America (RSNA), Keith Dreyer proposed that the future model of the radiologist is the “centaur diagnostician”; such a physician would team up with the ML system to optimize patient care.⁶¹ This idea follows the observation that the performance of human-machine teams in playing chess could exceed that of a human or a machine system alone.⁶² This partnership would yield greater precision and detail in their imaging-based report, including more quantitative information and evidence-based recommendations.⁶¹ In addition, this could help facilitate advanced visualization techniques, refine clinical-radiological work procedures, and improve the timeliness and quality in communication between the radiologist and referring physician, as well as between the radiologist and patient. By viewing ML systems as a collaborator, not as a competitor, future radiologists could benefit from a partnership where the combined performance of the radiologist-computer team would likely be superior to either one alone, and feel enriched by the “luxury” of working with the advanced technological support offered by machine learning. In addition, the computer could allow the human to do more of what he or she does best – such as judicious use of the cognitive abilities associated with curiosity, experimentation, and insight. Just as in the example of Advanced Chess, it seems likely that the ability to work effectively with the computer will become a distinct competitive advantage. The futurist Kevin Kelly suggests that we cannot race against the machines, but that we can race with the machines. His conclusion is even more succinct: “You’ll be paid in the future based on how well you work with robots.”⁶³ This whole concept is also being embraced in various industries, as

well as in medicine, including the explicit re-definition of “AI” by the American Medical Association as standing for “augmented intelligence” rather than “artificial intelligence.”⁶⁴

We believe it likely that machine-based learning systems will need oversight for a great many years because of the potential for many different kinds of errors on various kinds of imaging studies. In addition, most current medical imaging algorithms are not equipped with the basic knowledge and skills in human anatomy, physiology, and pathology. If we do reach a point when we might expect that machine-based systems approach the accuracy and reliability of a practicing radiologist, then it will become a societal issue as to whether or not diagnoses based solely upon machine learning are acceptable. If this is viewed solely as a technological upgrade, and if society has already accepted other innovations such as self-driving cars, then this change may not be controversial. On the other hand, if there is significant adverse public reaction to the loss of human interaction in the realm of medicine, then it is possible that radiologists may not be displaced for a very long time, if at all. Along these lines, Verghese et al have issued a strong call for the computer and the physician to be working together for the foreseeable future and have given a warning about the unintended consequences of the implementation of new technology.⁶⁵

Given the expected retirement of increasing numbers of baby-boomer radiologists over the next two decades and the growing emphasis on screening and maintenance of health, it is likely that there will be a need for more radiologists over the next 20 years, and that computers will increasingly be regarded by those radiologists as trusted partners. The ML systems will be able to help create preliminary reports and note additional findings that may not make it into the final report, but, as is true of CAD today, computers would not be primarily responsible for the final reports. There will be a requirement for much more academic work to be done by human radiologists, including knowledge sharing and transfer learning, even before reaching the stage where the machine-based learning programs can become true partners in the imaging interpretation process.^{66,67} Over the last few years, the RSNA R&E Foundation has received an increasing number of submissions of research and education grant applications (1 in 2015, 3 in 2016, 9 in 2017, and 27 in 2018) which involve the development of artificial intelligence in radiology, including machine learning.⁶⁸ In addition to educational offerings at various universities in the US and around the world (whether as part of degree-granting programs, certificate-based programs, or online training), there are also several developmental opportunities for physicians (whether internships or jobs) at various technology-based corporations in the US. The involvement of radiologists in machine-based learning in radiology will be critical in assuring that the care of future patients is not compromised by errors of commission or omission. While not yet part of the radiology curriculum for trainees, it is not hard to imagine that training in radiology informatics is likely to become an even more central component of radiology residency education. The first step has been taken by organized radiology with the development of a specific training program in radiology informatics geared towards fourth-year radiology

residents that was funded by the Association of University Radiologists (AUR) and is co-sponsored by RSNA and the Society for Imaging Informatics in Medicine (SIIM).⁶⁹ In the US, curricular issues in radiology education are still governed by the American Board of Radiology (ABR), but educational initiatives to incorporate informatics training for all radiology trainees are likely to be in line with future developments in diagnostic imaging. Current practicing radiologists will also need to be proactive in ensuring that they are full partners in these endeavors, rather than serving as “hand-maidens” to the other investigators who “just want their images labeled”. For those individuals wishing to learn more about machine learning without having to abandon busy clinical careers for any length of time, there are several recommended online courses offered by various academic institutions (including Stanford, MIT, and Columbia) and by certain corporate entities (including Google and Nvidia) that have been available to the public at no charge.⁷⁰ Academic medical centers and other radiology organizations will need to provide environments where radiologists, machine learning experts, and other computer scientists can interact on a continual basis. As Davenport and Dreyer point out: “If the predicted improvements in deep learning image analysis are realized, then providers, patients, and payers will gravitate toward the radiologists who have figured out how to work effectively alongside AI”.⁷¹ Along those lines, we find that the creation of the ACR Data Science Institute is a strong indication that radiology organizations have recognized machine learning as a potential disruptive technology and are getting prepared to respond to this threat by investing resources to help develop, adapt, and deploy this new technology in the radiology workspace over the coming years.⁷² In addition, several radiology-based organizations have started collaborations with major technology companies to develop ML algorithms and platforms.^{73–75} We believe that this is just the beginning of a major trend in radiology, and that it behooves radiologists to participate in such endeavors for the betterment of radiology practice and the welfare of the patients that we serve.

CONCLUSION

We agree that machine learning will continue to make major advances in radiology over the next 5 to 10 years, but we completely disagree that there is any real possibility that radiologists will be replaced in that time frame, or even during the careers of our current trainees. In spite of all the advances of machine learning in the fields of self-driving cars, robotic surgery, and language translation, we believe that the work performed by radiologists is more complex than is thought by non-radiologists, and therefore more difficult to replicate by machine learning. The emergence of deep learning algorithms will help radiologists broaden the kinds of activities that establish their value in clinical care (such as routinely providing quantitative analysis), and to enhance the proportion of cognitive work (e.g. formulation of diagnosis) relative to visual search work (e.g. detection of imaging abnormalities). Imaging modalities will increasingly utilize deep learning to reduce image noise and enhance image quality overall. Since the potential for disruption of the radiology industry by machine learning does remain latent, it would be wise for various radiology organizations—especially academic institutions - to participate in research and development of this technology, and not leave the arena solely to corporate entities in the information technology sphere. While the economic environment of healthcare will continue to bring change to the practice of medicine and radiology, we believe that machine learning will not bring about the imminent doom of the radiologist. Instead, we foresee an intellectually vibrant future in which radiologists will continue to thrive professionally and benefit substantially from increasingly sophisticated and useful ML systems over the next few decades. Therefore, we would certainly encourage medical students and others interested in radiology as a profession—especially those with expertise in computer science—to pursue, enjoy, and look forward to a long career in diagnostic radiology, nuclear medicine, molecular imaging, and/or interventional radiology. This would provide benefits not only for the practitioners of diagnostic radiology, but even more importantly for our patients and for society.

REFERENCES

- Lewis-Kraus G. The great A.I. awakening. 2016. Available from: https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0.
- Muhkerjee S. A.I. versus M.D. What happens when diagnosis is automated? 2017. Available from: http://www.newyorker.com/magazine/2017/04/03/ai-versus-md?utm_term=0_e44ef774d4-947de57fec-92238377&mc_cid=947de57fec&mc_eid=6e3f8f4920&utm_content=buffera21af&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer [Published on April 3, 2017. Accessed in pre-publication on March 30, 2017].
- Christiansen CM. When new technologies cause great firms to fail. In: *The innovator's dilemma*: Boston; 1997.
- Christiansen CM, Raynor ME, McDonald R. What is disruptive innovation? *Harvard Bus Rev* 2015;.
- Chan S. Strategy development for anticipating and handling a disruptive technology. *J Am Coll Radiol* 2006; 3: 778–86. doi: <https://doi.org/10.1016/j.jacr.2006.03.014>
- Chockley K, Emanuel E. The end of radiology? Three threats to the future practice of radiology. *J Am Coll Radiol* 2016; 13(12 Pt A): 1415–20. doi: <https://doi.org/10.1016/j.jacr.2016.07.010>
- Dargan R. Not so elementary: experts debate the takeover of radiology by machines. 2017. Available from: <http://www.rsna.org/News.aspx?id=21017> [Published January 1. Accessed January 2, 2017].
- Video from RSNA 2016: AI and radiology -- separating hope from hype. 2016. Available from: <https://www.youtube.com/watch?v=4b0t7TzjZRE> [Published November 29, 2016. Accessed August 26, 2018].
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* 2015; 115: 211–. doi: <https://doi.org/10.1007/s11263-015-0816-y>
- Koch C. How the computer beat the Go master. *Scientific American*. 2016. Available from: <https://www.scientificamerican.com/article/how-the-computer-beat-the-gomaster/>. [Accessed May 5, 2018].
- Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al.

- Mastering the game of Go without human knowledge. *Nature* 2017; **550**: 354–9. doi: <https://doi.org/10.1038/nature24270>
12. Freer TW, Ulisse MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001; **220**: 781–6. doi: <https://doi.org/10.1148/radiol.2203001282>
 13. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology* 2009; **253**: 9–16. doi: <https://doi.org/10.1148/radiol.2531090611>
 14. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. *Arch Gynecol Obstet* 2009; **279**: 881–90. doi: <https://doi.org/10.1007/s00404-008-0841-y>
 15. Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? *J Am Coll Radiol* 2010; **7**: 802–5. doi: <https://doi.org/10.1016/j.jacr.2010.05.019>
 16. Fenton JJ, Abraham L, Taplin SH, Geller BM, Carney PA, D'Orsi C, et al. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst* 2011; **103**: 1152–61. doi: <https://doi.org/10.1093/jnci/djr206>
 17. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology* 2009; **253**: 17–22. doi: <https://doi.org/10.1148/radiol.2531090689>
 18. Mezrich JL, Siegel EL. Legal ramifications of computer-aided detection in mammography. *J Am Coll Radiol* 2015; **12**: 572–4. doi: <https://doi.org/10.1016/j.jacr.2014.10.025>
 19. Jackson WL, Jackson WL. *The state of CAD for mammography*. 2014. Available from: <http://www.diagnosticimaging.com/cad/state-cad-mammography>. [Published February 18. Accessed February 4, 2017].
 20. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD. High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology* 2018; **286**: 810–8. doi: <https://doi.org/10.1148/radiol.2017170549>
 21. Masarie FE, Miller RA, Myers JD. INTERNIST-I properties: representing common sense and good medical practice in a computerized medical knowledge base. *Comput Biomed Res* 1985; **18**: 458–79. doi: [https://doi.org/10.1016/0010-4809\(85\)90022-9](https://doi.org/10.1016/0010-4809(85)90022-9)
 22. Domingos P. A few useful things to know about machine learning. *Commun ACM* 2012; **55**: 78–87. doi: <https://doi.org/10.1145/2347736.2347755>
 23. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep Learning in Radiology: Does One Size Fit All? *J Am Coll Radiol* 2018; **15**(3 Pt B): 521–6. doi: <https://doi.org/10.1016/j.jacr.2017.12.027>
 24. Tanz J. *Soon we won't program computers. We'll train them like dogs*. 2016. Available from: <https://www.wired.com/2016/05/the-end-of-code> [Published May 17, 2016. Accessed August 25, 2018].
 25. Ruder S. Transfer learning – machine learning's next frontier. 2017. Available from: <http://ruder.io/transfer-learning> [Published March 21, 2017. Accessed August 25, 2018].
 26. Benjamini Y Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 1995; **57**: 289–300.
 27. NIH Clinical Center. NIH Clinical Center provides one of the largest publicly available chest x-ray datasets to scientific community. 2017/2018. Available from: <https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community> [Published September 27, 2017. Accessed October 4, 2018].
 28. The Cancer Imaging Archive (TICA). TICA Collections. 2018. Available from: <http://www.cancerimagingarchive.net/> [Accessed October 4, 2018].
 29. Kohli MD, Summers RM, Geis JR. Medical Image Data and Datasets in the Era of Machine Learning-Whitepaper from the 2016 C-MIMI Meeting Dataset Session. *J Digit Imaging* 2017; **30**: 392–9. doi: <https://doi.org/10.1007/s10278-017-9976-3>
 30. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; **284**: 574–82. doi: <https://doi.org/10.1148/radiol.2017162326>
 31. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kansagra AP. Deep learning in radiology. *Academic Radiology* 2018 in press. 2018. Available from: <https://doi.org/10.106/j.acra.2018.02.018> [Accessed August 25, 2018].
 32. Lusted LB. Signal detectability and medical decision-making. *Science* 1971; **171**: 1217–9. doi: <https://doi.org/10.1126/science.171.3977.1217>
 33. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978; **8**: 283–98. doi: [https://doi.org/10.1016/S0001-2998\(78\)80014-2](https://doi.org/10.1016/S0001-2998(78)80014-2)
 34. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Med Decis Making* 1984; **4**: 137–50. doi: <https://doi.org/10.1177/0272989X8400400203>
 35. Oakden-Rayner L. The philosophical argument for using ROC curves. 2018. Available from: <https://lukeoakdenrayner.wordpress.com/2018/01/07/the-philosophical-argument-for-using-roc-curves> [Published January 7, 2018. Accessed August 25, 2018].
 36. Keats TE, Anderson MW. Elsevier, ed. *Atlas of normal roentgen variants that may simulate disease*. 9th ed. Philadelphia, PA: Elsevier; 2013.
 37. Ding J, Bashashati A, Roth A, Oloumi A, Tse K, Zeng T, et al. Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* 2012; **28**: 167–75. doi: <https://doi.org/10.1093/bioinformatics/btr629>
 38. Knight W. *The dark secret at the heart of AI*. *MIT Technology Review*. 2018. Available from: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/> [Accessed August 26, 2018].
 39. Zhang Q, Zhu SC. Visual interpretability for deep learning. 2018. Available from: <https://arxiv.org/abs/1802.00614> [Published February 2, 2018. Accessed August 26, 2018].
 40. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1998; **20**: 1254–9. doi: <https://doi.org/10.1109/34.730558>
 41. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualizing image classification models and saliency maps. 2013. Available from: <https://arxiv.org/abs/312.6034> [Published December 20. Accessed August 26, 2018].
 42. Zhang Q, Cao F, Shi F, YN W, Zhu SC. Interpreting CNN knowledge via an explanatory graph. 2017. Available from: <https://arxiv.org/pdf/1708.01785> [Published November 22. Accessed October 4, 2018].
 43. Polikar R. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine* 2015; **6**: 21–45. doi: <https://doi.org/10.1109/MCAS.2006.1688199>
 44. Acemoglu D, Autor D. Skills, tasks and technologies: implications for employment and earnings. Working paper, National Bureau of Economic Research. 2017. Available from: <http://www.nber.org/papers/w16082> [Issued June 2010. Accessed January 2, 2017].

45. Kim C. How will AI manage your money? *Barron's – Lipper Mutual Fund Quarterly* 2018; L7–L12.
46. McGinty G, Allen Jr B, Wald C. Imaging 3.0TM. American College of Radiology. 2013. Available from: <https://www.acr.org/~media/ACR/Documents/PDF/Advocacy/IT%20Reference%20Guide/IT%20Ref%20Guide%20Imaging3.pdf> [Accessed January 2, 2017].
47. McKinsey Global Institute. A future that works: automation, employment, and productivity. January 2017. Available from: <http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works> [Accessed July 2, 2017].
48. Brynjolfson E, McAfee A. Work, progress, and prosperity in a time of brilliant technologies. In: *The second machine age*. New York: Norton; 2014. pp. 192–3.
49. Edwards B. FDA guidance on clinical decisions: Peering inside the black box of algorithmic intelligence. 2017. Available from: <https://www.chilmarkresearch.com/fda-guidance-clinical-decision-support/> [Published December 19, 2017. Accessed August 26, 2018].
50. U.S. Food and Drug Administration. Expansion of the abbreviated 510(k) program: demonstrating substantial equivalence through performance criteria. Draft guidance for industry and Food and Drug Administration. 2018. Available from: <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM604195.pdf> [Published April 12, 2018. Accessed August 26, 2018].
51. U.S. Food Drug Administration. *In Brief: FDA to offer a voluntary, more modern 510(k) pathway for enabling moderate risk devices to more efficiently demonstrate safety and effectiveness*. 2018. Available from: <https://www.fda.gov/NewsEvents/Newsroom/FDAInBrief/ucm604348.htm> [Published April 12, 2018. Accessed August 26, 2018].
52. News FDA. Arterys wins FDA clearance for oncology imaging software. 2018. Available from: <https://www.fdanews.com/articles/185673-arterys-wins-fda-clearance-for-oncology-imaging-software> [Published February 20, 2018. Accessed August 26, 2018].
53. U.S. Food and Drug Administration. FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients. 2018. Available from: <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm596575.htm> [Published February 13, 2018. Accessed August 26, 2018].
54. U.S. Food and Drug Administration. FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures. 2018. Available from: <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm608833.htm> [Accessed August 26].
55. U.S. Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. 2018. Available from: <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm> [Published April 11, 2018. Accessed August 26, 2018].
56. Enzmann DR, Feinberg DT. The nature of change. *J Am Coll Radiol* 2014; **11**: 464–70. doi: <https://doi.org/10.1016/j.jacr.2013.12.006>
57. Domingos P. *The master algorithm; how the quest for the ultimate learning machine will remake our world*. New York, NY: Basic; 2015. pp. 25–6.
58. Grove AS. *Only the paranoid survive; how to exploit the crisis points that challenge every company*. New York, NY: Currency; 1996.
59. Bloch A. *Murphy's Law and other reasons why things go wrong*. Los Angeles. **41**. CA: Price/Stern/Sloan; 1979.
60. Henson J. The Paperwork Explosion. The Jim Henson Company (for IBM). 1967/20102018. Available from: www.youtube.com/watch?v=_JZw2CoYztk [Produced October 1967. Published March 30, 2010. Accessed August 26, 2018].
61. Dreyer K. When machines think: radiology's next frontier. In: *Presented at the 2016 annual meeting of the Radiological Society of North America*. Chicago, Ill; 2016.
62. Thompson C. In: Smarter than you think; how technology is changing our minds for the better. In: *The rise of the centaurs*. NY: Penguin; 2013. pp. 1–5.
63. Kelly K. *The inevitable; understanding the 12 technological forces that will shape our future*. New York, NY: Viking; 2016.
64. American Medical Association. AMA Passes First Policy Recommendations on Augmented Intelligence. 2018. Available from: <https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence> [Published June 14, 2018. Accessed August 26, 2018].
65. Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. *JAMA* 2018; **319**: 19–20. doi: <https://doi.org/10.1001/jama.2017.19198>
66. Gunderman R, Chan S. Knowledge sharing in radiology. *Radiology* 2003; **229**: 314–7. doi: <https://doi.org/10.1148/radiol.2292030030>
67. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 2017; **208**: 754–60. doi: <https://doi.org/10.2214/AJR.16.17224>
68. Walter S. Private communication: unpublished observation on trends in RSNA R&E Foundation grants applications 2014–2018. February 2018
69. Cook T. <https://sites.google.com/view/imaging-informatics-course/resources>. 2017. Available from: <https://sites.google.com/view/imaging-informatics-course/resources> [Start date: October 2, 2017. Accessed August 26, 2018].
70. Marr B. The 6 best free online artificial intelligence courses for 2018. 2018. Available from: <https://www.forbes.com/sites/bernardmarr/2018/04/16/the-6-best-free-online-artificial-intelligence-courses-for-2018/#6836668959d7> [Published April 16, 2018. Accessed August 26, 2018].
71. Davenport TH, Dreyer KJ. AI will change radiology, but it won't replace radiologists. 2018. Available from: <https://hbr.org/2018/03/ai-will-change-radiology-but-it-wont-replace-radiologists> [Rev March 27, 2018. Accessed April 19, 2018].
72. American College of Radiology (ACR) Data Science Institute. 2018. Available from: <http://www.acrdsi.org/index.html> [Accessed May 5, 2018].
73. Healthcare P. Partners HealthCare and GE Healthcare launch 10-year collaboration to integrate Artificial Intelligence into every aspect of the patient journey. 2017. Available from: <https://www.partners.org/Newsroom/Press-Releases/Partners-GE-Healthcare-Collaboration.aspx> [Published May 17, 2017. Accessed August 26, 2018].
74. Terdiman T. Facebook and NYU believe AI can make MRIs way, way faster. 2018. Available from: <https://www.fastcompany.com/90220180/facebook-and-nyu-believe-ai-can-make-mris-way-way-faster> [Published August 20, 2018. Accessed August 26, 2018].
75. Partners R. Radiology Partners to join the IBM Watson Health Medical Imaging Collaborative. 2017. Available from: <http://www.radpartners.com/events/j4ss9xuz22/Radiology-Partners-to-join-the-IBM-Watson-Health-Medical-Imaging-Collaborative> [Accessed August 26, 2018].



Radiomics in prostate cancer: basic concepts and current state-of-the-art

Shan Yao¹ · Hanyu Jiang¹ · Bin Song¹

Received: 1 September 2019 / Revised: 17 October 2019 / Accepted: 30 October 2019
© The Author(s) 2019

Abstract

Prostate cancer (PCa) is the second most common type of cancer among males and the fifth major contributor to cancer-related mortality and morbidity worldwide. Radiomics, as a superior method of mining big data in medical imaging, has enormous potential to assess PCa from diagnosis to prognosis to treatment response, empowering clinical medical strategies accurately, reliably, and effectively. Hence, this article reviews the basic concepts of radiomics and its current state-of-the-art in PCa as well as put forwards the prospects of future directions.

Keywords Radiomics · Prostate cancer · Multi-parametric magnetic resonance imaging · Artificial Intelligence

Introduction

Prostate cancer (PCa) is the second most common type of cancer among males and the fifth major contributor to cancer-related mortality and morbidity worldwide [1–3]. However, accurate identification and effective treatment of PCa remain a major public health challenge, largely due to its substantial heterogeneity which often leads to imprecise diagnosis and suboptimal disease management.

Digital rectal examination (DRE), prostate-specific antigen (PSA) test, and transrectal ultrasound (TRUS)-guided prostate biopsy are currently the most widely used diagnostic methods of PCa in clinical practices. However, each of these methods has some limitations [4], including different suitable conditions, unstable accuracy, sampling error, over-diagnosis, etc. The current paradigm for screening and diagnosis is imperfect, with relatively low specificity, high cost, and high morbidity. Meanwhile, the optimal clinical management which may include watchful waiting, active surveillance, open, laparoscopic or robotic-assisted radical prostatectomy, external beam radiation therapy (EBRT), and brachytherapy [5], is highly dependent on accurate diagnosis. Early detection of PCa enables radical treatment and

long-term patient survival. Nevertheless, once the tumor infiltrates out of the prostate capsule, the treatment effect and prognosis are often poor.

With the rapid development of medical imaging techniques, many imaging modalities have demonstrated great value in the screening, diagnosis, treatment response measurement, and prognosis evaluation of PCa. Magnetic resonance imaging (MRI) could provide the advantage of detecting prostate and periprostatic characterization and structures with high spatial resolution, superior contrast resolution in soft tissue, multiplanar imaging capabilities, and larger field of view (FOV) [6, 7]. Multi-parametric magnetic resonance imaging (mpMRI) has shown promise to improve detection and characterization of PCa considerably with more seminal information combining structure and function, which plays an extremely crucial role in tumor detection and localization, staging, aggressiveness assessment, treatment option assistant, and patient follow-up of PCa [8–10]. Besides, to standardize the use of mpMRI, the Prostate Imaging Reporting and Data System (PI-RADS) was presented by the European Society of Urogenital Radiology (ESUR) in 2013 [8] and an updated version (PI-RADS v2) in 2015 [11] which has been keeping updating and supplementing up to now [12, 13]. Nevertheless, there are also some limitations, such as invasive and with biopsy errors of MR-directed biopsy (MRDB), the lack of consistency and nonquantitative nature of dynamic contrast-enhancement-MRI (DCE-MRI), not providing recommendation regarding the best threshold, unavailable 3D tumor volume delineation, and a large degree

✉ Bin Song
songlab_radiology@163.com

¹ Department of Radiology, West China Hospital, Sichuan University, No. 37 Guo Xue Alley, Wu Hou District, Chengdu 610041, China

of subjectivity related to imaging quality, radiologists, and urologist with PI-RADS [14–18].

Compared with traditional medical imaging, radiomics has the strong ability of extracting more critical and comprehensive information of lesions with high throughput by quantitative methods [19–21]. It enables automatic localization and characterization of PCa as well as identifies the great value of grading and staging, therapeutic evaluation, prognostic analysis, and even genomics, helping a lot in clinical diagnosis and treatment decisions. Hence, this article reviews the basic concepts of radiomics and its current state-of-the-art in PCa.

Basic concepts of radiomics in PCa

Definition

“Radiomics” was first mentioned by Gillies et al. [22] in 2010 to describe the extraction of quantitative features from image images. In 2012, Lambin et al. [19] formally put forward the definition of “Radiomics” for the first time, as analyzing medical image data quantitatively that extracting a large number of features from medical images with high throughput and then transforming them into high resolution and deep-going mineable database with automatic or semi-automatic software. In the same year, Kumar et al. [23] expanded the definition of radiomics to extraction and analysis of a large number of advanced and quantitative image features from medical imaging such as computed tomography (CT), positron emission tomography (PET), or MRI with high throughput.

Process of radiomics

Radiomics is a multi-disciplinary technology, of which the core steps include data acquisition, features selection, model building, and analysis, aiming at converting routine clinical images into mineable data, with high fidelity and high throughput.

The process of radiomics generally consists of several closely related steps as followed:

1. acquiring high-quality standardized imaging data and reconstruction;
2. segmentation of the region of interest (ROI) or the volume of interest (VOI) manually or automatically with computer-assisted contouring;
3. high-throughput features extraction and quantification;
4. feature selection and construction of clinical prediction models;
5. validation of the models and establishment of shared databases [19, 23, 24].

Image acquisition and reconstruction

Acquisition of high-quality images is the basis of radiomics, thus, it is pivotal to standardize the process of data acquisition and reconstruction. Those imaging data are obtained with CT, MRI, PET/CT, or PET/MRI. CT is mainly used to evaluate the density, shape, and texture characteristics of lesions due to its high spatial resolution, while it is not recommended for PCa because of without characteristic manifestation. MRI, especially mpMRI is widely used for the analysis of PCa lesions because of its better soft-tissue resolution and comprehensive information. Functional MRI such as diffusion weighted imaging (DWI) and DCE-MRI extracts more image features about cell structure and microvascular perfusion, meanwhile, tissue metabolism information can be provided by PET/CT or PET/MRI [10, 25–27].

However, the robustness could be affected by many factors, such as pulse sequence, FOV, slice thickness et al. of PCa-widely-used MRI. The reproducibility and repeatability of image data characters rely heavily on standardized image acquisition protocols. In addition, calibration of imaging settings is crucial as images acquired at different imaging settings may have poor repeatability [28]. Therefore, great efforts have been made by many international organizations, such as Radiological Society of North America, the Society of Nuclear Medicine and Molecular Imaging, the International Society of Magnetic Resonance in Medicine, and the World Molecular Imaging Society [24] to define the acquisition and reconstruction standards for radiomics.

Segmentation

Image segmentation, referred as delineation of the target area (such as tumor), is the premise of data extraction to ensure that the follow-up work goes on well. There are generally three ways of segmentation: manual, semi-automatic, and automatic, of which the former two are mostly used at present. Among these methods, manual segmentation has the advantage of high accuracy, especially for most tumors with clear boundaries but irregular shape. However, manual segmentation is time-consuming with low efficiency and inter-operator variability. For PCa tumors with blurred margins, the heterogeneity in locating the tumor boundaries by different radiologists can cause limited data repeatability. Automatic or semi-automatic segmentation, on the contrary, can reduce this heterogeneity. Nevertheless, they are not precise enough in some confusing components with limited interpretability of models that need further improvement. There are many algorithms developed for segmentation, such as region-growing method [29], graph-cuts algorithm, atlas-based segmentation [30], volumetric CT-based segmentation [31], semi-automatic segmentation [32], active contours algorithm [33], live-wire-based segmentation [34],

etc. Currently, several software packages are available for segmentation, including ITK-SNAP (www.itksnap.org), 3DSlicer (www.slicer.org), MIM (www.mimsoftware.com) and ImageJ (<https://imagej.nih.gov/ij/>), etc. Automatic segmentation will be encouraged strongly in the future while requires large data sets for training.

Feature extraction and quantification

Extraction and quantification of the imaging features which could characterize the attributes of the target area are the heart of radiomics. There are two types of features extracted in radiomics: “semantic” and “agnostic” features [24]. The former “semantic” is used to describe qualitative morphological features such as size, shape, location, vascularity, speculation, necrosis, and attachments or lepidics. The latter “agnostic” refers to invisibly quantitative description of heterogeneity of lesions such as textures, histogram, wavelets, Laplacian transforms, Minkowski functionals, and fractal dimensions. Textures can be obtained through first-, second-, and high-order statistical methods generally. The first-order features based on histogram mainly include maximum, minimum, average, standard deviation, variance, energy, entropy, sharpness, skewness, and kurtosis, gray-scale, which acquire relevant statistical information by frequency distribution of different gray levels in ROI. Second-order texture feature algorithms include gray-level co-occurrence matrix (GLCM) [35] and gray-level run-length matrix (GLRLM) [36]. High-order algorithms customarily make use of neighborhood gray-tone difference matrix (NGTDM) [37] and gray-level size zone matrix (GLSZM) [38]. As for methods based on models or transformation, Laplacian transforms are often utilized in image preprocessing and wavelet transform is in extracting texture features from sub-images to mine information more deeply. Similarly, a lot of software packages have been put into features extraction such as IBEX [39], MaZda [40], Pyradiomics [41], CERR [42], ePAD [43], LifeX [44], and some other R-based or MATLAB-based programs. Cooperative use of different software may help to acquire more comprehensive radiomics features.

Feature selection and construction of clinical prediction models

To avoid some algorithms failure caused by high dimensionality of feature space, reduce over-fitting, improve the model stability, and shorten the training time, feature selection will be carried out before modeling. Fisher’s discriminant ratio, mutual information feature selection (MIFS), maximal relevance and minimum redundancy (mRMR), principal component analysis (PCA), consensus clustering (CC), locally linear embedding (LLE), etc., are common feature-selecting methods [45–47]. Database and model

construction are a breakthrough point of radiomics analysis that could be applied as a powerful assistant tool for diagnosis and treatment effect prediction. After that, the classifier or prediction model is usually built with machine learning algorithms, which mainly known as Support Vector Machine (SVM) [47, 48], Logistic Regression [49], Random Forest (RF), Decision Tree (DT), clustering analysis, etc. Besides, Convolutional Neural Network (CNN), Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Holistically Nested Network (HNN) [50, 51], etc., which belong to rapid-developing deep learning, really accelerated the pace of radiomics progress. The establishment of database and modeling is a complex and challenging process, which is necessary to strengthen the cooperation of multi-disciplinary and multi-team especially medical science and engineering, so as to standardize management and make efficient use of images feature data, as well as to build stable and accurate models.

Data sharing and mining

Radiomics is a bigdata analysis method, inevitably, whose results may be affected by some relevant factors such as the single source of research objects, different imaging equipment and parameters, complexity of image segmentation and feature extraction, etc. Thus, validation in multiple centers is quite of necessary, so as to improve the stability and representativeness of data. Though it is really hard to work radiomics done, we need to capture valuable data and share them across institutions to accumulate sufficient numbers for statistical power, as the QIN [52] proposing. Also, it is quietly important to make great efforts to mine data more deeply.

The current state-of-the-art of radiomics in PCa

In PCa, radiomics has been intensively applied to tumor detection, localization, staging, aggressiveness assessment, treatment decision-making assistant, and patient follow-up.

Detection and diagnosis

Accurate tumor diagnosis and staging is the cornerstone of proper patient management. Cameron et al. [53] proposed a quantitative comprehensive feature model called MAPS based on radiomics for automatic detection of PCa and achieved an accuracy (ACC) of 87%. Furthermore, Khalvati et al. [54] designed a new automatic mpMRI texture feature models incorporating computed high-b (CHB-DWI) and correlated diffusion imaging (CDI). It helped to improve radiomics-driven detection of PCa significantly compared to conventional mpMRI models. And the ACC and area under the curve (AUC) of the receiver-operating

characteristic (ROC) of the full modalities model reached 0.82, 0.86 and 0.88, 0.88 using sensitivity and specificity, respectively, as performance criteria. Another study by Wibmer et al. [55] using MRI in 147 patients with PCa confirmed by biopsy showed that Haralick texture features derived from T2-weighted images and apparent diffusion coefficient (ADC) maps had the potential to differentiate PCa and non-cancerous prostate tissue. In the discrimination between clinically significant PCa (csPCa) and clinically insignificant PCa (ciPCa), Min et al. [56] demonstrated that mpMRI-based radiomics signature had the potential to noninvasively work it done using a cross-validation of a machine learning method, which may help clinicians to facilitate prebiopsy and pre-treatment risk stratification (AUC, sensitivity, and specificity are 0.823, 0.841, and 0.727, respectively). Furthermore, more useful parameters with good performance are being excavated. For instance, Cuocolo et al. [57] thought that the surface area-to-volume ratio (SAVR) derived from ADC maps was recognized as the most promising tool in the discrimination of csPca from non-csPca, outperforming other shape features even such as lesion volume and maximum diameter (AUC = 0.78). As for identifying lesions in transition zone (TZ) and peripheral zone (PZ), Ginsburg et al. [58] suggested that a zone-aware classifier C^{PZ} significantly improved the accuracy of cancer detection in the PZ, with the AUC of 0.71.

There are also PI-RADS related studies pointing out that MR radiomics could help to improve the performance of PI-RADS v2 in clinically relevant PCa [59], with the aid of which the sensitivity significantly increased (79–94.4% in PZ PCa, 73.4–91.6% in TZ PCa). Though the samples were small (< 100). Similarly, Chen et al. [60] compared radiomics-based analysis with PI-RADS v2, which indicated that T2 W- and ADC-based radiomics models showed high diagnostic efficacy in distinguishing PCa vs. non-PCa at a high ACC of 0.991, as well as in high-grade vs. low-grade (ACC 0.867). Those are complementary to the refinement of specific standards and optimization model both each other.

Aggressiveness evaluation and staging

As the gold standard for PCa aggressiveness assessment [61], Gleason grading system plays an important role in the stratification of risk for PCa. Radiomics-combined patterns can impact clinical outcomes, treatment selection, and the determination of disease status noninvasively. In this aspect, Wibmer et al. [55] reported that entropy derived from the ADC map is significantly associated with PCa Gleason score (GS) in PZ, independently from the median ADC value ($P < 0.05$). Nketiah et al. [62] worked on distinguishing GS3 + 4 from GS4 + 3 PCa with several T2 W MRI-derived textural features and MRI parameters, among which angular second moment (ASM) and entropy produced the

best results (AUC = 0.83, both). As the first study that had implemented cross-modality intensity statistics for identifying radiomic features associated with GS, Chaddad et al. [63] presented a novel type of radiomic analysis model based on joint intensity matrices (JIMs), then evaluated its ability of predicting the GS in PCa patients, and compared it with GLCM. Final results showed that JIMs, which were suggested as a complementary biomarker to predict PCa GS, described the heterogeneity across mpMRI images better than GLCM (AUC of 78.37% vs 68.62% for GS ≤ 6, 80.54% vs 71.09% for GS3 + 4, and 62.65% vs 60.39% for GS ≥ 4+3, respectively). Then, they tested and confirmed the hypothesis that radiomic features extracted from mpMRI could predict the GS of patients with PCa in the same year [64]. Their research provided a reference for guiding the treatment planning of PCa, and also enlightened a new way for our future studies that multi-classification method can be applied to extract and analyze new multi-modal features.

Treatment evaluation and prognosis analysis

The management of advanced PCa has changed substantially with the availability of multiple effective novel treatments, which has led to improved disease survival. The imaging more precise, the earlier detection of metastatic disease and identification of oligometastatic disease more accurate are, so as to optimal assessment of treatment response. In prostate focal therapy, it is of great importance to localize malignant lesions accurately to increase biological effect of the tumor region while achieving a reduction in dose to non-cancerous tissue. Thus, a radiomics-based radiotherapy planning framework had been presented by Shiradkar et al. to generate targeted focal treatment plans [65]. It could boost dose to the cancerous lesions whilst minimize damage to the surrounding structures for brachytherapy and EBRT, as well as reduce treatment related side effects. Walsh et al. [66] provided a ‘proof-of-concept’ methodology enabling the determination of a threshold 5% that would most likely benefit from proton therapy prospectively. It justified the selection of proton-EBRT (P-EBRT) or photon-EBRT (X-EBRT) for PCa patients in a clinical decision support system (CDSS). For monitoring treatment changes, radiomics also plays a unique role. Abdollahi et al. compared radiomics features between pre- and post-radiotherapy and final results told that radiomics was being potentially useful imaging biomarkers for predicting the complications and structural changes in the bladder wall of PCa after RT (the highest AUC_{mean} 0.68, of pre-IMRT T2W radiomics). Feature changes had a good correlation with radiation dose and radiation-induced urinary toxicity [67, 68].

In addition, besides the lesion itself, the PCa-associated diseases with high risk and bad prognosis should not

be underestimated. A model combining texture analysis (TA) and machine learning for predicting the presence of histopathological extraprostatic extension (EPE) in PCa was suggested by Stanzone et al. [69], of which classifier Bayesian network (BN) showed high diagnosis ACC (82.3%). Besides, extracapsular extension (ECE) may affect clinical decisions and prognosis, which needs to be predicting to help on surgical planning and reduce the risk. Ma et al. had proved the value of radiomics in preoperative prediction of ECE with a high ACC at 83.58% better than radiologists, and demonstrated the radiomics signature yielded a good performances for discrimination, calibration, and clinical usefulness [70, 71].

Radiogenomics

Radiogenomics is an encouraging field that combines genomics and medical imaging techniques, considered as a bridge connecting radiomics with genomics [72], while some challenges still need to be addressed. At present, the application of this technique in PCa is relatively less extensive and in-depth than that in other organs tumor such as brain, lung, or liver [72, 73]. Since PCa clinical results are closely related to phosphatase and tensin homolog (PTEN), loss of which is associated with increased clinical aggressive phenotype and mortality, related studies are giving out valuable potential. For example, McCann et al. [74] investigated the association of mpMRI features and PZ PCa, as a result of weak but significant negative correlation between GS and PTEN expression ($r = -0.30$, $p = 0.04$) and between k_{ep} and PTEN expression ($r = -0.35$, $p = 0.02$). Similarly, Switlyk et al. [75] explored the relationship between clinicopathologic and mpMRI features in 43 PCa patients underwent radical prostatectomy. They found that low PTEN expression significantly corresponded to low ADC value in PCa, whilst PTEN expression was negatively associated with lymph-node metastasis (bead arrays, $p = 0.008$; RT-qPCR, $p < 0.001$). On the other side, Stoyanova et al. [76, 77] adopted a unique approach and performed radiogenomic analysis on PCa patients underwent MR-guided biopsies. Radiomics features associated with prognostic biomarkers were first identified in that approach, allowing a more accurate radiomic–biological correlation significantly (≥ 0.9 in TRPM8, DPP4, and GCNT1). While the samples were small (6 patients, 17 biopsy samples), further large-scale repeatable research is needed. As a relatively new imaging branch, radiogenomics is evolving and expected to play an important role in the clinical treatment of PCa, with an ultimate goal to predict prognosis and treatment response.

Habitat-based radiomics

Habitat imaging has enormous utility to get insights of tumor phenotype and microenvironment quantitatively [78, 79]. And as we know, intratumoral heterogeneity has long been a tricky obstacle in the diagnosis and management. For answering that, habitat-based radiomics was born at the right moment. Defining sub-regions and extracting habitat-based features will be added into the conventional process.

In 2018, Parra et al. [80] took use of perfusion curve patterns defined by DCE of mpMRI to identify the habitat of PCa. They evaluated both DCE and ADC features and affirmed the DCE features' value for discriminating csPCa and ciPCa (with AUC of 0.82). Then, in the next year, they investigated prostate habitats by analyzing seven quantitative DCE features based on the late area under the DCE time-activity curve (H-AUCf) [81], which was found of great value for predicting the csPCa (with best AUC of 0.82, 95% confidence interval (CI) [0.81–0.83]). Habitat-based radiomics may be a hot trend, though there is very little research on PCa now. Thus, well-designed prospective studies with high-quality data are required to strengthen it in future work.

Deep learning

Deep learning, as the best promising method for radiomics, has been putting a step forward in radiomics. For instance, several studies focusing on PCa segmentation relying on deep learning have shown promising results recently. Actually, Liao et al. [82] have attempted for automatic MRI prostate segmentation by deep learning framework in 2013. In 2017, Cheng et al. [51] achieved automated MRI prostate segmentation using HNN and fivefold cross-validation, with Dice similarity coefficient (DSC) of $(89.77\% \pm 3.29\%)$ and a mean Jaccard similarity coefficient (IoU) of $(81.59\% \pm 5.18\%)$. In 2019, Zhu et al. [83] proposed a boundary-weighted domain adaptive neural network (BOWDA-Net), which overcame the complexity between prostate and other structures and helped to segment prostate more accurate and sensitive (with high DSC of 89.67% and overperformed other methods, $p < 0.05$). However, it was limited as it worked on small data sets. Otherwise, to improve the performance in PCa diagnosis and treatment planning, Alkadi et al. [84] proposed a deep encoder–decoder CNN for detection and localization of PCa in T2WI images with gratifying results (average AUC of 0.995, ACC of 0.894, and recall of 0.928). Song et al. [85] also proposed deep CNN but in mpMRI for PCa diagnosis and prediction, with AUC of 0.944 (95% CI 0.876–0.994). However, mono-modality system was not as superior as multi-modality in model performance and generalization, which require larger data sets to validate in. At the same time, artificial intelligence (AI) provides

benefits at the expense of a high false-positive rate [86, 87] that needs to be under consideration and optimized.

Future directions, development, and potential issues

The application of AI in PCa is supposed to meet the clinical demands closely and transformation of radiomics into the clinic may require a more comprehensive understanding of the underlying morphologic tissue characteristics they reflect.

As heterogeneity is a well-known chasm of PCa, persistent action should be taken to reduce the impact of heterogeneity, as well as improve the accuracy and objectivity in the further work. In addition, its multifocal nature prompts us to concentrate on PZ, TZ, surrounding tissue, and tumor microenvironment. Additionally, a minority of the prior studies focus on radiomics-guided treatment, which needs to be supported in further work.

Moreover, images' differences can be tough due to the lack of uniform standard in scanning parameters and reconstruction algorithms for imaging equipment. Even in the same equipment, differences in contrast agent, scanning thickness, convolution kernel, and even coils (body or endorectal), etc., will have potential influences on data analysis. Most of the existing studies are small sample exploration in a single institution, of which conclusions are short of extensive validation. Therefore, radiomics on PCa must be repeatedly refined and externally validated in multi-center, large-sample, randomized-controlled clinical trials, which can better interpret the complexity of PCa, by the way, meet the requirements of precision medicine. Perhaps, it is a good choice to unify standards, share data, or open source.

In addition, the application of AI in PCa should not be limited to simple computer-aided diagnosis (CADx) or machine learning, but deep learning to assist the completion of large data analysis to create more value and more radiologists should be involved in the sustainable development task of AI. However, information security and privacy and the ethical issues of AI may pose a barrier when mining data depth by depth.

At present, radiomics alone is facing a number of great challenges. For the foreseeable future, the multi-dimensional and multi-model radiomics combined with clinical and laboratory information and other omics has become the next trend of AI-driven medicine. And that is exactly what the modern imaging rapidly evolving and expanding aiming at.

Conclusion

In conclusion, radiomics has the potential to become a useful assistant tool in clinical oncology imaging, providing important information with the characters, prognosis, treatment prediction, and response of tumors in PCa. However, the potential value of radiomics in PCa has not been fully investigated. In the face of great opportunities and challenges, we need to spare no efforts to expand it and derive more clinically meaningful trends, as well as to meet the developing needs of precision medicine and enhance precision medicine initiatives.

Acknowledgements This manuscript was supported by Research Grant of National Natural Science Foundation of China, No. 81771797.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Torre LA, Freddie B, Siegel RL, Jacques F, Joannie LT, Ahmedin J. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87–108. <https://doi.org/10.3322/caac.21262>.
2. McGuire, S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Adv Nutr Int Rev J*. 2016;7(2):418–9. <https://doi.org/10.3945/an.116.012211>.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;60(5):277–300. <https://doi.org/10.3322/caac.21442>.
4. Graham J, Kirkbride P, Cann K, Hasler E, Prettyjohns M. Prostate cancer: summary of updated NICE guidance. *BMJ (Clin Res Ed)*. 2014;348:f7524. <https://doi.org/10.1136/bmj.f7524>.
5. Horwich A, Parker C, Reijke TD, Kataja V. Prostate cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol Off J Eur Soc Med Oncol*. 2013;24(Suppl 6):vi106. <https://doi.org/10.1093/annonc/mdt208>.
6. Woodrum DA, Kawashima A, Gorny KR, Mynderse LA. Prostate cancer: state of the art imaging and focal treatment. *Clin Radiol*. 2017;72(8):665–79. <https://doi.org/10.1016/j.crad.2017.02.010>.
7. Mendhiratta N, Taneja SS, Rosenkrantz AB. The role of MRI in prostate cancer diagnosis and management. *Future Oncol (Lond Engl)*. 2016;12(21):2431–43. <https://doi.org/10.2217/fon-2016-0169>.
8. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, et al. ESUR prostate MR guidelines 2012. *Eur Radiol*. 2012;22(4):746–57. <https://doi.org/10.1007/s00330-011-2377-y>.

9. Sciarra A, Barentsz J, Bjartell A, Eastham J, Hricak H, Panebianco V, et al. Advances in magnetic resonance imaging: how they are changing the management of prostate cancer. *Eur Urol*. 2011;59(6):962–77. <https://doi.org/10.1016/j.eururo.2011.02.034>.
10. Johnson LM, Turkbey B, Figg WD, Choyke PL. Multiparametric MRI in prostate cancer management. *Nat Rev Clin Oncol*. 2014;11(6):346–53. <https://doi.org/10.1038/nrclinonc.2014.69>.
11. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS prostate imaging—reporting and data system: 2015, version 2. *Eur Urol*. 2016;69(1):16–40. <https://doi.org/10.1016/j.eururo.2015.08.052>.
12. Barentsz J, de Rooij M, Villeirs G, Weinreb J. Prostate imaging-reporting and data system version 2 and the implementation of high-quality prostate magnetic resonance imaging. *Eur Urol*. 2017;72(2):189–91. <https://doi.org/10.1016/j.eururo.2017.02.030>.
13. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur Urol*. 2019;76(3):340–51. <https://doi.org/10.1016/j.eururo.2019.02.033>.
14. Muthigi A, George AK, Sidana A, Kongnyuy M, Simon R, Moreno V, et al. Missing the mark: prostate cancer upgrading by systematic biopsy over magnetic resonance imaging/transrectal ultrasound fusion biopsy. *J Urol*. 2017;197(2):327–34. <https://doi.org/10.1016/j.juro.2016.08.097>.
15. Giganti F, Moore CM. A critical comparison of techniques for MRI-targeted biopsy of the prostate. *Transl Androl Urol*. 2017;6(3):432–43. <https://doi.org/10.21037/tau.2017.03.77>.
16. Hamoen EHJ, de Rooij M, Witjes JA, Barentsz JO, Rovers MM. Use of the prostate imaging reporting and data system (PI-RADS) for prostate cancer detection with multiparametric magnetic resonance imaging: a diagnostic meta-analysis. *Eur Urol*. 2015;67(6):1112–21. <https://doi.org/10.1016/j.eururo.2014.10.033>.
17. Padhani AR, Weinreb J, Rosenkrantz AB, Villeirs G, Turkbey B, Barentsz J. Prostate imaging-reporting and data system steering committee: PI-RADS v2 status update and future directions. *Eur Urol*. 2019;75(3):385–96. <https://doi.org/10.1016/j.eururo.2018.05.035>.
18. Padhani AR, Haider MA, Villers A, Barentsz JO. Multiparametric magnetic resonance imaging for prostate cancer detection: what we see and what we miss. *Eur Urol*. 2019;75(5):721–2. <https://doi.org/10.1016/j.eururo.2018.12.004>.
19. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer (Oxf Engl 1990)*. 2012;48(4):441–6. <https://doi.org/10.1016/j.ejca.2011.11.036>.
20. Patel N, Henry A, Scarsbrook A. The value of MR textural analysis in prostate cancer. *Clin Radiol*. 2018. <https://doi.org/10.1016/j.crad.2018.11.007>.
21. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. *Cancer Imaging Off Publ Int Cancer Imaging Soc*. 2013;13:140–9. <https://doi.org/10.1102/1470-7330.2013.0015>.
22. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin Radiol*. 2010;65(7):517–21. <https://doi.org/10.1016/j.crad.2010.04.005>.
23. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30(9):1234–48. <https://doi.org/10.1016/j.mri.2012.06.010>.
24. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–77. <https://doi.org/10.1148/radiol.2015151169>.
25. Umbehr MH, Muntener M, Hany T, Sulser T, Bachmann LM. The role of 11C-choline and 18F-fluorocholine positron emission tomography (PET) and PET/CT in prostate cancer: a systematic review and meta-analysis. *Eur Urol*. 2013;64(1):106–17. <https://doi.org/10.1016/j.eururo.2013.04.019>.
26. Schick U, Lucia F, Dissaux G, Visvikis D, Badic B, Masson I, et al. MRI-derived radiomics: methodology and clinical applications in the field of pelvic oncology. *Br J Radiol*. 2019;92:20190105. <https://doi.org/10.1259/bjr.20190105>.
27. Liu B, Cheng J, Guo DJ, He XJ, Luo YD, Zeng Y, et al. Prediction of prostate cancer aggressiveness with a combination of radiomics and machine learning-based analysis of dynamic contrast-enhanced MRI. *Clin Radiol*. 2019. <https://doi.org/10.1016/j.crad.2019.07.011>.
28. Larue RT, Defraene G, De Ruyscher D, Lambin P, van Elmpot W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br J Radiol*. 2017;90(1070):20160665. <https://doi.org/10.1259/bjr.20160665>.
29. Ma J, Son JB, Hazle JD. An improved region growing algorithm for phase correction in MRI. *Magn Reson Med*. 2016;76(2):519–29. <https://doi.org/10.1002/mrm.25892>.
30. Kwak K, Yoon U, Lee DK, Kim GH, Seo SW, Na DL, et al. Fully-automated approach to hippocampus segmentation using a graph-cuts algorithm combined with atlas-based segmentation and morphological opening. *Magn Reson Imaging*. 2013;31(7):1190–6. <https://doi.org/10.1016/j.mri.2013.04.008>.
31. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep*. 2013;3:3529. <https://doi.org/10.1038/srep03529>.
32. Scholten ET, Jacobs C, van Ginneken B, van Riel S, Vliegenthart R, Oudkerk M, et al. Detection and quantification of the solid component in pulmonary subsolid nodules by semiautomatic segmentation. *Eur Radiol*. 2015;25(2):488–96. <https://doi.org/10.1007/s00330-014-3427-z>.
33. Farmaki C, Marias K, Sakkalis V, Graf N. Spatially adaptive active contours: a semi-automatic tumor segmentation framework. *Int J Comput Assist Radiol Surg*. 2010;5(4):369–84. <https://doi.org/10.1007/s11548-010-0477-9>.
34. Farber M, Ehrhardt J, Handels H. Live-wire-based segmentation using similarities between corresponding image structures. *Comput Med Imaging Graph Off J Comput Med Imaging Soc*. 2007;31(7):549–60. <https://doi.org/10.1016/j.compmedimag.2007.06.005>.
35. Haralick RM, Shanmugam K, Dinstein IH. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern*. 1973;3(6):610–21.
36. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process*. 1975;4(2):172–9.
37. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern*. 1989;19(5):1264–74.
38. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Shape and texture indexes application to cell nuclei classification. *Int J Pattern Recognit Artif Intell*. 2013;27(01):1545.
39. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys*. 2015;42(3):1341–53. <https://doi.org/10.1118/1.4908210>.
40. Szczypinski PM, Strzelecki M, Materka A, Klepaczko A. MaZda—a software package for image texture analysis. *Comput Methods Progr Biomed*. 2009;94(1):66–76. <https://doi.org/10.1016/j.cmpb.2008.08.005>.
41. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Can Res*. 2017;77(21):e104–7. <https://doi.org/10.1158/0008-5472.can-17-0339>.

42. Apte AP, Iyer A, Crispin-Ortuzar M, Pandya R, van Dijk LV, Spezi E, et al. Technical note: extension of CERR for computational radiomics: A comprehensive MATLAB platform for reproducible radiomics research. *Med Phys*. 2018. <https://doi.org/10.1002/mp.13046>.
43. Rubin DL, Ugur Akdogan M, Altindag C, Alkim E. ePAD: an image annotation and analysis platform for quantitative imaging. *Tomography (Ann Arbor, Mich)*. 2019;5(1):170–83. <https://doi.org/10.18383/j.tom.2018.00055>.
44. Nioche C, Orlhac F, Boughdad S, Reuze S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res*. 2018;78(16):4786–9. <https://doi.org/10.1158/0008-5472.can-18-0125>.
45. Doyle S, Hwang M, Shah K, Madabhushi A, Feldman M, Tomaszewski J, editors. Automated grading of prostate cancer using architectural and textural image features. In: *IEEE international symposium on biomedical imaging: from nano to macro*; 2007.
46. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. 2015;5:13087. <https://doi.org/10.1038/srep13087>.
47. Li J, Weng Z, Xu H, Zhang Z, Miao H, Chen W, et al. Support vector machines (SVM) classification of prostate cancer Gleason score in central gland using multiparametric magnetic resonance images: a cross-validated study. *Eur J Radiol*. 2018;98:61–7. <https://doi.org/10.1016/j.ejrad.2017.11.001>.
48. Niaf E, Flamary R, Rouviere O, Lartizien C, Canu S. Kernel-based learning from both qualitative and quantitative labels: application to prostate cancer diagnosis based on multiparametric MR imaging. *IEEE Trans Image Process Publ IEEE Signal Process Soc*. 2014;23(3):979–91. <https://doi.org/10.1109/tip.2013.2295759>.
49. Wu M, Krishna S, Thornhill RE, Flood TA, McInnes MDF, Schieda N. Transition zone prostate cancer: logistic regression and machine-learning models of quantitative ADC, shape and texture features are highly accurate for diagnosis. *J Magn Reson Imaging JMRI*. 2019. <https://doi.org/10.1002/jmri.26674>.
50. Mayerhoefer ME, Schima W, Trattnig S, Pinker K, Berger-Kulemann V, Ba-Ssalamah A. Texture-based classification of focal liver lesions on MRI at 3.0 Tesla: a feasibility study in cysts and hemangiomas. *J Magn Reson Imaging JMRI*. 2010;32(2):352–9. <https://doi.org/10.1002/jmri.22268>.
51. Cheng R, Roth HR, Lay N, Lu L, Turkbey B, Gandler W, et al. Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks. *J Med Imaging (Bellingham, Wash)*. 2017;4(4):041302. <https://doi.org/10.1117/1.jmi.4.4.041302>.
52. Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, et al. The quantitative imaging network: NCI's historical perspective and planned goals. *Transl Oncol*. 2014;7(1):1–4. <https://doi.org/10.1593/tlo.13832>.
53. Cameron A, Khalvati F, Haider MA, Wong A. MAPS: a quantitative radiomics approach for prostate cancer detection. *IEEE Trans Biomed Eng*. 2016;63(6):1145–56. <https://doi.org/10.1109/tbme.2015.2485779>.
54. Khalvati F, Wong A, Haider MA. Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *BMC Med Imaging*. 2015;15:27. <https://doi.org/10.1186/s12880-015-0069-9>.
55. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol*. 2015;25(10):2840–50. <https://doi.org/10.1007/s00330-015-3701-8>.
56. Min X, Li M, Dong D, Feng Z, Zhang P, Ke Z, et al. Multiparametric MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: cross-validation of a machine learning method. *Eur J Radiol*. 2019;115:16–21. <https://doi.org/10.1016/j.ejrad.2019.03.010>.
57. Cuocolo R, Stanzione A, Ponsiglione A, Romeo V, Verde F, Creta M, et al. Clinically significant prostate cancer detection on MRI: a radiomic shape features study. *Eur J Radiol*. 2019;116:144–9. <https://doi.org/10.1016/j.ejrad.2019.05.006>.
58. Ginsburg SB, Algohary A, Pahwa S, Gulani V, Ponsky L, Aronen HJ, et al. Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: preliminary findings from a multi-institutional study. *J Magn Reson Imaging JMRI*. 2017;46(1):184–93. <https://doi.org/10.1002/jmri.25562>.
59. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol*. 2017;27(10):4082–90. <https://doi.org/10.1007/s00330-017-4800-5>.
60. Chen T, Li M, Gu Y, Zhang Y, Yang S, Wei C, et al. Prostate cancer differentiation and aggressiveness: assessment with a radiomic-based model vs. PI-RADS v2. *J Mag Reson Imaging JMRI*. 2019;49(3):875–84. <https://doi.org/10.1002/jmri.26243>.
61. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol*. 2016;40(2):244–52. <https://doi.org/10.1097/pas.0000000000000530>.
62. Nketiah G, Elschot M, Kim E, Teruel JR, Scheenen TW, Bathen TF, et al. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: preliminary results. *Eur Radiol*. 2017;27(7):3050–9. <https://doi.org/10.1007/s00330-016-4663-1>.
63. Chaddad A, Kucharczyk MJ, Niazi T. Multimodal radiomic features for the predicting gleason score of prostate cancer. *Cancers*. 2018. <https://doi.org/10.3390/cancers10080249>.
64. Chaddad A, Niazi T, Probst S, Bladou F, Anidjar M, Bahoric B. Predicting gleason score of prostate cancer patients using radiomic analysis. *Front Oncol*. 2018;8:630. <https://doi.org/10.3389/fonc.2018.00630>.
65. Shiradkar R, Podder TK, Algohary A, Viswanath S, Ellis RJ, Madabhushi A. Radiomics based targeted radiotherapy planning (RadTRaP): a computational framework for prostate cancer treatment planning with MRI. *Radiat Oncol (Lond Engl)*. 2016;11(1):148. <https://doi.org/10.1186/s13014-016-0718-3>.
66. Walsh S, Roelofs E, Kuess P, van Wijk Y, Vanneste B, Dekker A, et al. Towards a clinical decision support system for external beam radiation oncology prostate cancer patients: proton vs photon radiotherapy? A radiobiological study of robustness and stability. *Cancers*. 2018. <https://doi.org/10.3390/cancers10020055>.
67. Abdollahi H, Mahdavi SR, Shiri I, Mofid B, Bakhshandeh M, Rahmani K. Magnetic resonance imaging radiomic feature analysis of radiation-induced femoral head changes in prostate cancer radiotherapy. *J Cancer Res Ther*. 2019;15(Supplement):S11. https://doi.org/10.4103/jcrt.JCRT_172_18.
68. Abdollahi H, Tanha K, Mofid B, Razzaghdoust A, Saadipoor A, Khalafi L, et al. MRI radiomic analysis of IMRT-induced bladder wall changes in prostate cancer patients: a relationship with radiation dose and toxicity. *J Med Imaging Radiat Sci*. 2019;50(2):252–60. <https://doi.org/10.1016/j.jmir.2018.12.002>.
69. Stanzione A, Cuocolo R, Coccozza S, Romeo V, Persico F, Fusco F, et al. Detection of extraprostatic extension of cancer on biparametric MRI combining texture analysis and machine learning: preliminary results. *Acad Radiol*. 2019. <https://doi.org/10.1016/j.acra.2018.12.025>.

70. Ma S, Xie H, Wang H, Yang J, Han C, Wang X, et al. Preoperative prediction of extracapsular extension: radiomics signature based on magnetic resonance imaging to stage prostate cancer. *Mol Imaging Biol MIB Off Publ Acad Mol Imaging*. 2019. <https://doi.org/10.1007/s11307-019-01405-7>.
71. Ma S, Xie H, Wang H, Han C, Yang J, Lin Z, et al. MRI-based radiomics signature for the preoperative prediction of extracapsular extension of prostate cancer. *J Magn Reson Imaging JMRI*. 2019. <https://doi.org/10.1002/jmri.26777>.
72. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: bridging imaging and genomics. *Abdom Radiol (New York)*. 2019;44(6):1960–84. <https://doi.org/10.1007/s00261-019-02028-w>.
73. Pinker K, Shitano F, Sala E, Do RK, Young RJ, Wibmer AG, et al. Background, current role, and potential applications of radiogenomics. *J Magn Reson Imaging JMRI*. 2018;47(3):604–20. <https://doi.org/10.1002/jmri.25870>.
74. McCann SM, Jiang Y, Fan X, Wang J, Antic T, Prior F, et al. Quantitative multiparametric MRI features and PTEN expression of peripheral zone prostate cancer: a pilot study. *AJR Am J Roentgenol*. 2016;206(3):559–65. <https://doi.org/10.2214/ajr.15.14967>.
75. Switlyk MD, Salberg UB, Geier OM, Vlatkovic L, Lilleby W, Lyng H, et al. PTEN expression in prostate cancer: relationship with clinicopathologic features and multiparametric MRI findings. *AJR Am J Roentgenol*. 2019. <https://doi.org/10.2214/ajr.18.20743>.
76. Stoyanova R, Pollack A, Takhar M, Lynne C, Parra N, Lam LL, et al. Association of multiparametric MRI quantitative imaging features with prostate cancer gene expression in MRI-targeted prostate biopsies. *Oncotarget*. 2016;7(33):53362–76. <https://doi.org/10.18632/oncotarget.10523>.
77. Stoyanova R, Takhar M, Tschudi Y, Ford JC, Solorzano G, Erho N, et al. Prostate cancer radiomics and the promise of radiogenomics. *Transl Cancer Res*. 2016;5(4):432–47.
78. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology*. 2013;269(1):8–15. <https://doi.org/10.1148/radiol.13122697>.
79. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2015;21(2):249–57. <https://doi.org/10.1158/1078-0432.ccr-14-0990>.
80. Parra NA, Lu H, Li Q, Stoyanova R, Pollack A, Punnen S, et al. Predicting clinically significant prostate cancer using DCE-MRI habitat descriptors. *Oncotarget*. 2018;9(98):37125–36. <https://doi.org/10.18632/oncotarget.26437>.
81. Parra NA, Lu H, Choi J, Gage K, Pow-Sang J, Gillies RJ, et al. Habitats in DCE-MRI to predict clinically significant prostate cancers. *Tomography (Ann Arbor, Mich)*. 2019;5(1):68–76. <https://doi.org/10.18383/j.tom.2018.00037>.
82. Liao S, Gao Y, Oto A, Shen D. Representation learning: a unified deep learning framework for automatic prostate MR segmentation. *Med Image Comput Comput Assist Interv MICCAI Int Conf Med Image Comput Comput Assist Interv*. 2013;16(Pt 2):254–61.
83. Zhu Q, Du B, Yan P. Boundary-weighted domain adaptive neural network for prostate MR image segmentation. *IEEE Trans Med Imaging*. 2019. <https://doi.org/10.1109/tmi.2019.2935018>.
84. Alkadi R, Taher F, El-Baz A, Werghi N. A deep learning-based approach for the detection and localization of prostate cancer in T2 magnetic resonance images. *J Digit Imaging*. 2018. <https://doi.org/10.1007/s10278-018-0160-1>.
85. Song Y, Zhang YD, Yan X, Liu H, Zhou M, Hu B, et al. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *J Magn Reson Imaging JMRI*. 2018;48(6):1570–7. <https://doi.org/10.1002/jmri.26047>.
86. Giannini V, Mazzetti S, Armando E, Carabalona S, Russo F, Giacobbe A, et al. Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study. *Eur Radiol*. 2017;27(10):4200–8. <https://doi.org/10.1007/s00330-017-4805-0>.
87. Gaur S, Lay N, Harmon SA, Doddakashi S, Mehralivand S, Argun B, et al. Can computer-aided diagnosis assist in the identification of prostate cancer on prostate MRI? A multi-center, multi-reader investigation. *Oncotarget*. 2018;9(73):33804–17. <https://doi.org/10.18632/oncotarget.26100>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received:
16 March 2018Revised:
02 July 2018Accepted:
30 July 2018<https://doi.org/10.1259/bjr.20180270>

Cite this article as:

Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning—a review of recent innovations. *Br J Radiol* 2018; **91**: 20180270.

REVIEW ARTICLE

Automation in intensity modulated radiotherapy treatment planning—a review of recent innovations

¹MOHAMMAD HUSSEIN, PhD, ²BEN J M HEIJMEN, PhD, ^{3,4}DIRK VERELLEN, PhD and ^{5,6}ANDREW NISBET, PhD¹Metrology for Medical Physics Centre, National Physical Laboratory, Teddington, UK²Division of Medical Physics, Erasmus MC Cancer Institute, Rotterdam, The Netherlands³Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel (VUB), Brussels, Belgium⁴Radiotherapy Department, Iridium Kankernetwerk, Antwerp, Belgium⁵Department of Medical Physics, Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK⁶Department of Physics, University of Surrey, Guildford, UK

Address correspondence to: Dr Mohammad Hussein

E-mail: mohammad.hussein@npl.co.uk

ABSTRACT

Radiotherapy treatment planning of complex radiotherapy techniques, such as intensity modulated radiotherapy and volumetric modulated arc therapy, is a resource-intensive process requiring a high level of treatment planner intervention to ensure high plan quality. This can lead to variability in the quality of treatment plans and the efficiency in which plans are produced, depending on the skills and experience of the operator and available planning time. Within the last few years, there has been significant progress in the research and development of intensity modulated radiotherapy treatment planning approaches with automation support, with most commercial manufacturers now offering some form of solution. There is a rapidly growing number of research articles published in the scientific literature on the topic. This paper critically reviews the body of publications up to April 2018. The review describes the different types of automation algorithms, including the advantages and current limitations. Also included is a discussion on the potential issues with routine clinical implementation of such software, and highlights areas for future research.

INTRODUCTION

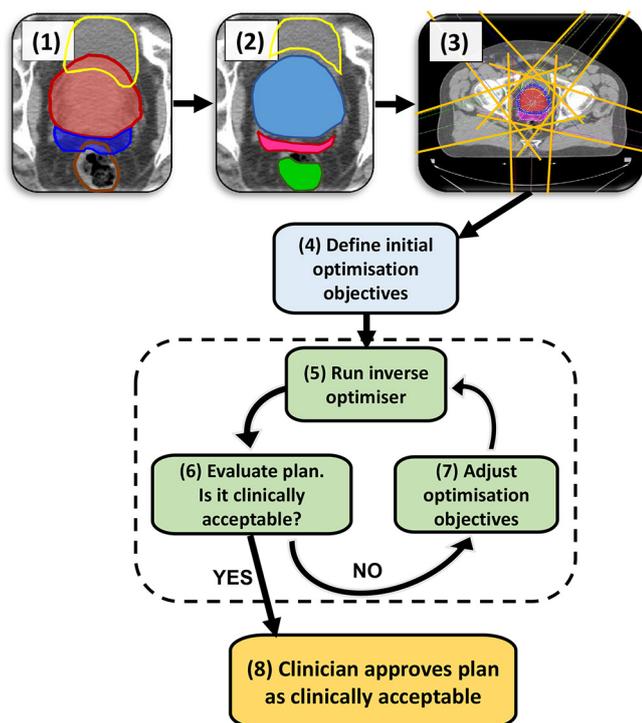
Radiotherapy is a major non-surgical technique in the treatment of cancer. A technology that was once expected to become obsolete in the face of chemotherapy and biological therapy is now used in around 40–60% of all cancer patient cases.^{1–3} This is partly due to technological advances, such as fixed- and rotational-field intensity modulated radiotherapy (IMRT) techniques and image-guided radiotherapy (IGRT), leading to more accurate radiotherapy.⁴ The field has entered an exciting era with rapidly evolving developments such as image-guided four-dimensional adaptive radiotherapy (ART), integration of novel & advanced quantitative imaging and future developments on the horizon such as stratified or personalised radiotherapy.^{5–10}

At the heart of these developments is the central role of optimised, high quality and efficient treatment planning. Complex IMRT techniques have to be inversely planned; this means that a computerised treatment plan is generated by defining a set of objectives and constraints for tumour coverage and healthy tissue sparing and the software uses these to generate a large number of radiation segments to deliver the required dose. Even though the inverse planning

process is highly computerised, it is still human resource intensive, and typically a high level of treatment planner intervention is required to ensure a high-quality plan is produced. A typical inverse IMRT planning pathway is shown schematically in [Figure 1](#) and described in detail in the figure legend. Especially, the depicted interactive feedback loop can lead to variability in inter- and intra-centre plan quality depending on the skills and experience of the operator, which could affect clinical outcome¹¹ and the efficiency in which a plan is produced. Such resource requirements may also limit access to advanced IMRT and emerging treatments such as adaptive radiotherapy, or to suboptimal usage of these techniques.

Efforts to streamline and standardise the treatment planning process are ongoing. In the last few years, there has been significant progress into research and development of automated inverse treatment planning approaches, with most commercial manufacturers now offering some form of solution. There is a rapidly growing body of research published in the literature. These algorithms could significantly improve the efficiency, consistency, and quality of

Figure 1. A typical manual IMRT treatment planning pathway. The example shown is for a prostate + seminal vesicle case. The steps are as follows: (1) CT scan with PTVs and OARs delineated; here the colours of ROIs are red: Prostate PTV, dark blue: SV PTV, yellow: bladder OAR, brown: rectum OAR. (2) create a range of “helper” (ROI) to aid the optimiser; e.g. the part of an OAR not overlapping with the PTV, PTVs overlapping with each other, ring structures to control dose spillage. In the example in Step 2, the ROIs shown are yellow: bladder cropped from prostate PTV, green: rectum cropped from seminal vesicle PTV, magenta: SV PTV cropped from prostate PTV, blue: prostate PTV unedited as it is the higher dose prescription than SV PTV. Step (3) set-up beam geometry. (4) Define the initial optimisation objectives either from scratch or from a class solution. (5) Run the inverse optimiser until it converges to a solution, calculate dose distribution. (6) Evaluate the resulting plan, if it is clinically acceptable proceed to Step 8, otherwise go to Step 7 to adjust the optimisation objectives. The part shaded in green (steps 5, 6, 7) is the iterative process of optimisation required by the planner to arrive at a clinically acceptable treatment plan to be approved by the clinician in Step 8. After this step, the plan will go through the quality control process and preparation for treatment, not shown on the flow chart. IMRT, intensity modulated radiotherapy; OAR, organ at risk; ROIs, regions of interest; PTV, planning target volume; SV, seminal vesicle.



treatment planning, leading potentially to improved patient access and improved patient outcome through maintaining and improving high-quality radiotherapy. In 2014, the National Health Service in England and Cancer Research UK published a 10 year Vision for Radiotherapy in the UK to allow patients to receive advanced and innovative radiotherapy that is cost-effective, and one suggestion to facilitate this is through the

implementation of software that automate parts of the planning process.¹

This paper critically reviews the body of publications up to April 2018. The review describes the different types of automation algorithms for IMRT planning, including the advantages and current limitations. Also included is a discussion on the potential issues with routine clinical implementation of such software, and highlights areas for future research.

LITERATURE SEARCH METHODOLOGY

The literature was searched using Elsevier Scopus®, MEDLINE, Web of Science™ using the following keywords and logic statements: (“automated planning” OR “automatic planning” OR “automation planning” OR “automate planning” OR “knowledge-based” OR “Pareto” OR “multicriteria optimisation” OR “multicriteria optimization” OR “template based optimization” OR “template based optimisation” OR “interactive optimization” OR “interactive optimisation” OR “artificial intelligence” OR “AI” OR “artificial neural network” OR “dose prediction” OR “machine learning” OR “RapidPlan” OR “AutoPlan” OR “rayNavigator”) AND (“radiotherapy treatment planning” OR “radiation therapy treatment planning” OR “IMRT treatment planning” OR “intensity modulated radiotherapy treatment planning” OR “VMAT treatment planning” OR “volumetric modulated arc therapy treatment planning” OR “Tomotherapy treatment planning” OR “radiotherapy planning” OR “radiation therapy planning” OR “IMRT planning” OR “intensity modulated radiotherapy planning” OR “VMAT planning” OR “volumetric modulated arc therapy planning” OR “Tomotherapy planning”).

The search was made on the 7 May 2018 and included articles published up until the end of April 2018. Only full peer-reviewed original research articles written in English were included. There was no specific limit set on the date of earliest publication. After filtering to remove journals unrelated to healthcare and merging the searches from the different databases, 342 eligible records remained. These records were manually scanned based on the title to highlight articles for inclusion. The criteria were to retain articles that clearly described either the idea, development or clinical application of automated inverse treatment planning for IMRT, VMAT, or tomotherapy. Articles that only described automatic selection of beam angles, and did not also describe subsequent automation of the inverse IMRT or VMAT treatment planning, were excluded from this review. While these studies are interesting, the decision was made to apply this criterion to focus the review on the current topical issue of automated inverse planning for IMRT, VMAT or tomotherapy. In articles where the title was deemed ambiguous as to whether it fit the criteria for inclusion, the abstract was read. In total, 171 peer-reviewed papers in scientific journals were included up until the end of April 2018. The earliest publication on the topic of automated planning in IMRT was in 2003; there were other publications pre-2003 related to three-dimensional conformal radiotherapy, however while important, these were excluded from this critical review. Henceforth, we will succinctly refer to automated inverse IMRT planning as “automated planning”.

Figure 2. Trend showing the number of peer-reviewed publications on innovations in automated planning software per year, and the cumulative number of publications. The graph shows a significant increase from 2011.

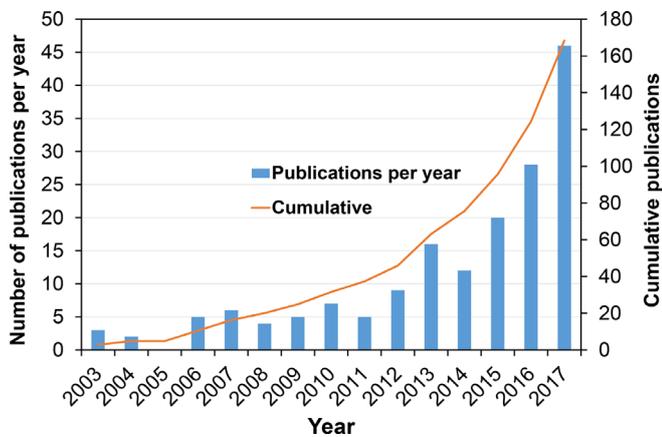


Figure 2 shows the number of publications per year and the cumulative number of papers over the years. The curve exhibits some of the characteristics described by Rogers.¹² There is the initial “lag” period with a steady trend until 2011. After this, there is a sharper uplift in the number of papers published per year; described as the “take-off” phase. Around 2008, major radiotherapy manufacturers began releasing commercial systems and this upward trend represents the effect of the early innovations of automated treatment planning being translated into widely available software. There is no evidence yet of the tail-off phase where the innovation has been so widely adopted that new research is limited. This graph, of course, does not demonstrate the rate of clinical adoption as some of the papers are multiple publications from the same group. Moreover, not all papers related to clinical application.

Through the literature search, three different paradigms of automated planning that were employed in clinical practice are apparent. These are: knowledge-based planning (KBP),^{13–85} protocol-based automatic iterative optimisation (PB-AIO)^{86–110} and multicriteria (or also called multiobjective) optimisation (MCO).^{111–183}

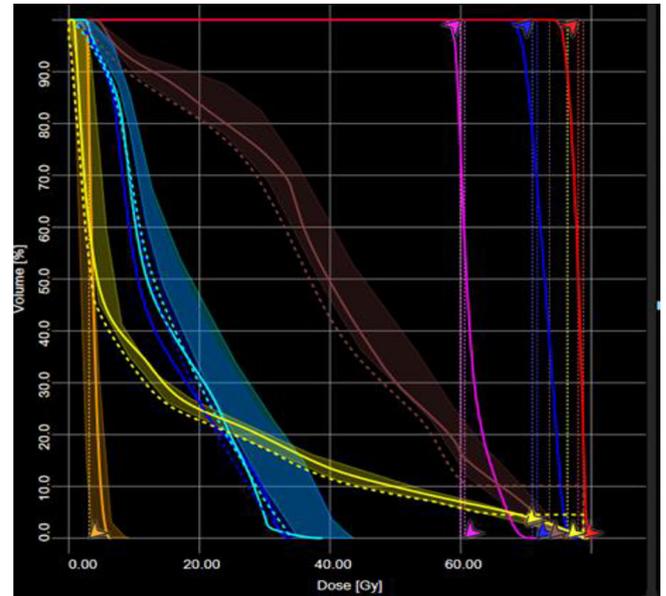
BACKGROUND ON THE DIFFERENT AUTOMATED PLANNING ALGORITHMS

Knowledge-based planning

An approach to improving the speed, efficiency and reducing variability in treatment planning is using so-called KBP approach. KBP is defined as any approach which directly utilises prior knowledge and experience to either predict an achievable dose in a new patient of a similar population or to derive a better starting point for further trial-and-error optimisation by a planner. There are two distinct approaches to this: the atlas-based approach and the model-based approach.

In the atlas-based method,^{13,14,25,36,47,58} the knowledge base is used to select the closest matching patient(s) to give a better starting point for the inverse optimisation than would be provided by conventional template-based approaches. Chanyavanich et al

Figure 3. An example of DVH prediction KBP in a 3-dose level localised prostate cancer case. The shaded lines are the predicted range of achievable DVHs for the different OARs. The solid lines are the actual achieved DVH in the plan. This example is from Varian RapidPlan and the dashed lines and arrows are the optimisation objectives that have been generated by RapidPlan. Courtesy: Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK. DVH, dose-volume histogram; KBP, knowledge-based planning; OAR, organ at risk.



investigated an approach of predicting the starting treatment machine parameters based on a database of prior prostate cancer fixed-field IMRT plans.⁵⁸

Dose-volume histogram (DVH)-guidance is one of the approaches of model-based KBP.^{15–24,26–32,69,80,84,85} In this approach, a large number of clinically accepted treatment plans and contours are used to characterise the relationships between anatomical and geometric features for a given anatomical site to build a predictive DVH model for that site. For any new patient treated in the same anatomical site, this knowledge can be used to predict the achievable DVH based on the features of similar contours and quality of treatment plan; see an example in Figure 3. A range of different implementations of DVH-guided KBP has been proposed and developed. Commercially, the DVH-guidance KBP approach is utilised by the Varian Eclipse® Treatment Planning System as RapidPlan™ (Varian Medical Systems, Palo Alto, CA).

A known limitation of the DVH-guidance approach is that the DVHs are only predicted for the regions of interest that are delineated. This means that regions of tissue outside of delineated regions of interest (ROIs), which a human planner may also optimise to reduce dose, may not be taken into account, e.g. to enhance conformality or avoid hot spots. Additionally, DVHs do not provide any spatial information. An interesting approach that has been investigated to overcome these issues is voxel-based dose prediction. Rather than predicting DVHs, the idea is to use

knowledge from prior plans to build a model that can predict doses to individual voxels within the patient's image.^{33–35,38,39} A limitation of the model-based dose prediction approach is that the plan quality for new patients strongly depends on the quality of plans generated in the past.

Protocol-based automatic iterative optimisation

The challenge with manually optimising a plan is that it is sometimes a time-consuming process to arrive at a clinically acceptable plan. Moreover, it is often not clear if the plan could be better if further adjustments of the optimisation criteria were made. The clinically optimal plan is one where there is the best trade-off between normal tissue sparing and target coverage, taking into account the clinical requirements and priorities regarding sparing of the various tissues. For example, one may consider a head and neck cancer where the PTV abuts the spinal cord. In this case, it is typically the highest clinical priority to keep the spinal cord within tolerance and therefore, requires compromise of PTV coverage. It is relatively straightforward to achieve a plan that meets the spinal cord tolerance (and may, *e.g.* still achieve the PTV D95% objective). However, a better plan may be one that keeps the spinal cord just within tolerance while maximising the coverage of the PTV, as well as pushing the lower doses to other healthy tissue as low as possible. Achieving this better plan manually would require significant time, effort, and planner experience, as it requires iterative adjustment of optimisation criteria to keep challenging the optimiser to achieve a progressively better plan until no improvement in plan quality is possible. The success of performing this process manually and efficiently will depend on the skill and experience of the treatment planner, and the time available to plan.

One approach to solve this is to automate the iterative adjustment of the optimisation objectives and constraints.^{86–88,98,104–110} The basic idea is to start with a user-defined template which has the required clinical objectives. The user can then input the priorities for mandatory (hard) constraints. The optimiser then generates a plan that meets all the objectives, at which point the high priority constraints are locked down and become hard constraints. A script is then put in motion which iteratively pushes the DVH of all of the structures to the point where the hard constraints were just breached, and then a step is taken back to the point where the breach did not occur. At this point, the plan cannot be pushed further and could be considered the optimum plan.

Tol *et al*¹⁰⁷ developed an interface to move the mouse cursor on the computer screen automatically and thus adapt the optimisation objectives in the Varian Eclipse VMAT optimiser. The interface detects the position of the DVH line for each ROI on the screen and iteratively moves fixed objectives to ones more challenging during the optimisation and was shown to be able to automate VMAT planning of head and neck cancer giving improved dosimetric results.¹⁰⁷

Various authors have developed artificial intelligence (AI) systems which simulate the reasoning behaviour of a human planner to automatically adjust the optimisation parameters during the optimisation process.^{98,104,105,109} Such AI systems

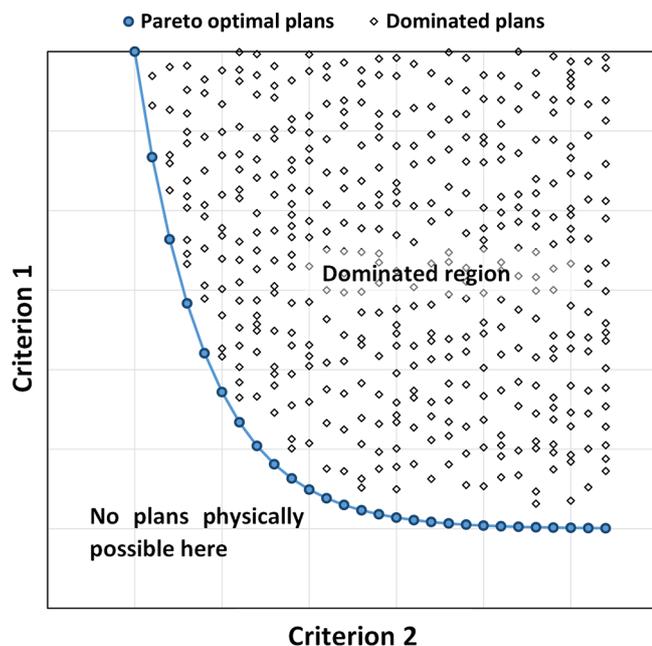
have been based on fuzzy logic theory whereby the trial-and-error actions of expert human planners were converted into binary “IF-THEN” logic statements.

A commercially available solution is AutoPlanning within the Philips Pinnacle³ TPS (Philips Radiation Oncology Systems, Fitchburg, WI). The user initially generates a template (“Treatment Technique”) which has the target prescriptions and the goals for organ at risk (OAR) sparing according to the required clinical protocol. For OARs, the user also specifies their clinical importance, from those that have low significance to those that have hard constraints. Based on the PTV(s) and OARs defined, the software automatically generates “dummy” optimisation structures such as those that take into account overlap between OAR and PTV, PTV ring structures to control dose fall, and various other “help” structures to control target uniformity and dose spillage to the rest of the body. There are also additional “advanced” settings to control dose fall-off, homogeneity and managing cold/hotspots, which initially have default factory set values or could be fine-tuned by the user. Based on the optimisation contours and the settings used, the software automatically generates the starting optimisation criteria. The software then enters into a 5-loop iterative optimisation cycle to gradually fine-tune the plan to achieve a solution based on the clinical protocol as defined by the user.^{89,103}

Multicriteria optimisation

Another approach which seeks to overcome the issue of finding the optimal trade-offs between target coverage and sparing of all normal tissues is called MCO (also sometimes called multiobjective optimisation).^{121,131,140,143,146,184} Central to MCO is the concept of the “pareto optimal solution”; which is a plan that cannot be improved in any of the objectives without degrading at least one of the other objectives. There are two approaches to MCO, *a posteriori* and *a priori* approach. In the *a posteriori* approach, rather than the optimiser generating a single plan, multiple plans are automatically generated where each criterion is optimised to the extent where it cannot be improved upon without affecting at least one other criterion; each of these plans is a so-called pareto optimal solution. The schematic in Figure 4 illustrates this concept with a graph of two competing criteria. The graph shows a large number of different feasible planning solutions, representing a variety of different permutations for criterion 1 and 2. The solid line represents the pareto front where improving one criterion inevitably leads to the worsening of the other and vice versa. Plans that lie on this front are the “pareto optimal solutions”; shown as blue circles in the schematic. The plans shown as diamonds are referred to as “dominated” because there is always a solution on the pareto front where at least one criterion can be improved. Pareto optimality by itself does not imply clinical optimality and pareto optimal plans can be clinically highly undesirable. On the other hand, the best clinically acceptable plan is pareto optimal. Therefore, in the *a posteriori* approach the database of pareto optimal plans is interactively (*a posteriori*) navigated by the treatment planner to choose a clinically optimal plan.^{127,131,140,144} The automation in this process is that the database of pareto optimal plans is automatically generated. The main issue with this approach is the number of

Figure 4. Schematic diagram of two competing criteria. The graph shows a large number of different feasible planning solutions, representing a variety of different permutations for criterion 1 and 2. The solid line represents the pareto front where improving one criterion inevitably leads to the worsening of the other and vice versa. Plans that lie on this front are the “pareto optimal solutions”, shown as blue circles in the schematic. The plans shown as diamonds are referred to as “dominated” because there is always a solution on the pareto front where at least one criterion can be improved.



plans that are generated, since mathematically there is an infinite number of pareto optimal plans, and the intensive computing resource required can be a limiting factor. Moreover, especially in the case of a large number of clinical objectives, selection of a plan may be difficult and operator-dependent. The *a posteriori* approach has been implemented in the commercial RayStation TPS, and recently as an option in the Varian Eclipse TPS. In this approach, the dimensionality of the pareto fronts is dictated by the number of objectives, and thus also the number of plans that are required to construct the fronts. Craft and Bortfield describe a method to estimate the number of plans that are sufficient¹²⁸ and suggest that $N+1$ plans are sufficient, where N is the number of objectives. At the time of writing, this recommendation is the basis of the number of plans generated in RayStation rayNavigator. Different methods were described in the literature for interactively navigating the pareto fronts, using a navigation star^{127,131,140,144} or sliders, however, the premise is still the same. The lowest and highest values for each objective is displayed visually along with a starting dose distribution and DVH line. The user then uses the mouse to drag the objective of interest and in near real-time, the software finds the plans in the database that are better in the selected criterion, and then, via fast interpolation, the dose distribution and DVH are updated.

In the *a priori* MCO approach, for each patient, only a single pareto-optimal plan is fully automatically generated. This plan

has clinically desired trade-offs between all objectives, in line with the institution’s clinical protocol and treatment tradition.^{143,146,185} Optimization is based on a treatment site-specific protocol, a so-called “wish-list”, containing the objective functions with assigned priorities and hard constraints that should never be violated. An example wish-list for automated plan generation with iCycle for localised prostate cancer patients is in Table 1. In an automatic multiobjective optimisation approach, the objectives are sequentially minimized according to their priorities to obtain a pareto optimal plan with favourable balances between all objectives. Wish-lists are treatment site specific, *i.e.* no patient-specific adaption is applied. They are generated in an iterative tuning process, involving the multidisciplinary planning team. In this process, a first estimate of the wish-list is made based on a review of plans of recently treated patients, the clinical protocol, and initial team discussions. This wish-list is then improved in several iterations consisting of (1) use of current wish-list to automatically generate treatment plans for CT-scans of a small group of previously treated patients (typically five), (2) evaluate the automatically generated plans (including comparison with clinically applied plan), (3) update the current wish-list (new estimate), (4) go back to (1). This iterative improvement of the wish-list is stopped when further improvements of the wish-list are deemed not possible. This iterative wish-list improvement has an intrinsic drive to improve the clinical plan quality. This *a priori* MCO approach has been developed and implemented in the Erasmus MC Cancer Institute in their “Erasmus-iCycle” software.^{129,143,146,162} Apart from beam profile optimisation, the system also features automated beam angle optimisation.¹⁶¹ As well as optimisation for regular linacs, Erasmus-iCycle has separate models for optimization of Cyberknife treatments¹⁸⁶ and proton treatments (intensity modulated proton therapy, IMPT).¹⁸⁷ Currently, Elekta AB (Stockholm, Sweden) is preparing a commercial implementation of the system for photon beams.¹⁵⁰

CLINICAL EVALUATION AND IMPLEMENTATION OF AUTOMATED PLANNING TECHNIQUES

There are many studies in the literature which have clinically implemented commercial and in-house implementations of PB-AIO, KBP, and MCO automated planning.^{42–46,48–57,59–68,70–79,81–83,89–97,99–103,150–155,157–166,168–177,179–181} Most of these articles tackle the current commercial implementations of Varian RapidPlan (KBP), Pinnacle AutoPlan (PB-AIO), Raystation (*a posteriori* MCO), and Erasmus-iCycle (*a priori* MCO).

From the present literature review, 81 (~43%) papers were reporting on the clinical evaluation, implementation, or application of automated planning. It was noted that most studies were retrospective and different methodologies were reported for evaluating automated plans and manual plans. The most popular method employed, in 67 of the 81 papers, was evaluation based on comparing DVH metrics for PTVs and OARs, or deriving other metrics from DVHs such as conformity index, homogeneity, tumour control probability, normal tissue complication probability. Some papers used these in conjunction with qualitative blinded clinician evaluation which took the format of

Table 1. An example wish-list for automated plan generation with Erasmus-iCycle for localised prostate cancer patients

Constraints				
	Volume	Type	Limit	
	PTV	Max dose	105% of D_{Px}	
	PTV	Mean dose	101% of D_{Px}	
	Rectum & anus	Max dose	102% of D_{Px}	
	PTV shell 50mm	Max dose	50% of D_{Px}	
	Unspecified tissues	Max dose	105% of D_{Px}	
Objectives				
Priority	Volume	Type	Goal	Parameters
1	PTV	↓LTCP	0.8	$D_{Px} = 78\text{Gy}$, $\alpha = 0.8$
2	Rectum	↓EUD	20Gy	$k = 12$
3	OAR 2	↓EUD	10Gy	$k = 8$
4	PTV shell 5mm Skin ring 20mm	↓Max dose ↓Max dose	80% of D_{Px} 20% of D_{Px}	
5	Rectum	↓Mean dose	5Gy	
6	Anus	↓Mean dose	5Gy	
7	Bladder	↓Max dose	5Gy	
8	PTV shell 15mm PTV shell 25mm	↓Max dose	50% of D_{Px} 30% of D_{Px}	
9	Left & right femoral heads	↓Max dose	50% of D_{Px}	

α , cell sensitivity; EUD, equivalent uniform dose; k , volume effect; LTCP, logarithmic tumour control probability; PTV, planning target volume; D_{Px} , prescribed dose.

The priorities assigned to the objectives are used in the *a priori* MCO, guaranteeing for each patient generation of a pareto-optimal plan with clinically favorable balances between all treatment objectives. (Courtesy: A.W. Sharfo).

either binary decisions on whether plans were clinically acceptable^{95,162,169,172} or giving plans a ranking.^{93,103,164} Most studies are retrospective, comparing already delivered clinical plans with automatic plans. Voet *et al* performed a prospective study on automated planning of head and neck cancer with Erasmus-iCycle giving, for each new patient, the treating clinician the choice between a plan made with routine trial-and-error planning and an automatically generated plan.¹⁶² Hansen *et al* prospectively compared Pinnacle³ AutoPlan vs manual planning in head and neck cancer with blinded review by three clinicians to select the better plan.⁸⁹

Knowledge-based planning

KBP has been clinically investigated in various clinical sites, with most studies reporting that KBP is at least as good, or (slightly) better, than manual planning with improved efficiency and consistency, without manual intervention.^{42–46,48–57,59–68,70–79,81–83} All of the published studies are via the use of the commercial implementation of KBP in the Varian RapidPlan solution. Reports have been published for head and neck cancer,^{72,79,82} prostate cancer,^{62,66,78,83,188} cervical cancer,⁷⁸ lung cancer,^{75,83,189} spinal metastasis,⁶³ breast cancer,⁶⁵ upper gastrointestinal (GI) cancers,^{61,70,76} and lower GI cancers.¹⁹⁰ Foy *et al* reported that KBP could reduce the VMAT planning of stereotactic body radiotherapy of the spine from 1–1.5 h to around 10–15 min.⁶³ Hussein *et al* reported on the clinical validation and benchmarking of the

commercial RapidPlan KBP system for both IMRT and VMAT planning in prostate and cervical cancer. The authors highlighted that using the software “out-of-the-box” with the default settings for training the KBP models lead to automated plans with poor conformity, coverage and plan generation efficiency compared to the original clinical plans, and that an iterative process is required to fine-tune and optimise the model. After this refinement of the model was performed, the authors showed that RapidPlan was able to achieve better or comparable plans when compared to the original clinical plans.⁷⁸

Typically, a dose prediction KBP model is trained for one particular technique and clinical site, meaning that the model has been characterised for that particular population of patients; take, *e.g.* a prostate static field IMRT model. Suppose that the treatment technique was changed to VMAT or the model was shared with a centre that does not have the capability for VMAT. An option is to create a VMAT specific model, which requires replanning of a large number of patients followed by the refinement of the model. However, as the KBP model predicts the dose based on the anatomy of the patient and not treatment technique, there is the potential that (in this example) the IMRT model could be used outside of its original scope for VMAT planning. This is an interesting research question to demonstrate how robust a model is to changing techniques and sharing between centres. Additionally, broadening the scope of the model further by including

both IMRT and VMAT plans may potentially improve plan quality. Cagni et al investigated the use of helical tomotherapy plans to create KBP models for prostate cancer and found that this could be successfully performed.⁵³ Other areas of interest are whether a model that was trained for a particular clinical site could be used in another site with similar anatomy and similar *relative* dose levels to the original model. Some of these areas have been addressed in a limited number of studies but is still an area of active research.^{61,66,68,74,76,78} Wu et al investigated using a RapidPlan model trained on VMAT rectal plans treated in the supine position to create plans in patients treated with IMRT and those who were set up in the prone position. The study found that OAR sparing and plan consistency was improved but that the optimiser needs to be readapted to IMRT planning and that manual hotspot reduction is required.⁷⁴ Most of the reported studies in the literature focussed on single institution analysis. The performance of a broad scope RapidPlan KBP model for oesophageal cancer was investigated by Fogliata et al,⁷⁶ whereby the model took into account different dose prescriptions and tumour locations with the ultimate aim of determining whether the model could be shared amongst different centres where variations in clinical protocols can occur. The authors carried out the study across three centres, where one centre did not contribute any data to the model. In the latter centre, KBP resulted in superior plan quality. The study highlights the potential benefit of a heterogeneous data set, and this has also been highlighted by other studies that suggest that keeping statistical (but clinically relevant) outliers may be an advantage to the model strength.^{57,78}

An apparent limitation of the KBP approach is that the models can only be as good as the training data that has been input in the first instance. Strictly speaking, the plans can be clinically acceptable but not the optimal plan. RapidPlan attempts to get around this by always placing the optimisation objectives for OARs lower than the predicted DVH such that it always tries to improve on the prediction.

Protocol-based automatic iterative optimisation

Papers in the literature have generally reported that PB-AIO, commercially implemented within the Pinnacle TPS, is either equivalent to or superior to manual planning regarding plan quality and efficiency in automatically generating IMRT or VMAT plans in various clinical sites.^{89–97,99–103} Hazell et al compared PB-AIO with manual planning of 26 IMRT head and neck cancer plans and evaluated 2 types of plans through DVH metrics and clinician-blinded reviews. They found comparable target coverage and better sparing of normal tissues, with all plans clinically acceptable without manual intervention.¹⁰³ Hansen et al extended the analysis to VMAT and found similar results and reported that planner time was halved from 64 min using PB-AIO.⁸⁹ Speer et al used a quantitative point-based scoring system where treatment plan parameters were scored to objectively judge plan quality of PB-AIO over manual planning in head & neck cancer, where a score of 100 points indicates an optimum plan. They demonstrated automated plans using PB-AIO were better with an average score of 62.3 points compared to manual plans with a score of 59.1 points.¹⁰⁰ Similar conclusions about the potential for PB-AIO to efficiently produce clinically acceptable

plans for head and neck cancer have been reported.^{97,101,91} Nawa et al compared PB-AIO with manual planning in 23 prostate cancer cases. Comparison was performed using DVH objectives, and the study found target coverage and rectal dose was comparable between PB-AIO and manual planning. There was a significant reduction of the dose to the bladder and femoral heads with PB-AIO compared to manual plans. This is potentially due to the manual plans not pushing these doses down further as they had already passed the clinical tolerances and more attention was paid to the harder to achieve rectal dose constraints. Additionally, the authors quantified a reduction in interoperator variability with PB-AIO.⁹¹ PB-AIO has been evaluated for hippocampal sparing whole brain radiotherapy^{87,99} and was found to result in comparable or better plans with minimal manual intervention and expedited planning time which is essential in this palliative group of patients. Studies in oesophagus,^{92,95} and rectal cancer⁹³ cases are also consistent in their conclusions about the potential benefit of PB-AIO over manual planning.

However, whilst the overall message in all of these papers is favourable for PB-AIO, some studies argue that PB-AIO is a tool to improve overall plan quality, but not necessarily to completely remove the need for manual optimisation¹⁰³ and that for some particular cases, experienced planners performed better than PB-AIO.¹⁰⁰ The quality of plans generated by Auto-Planning in Pinnacle has also been reported to be dependent on the input from experience treatment planners to set the initial user settings and define good clinical protocols which are also an important consideration.¹⁰⁰

Multicriteria optimization

A posteriori MCO clinical implementations (all in RayStation) have been investigated and validated in different clinical sites including prostate cancer,^{168,169,171,176,179} head and neck cancer,^{164,171} brain,¹⁶⁸ lung cancer,¹⁷² and lower GI cancers.¹⁷⁷ All studies report better or comparable plan quality with a reduction of planning time. Wala et al report that MCO using RayStation took approximately 1 h per case and achieved superior plan quality based on blinded review and DVH objective comparisons for localised prostate cancer.¹⁶⁹ Other studies also report comparable or superior plan quality of *a posteriori* plans in prostate cases. Muller et al report a reduction in planning time by around 10 min for post-prostatectomy cases and 45 min for brain tumour cases.¹⁶⁸ Chen et al used MCO to generate 20 field IMRT plans for prostate cancer and head and neck cancer to then use the resulting DVH information as the basis for defining optimisation objectives for VMAT plan optimisation. Using this method, they were able to match the quality of single arc VMAT with the quality of a 20 field IMRT plan (where there is high quality due to the large degree of freedom).¹⁷¹ Kamran et al evaluated the potential benefit of *a posteriori* MCO in 10 patients with non-small cell lung cancer who were eligible for the RTOG 1308 Phase II trial.¹⁷² Evaluation of plan quality between MCO and manual planning was performed via a double-blinded review and DVH metrics. While all the MCO plans passed the DVH objectives, it was noted that clinicians preferred 8/10 of the MCO plans. The two manual plans were chosen due to better skin sparing and a lower maximum dose to the spinal cord, even

though the oesophagus dose was lower. Planning time improved by a median of 88 min.¹⁷²

Navigation-based (*a posteriori*) MCO could potentially avoid the often iterative interaction between planner and physician to arrive at a clinically acceptable plan by involving the clinician at an earlier stage. However, it is not clear as to whether, practically, this results in an improvement in the planning workflow particularly given the often limited clinician time. Müller *et al* investigated this retrospectively in prostate and brain cancer planning and demonstrated potential time savings, but further prospective studies are required.¹⁶⁸ In the context of a technique like *a posteriori* MCO, how the clinician and planner role will evolve will likely vary between different countries. For example in the UK, the Clinical Oncologists are responsible for administering both radiotherapy and other non-surgical treatments such as chemotherapy whereas in other countries there is a different approach with Radiation Oncologists who are more focused in radiotherapy. Therefore, the practicality of *a posteriori* MCO in the hands of clinicians in the UK will likely be different. One approach may be treatment planners taking on an extended role, or formalising an existing role, in the decision-making aided by the availability of other relevant clinical data and with the appropriate training and qualifications.

A limitation of current implementations of *a posteriori* MCO is that the plans optimised are *near* pareto optimal in the fluence space and do not directly consider the machine parameter optimisation. The final navigated plan is then converted into a deliverable using direct aperture optimisation. However, McGarry *et al* and Kyroudi *et al* have shown that dosimetric discrepancies can occur between the conversion of the navigated plan into a deliverable plan, and therefore may not reflect the clinical preferences that resulted in the choice of the navigated plan and this may increase the uncertainty in the plan navigation process.^{165,179} In most cases, this dosimetric difference may not translate to a clinically significant difference, and the advantage in the ability for navigating the trade-offs was retained. However, for some cases where there are small targets on low-density tissues, the dosimetric difference can be significantly larger such that manual fine-tuning is likely required.¹⁶⁵

A priori MCO with Erasmus-iCycle for linacs was validated for head and neck cancer,¹⁶² prostate cancer,^{150,185} cervical cancer,¹⁷⁴ lung cancer,¹⁸⁰ spinal metastases,¹⁷⁵ and gastric cancer.¹⁵¹ By itself, Erasmus-iCycle had fully automated MCO for beam fluence optimization. For the generation of clinically deliverable plans, the system was used as a pre-optimizer for the commercial Monaco TPS (Elekta AB, Stockholm, Sweden), which generates a deliverable, segmented plan that mimics the pre-optimized Erasmus-iCycle dose distribution. In the first evaluation study (on head and neck cancer), this plan reconstruction in Monaco was performed by a planner.¹⁶² In all later studies, the reconstruction in Monaco was fully automated, *i.e.* for each patient the Erasmus-iCycle dose distribution was used to automatically create a *patient-specific* Monaco planning template, which was then used by Monaco for automated generation of a deliverable VMAT or IMRT plan, mimicking the Erasmus-iCycle plan. In all

validation studies, the automated Monaco plans were compared with manually generated Monaco plans. Manual fine-tuning of automatically generated plans was not performed. For all investigated treatment sites, there was a considerable reduction in hands-on planning time, which virtually reduced to zero with fully automated planning. For treatment of the prostate only or prostate with seminal vesicles, plan quality between automated and conventional plan generation was similar.¹⁸⁵ For treatment of the prostate with seminal vesicles and elective nodal irradiation and all other sites, the quality of the automatically generated plans was superior. In the prospective head and neck cancer study,¹⁶² treating clinicians could, for each patient, choose between an automatically and a conventionally generated plan. In 97% of cases, preference was given to the plan that was generated with Erasmus-iCycle. At the Erasmus MC Cancer Institute, fully automated planning with the combination Erasmus-iCycle/Monaco was in routine use for prostate, cervix, lung, and head and neck cancer patients.

NOVEL APPROACHES TO USING AUTOMATED PLANNING ALGORITHMS

Automated planning as a plan quality assessment and checking tool

It was recognised in several studies that the principle of KBP can be used both as an automated planning tool and as a plan quality assessment tool. This is because the first component of KBP is to use prior knowledge to predict the achievable dosimetry for a prospective patient. This information can be used to judge the quality of a plan and has been shown to be effective by several groups.^{32,36,59,65,71,77,188,191} Additionally, it has also been reported to be a useful tool for training new staff members and improve the quality of manual planning, and also useful in clinical trial QA.⁶¹

“Bias-free” comparison of different treatment techniques using automated planning

Comparing different treatment techniques (*e.g.* VMAT and IMRT) in a treatment planning study can be prone to human subjectivity and bias, particularly there can be questions to what extent there were differences in the optimality of the plans for the compared techniques. An interesting approach is to use automated planning in these studies.^{163,166,173,181,192} Boylan and Rowbottom developed a PB-AIO approach and applied it in comparing seven fixed-field IMRT with two arc VMAT for nasopharyngeal head and neck cancer patients using a standard protocol, and to investigate two experimental strategies (a parotid-sparing strategy and dose escalation strategy).¹⁹² They showed that the IMRT and VMAT techniques were clinically comparable for the standard and dose escalation protocols, whereas VMAT was better in the parotid sparing strategy. Lechner *et al*, used MCO to objectively compare the quality of flattening filter free IMRT and VMAT vs flattening filter plans in prostate and head and neck cancer patients.¹⁶³ Sharfo *et al* used automated planning for bias-free comparisons of IMRT and VMAT techniques for cervical cancer.¹⁸¹ They demonstrated that a 12 field IMRT technique had similar quality as a dual-arc VMAT technique. Sharfo *et al* also used bias-free automated planning for comparison of liver SBRT with a fully non-coplanar technique, coplanar

VMAT, or a new approach called VMAT+.¹⁶¹ The latter was defined as VMAT supplemented with 1–5 computer-optimized non-coplanar beams. Regarding plan quality, they demonstrated that VMAT+ was superior to VMAT, and almost as good as fully non-coplanar. Treatment times with VMAT+ were much shorter than with fully non-coplanar treatment.

In prostate cancer (and other pelvic malignancies), one of the common challenges faced by treatment planners is the scenario where a patient has an artificial metallic hip implant. This presents a challenge to the planner as a typical technique is to avoid beams entering through the implant due to uncertainties in the density, which limits the permissible beam directions. The difficulty is amplified if the patient has bilateral implants. Voet et al, reported on using *a priori* MCO to automatically investigate different fixed-field IMRT strategies using the iCycle software which was able to optimise both beam angles and fluence profiles.¹⁷³

Automated planning as a decision support tool for treatment selection and personalised treatment

Studies have reported on the use of automated planning as a useful tool for making informed decisions on a patient's eligibility for specific novel radiotherapy techniques. The potential advantage of automated planning in this context is the quick production of plans for different techniques and clinical scenarios, where otherwise resource-intensive manual plan generations would be required.

One application of this is patient selection for proton therapy. Automated planning is used to select patients suitable for whom a proton treatment plan may be suitable and to promote that only patients that would clinically benefit from the treatment are selected and to thus avoid using this limited resource in patients where potentially a faster photon-based plan would be equally or more clinically useful. Delaney et al found that a KBP model based only on prior photon VMAT plans was able to predict proton DVHs and therefore, may be used in identifying patients for proton therapy.⁶⁴ Bijman et al used MCO to generate photon and proton plans, and while the primary focus of their study was in the context of analyzing the uncertainty of using normal tissue complication probability models for patient selection, the use of MCO for fast plan generation was still demonstrated.¹⁹³

In patients with liver tumours (such as from hepatocellular carcinoma or oligometastatic disease) and who are contraindicated for surgical intervention, stereotactic ablative radiotherapy (SABR) is a promising treatment modality. However, there are some limiting factors and criteria which dictate whether a patient is eligible for SABR. The most significant of these is the dose to the healthy liver tissue, which will vary depending on the tumour volume and liver volume. In practice, the decision on eligibility can only be determined once the trial-and-error planning has been attempted, which is an inefficient use of resources. Tran et al, reported on using KBP as a tool for predicting patient eligibility for liver SABR and to also determine whether the patient would benefit from a more complex non-coplanar technique than a standard coplanar VMAT technique.²⁴

Ronde et al investigated the feasibility of using MCO for shared decision-making in anal cancer and conclude that patient-clinician preference-informed plan selection is feasible.¹⁷⁷ Smith et al described a novel approach to personalized treatment planning by integrating a model of radiotherapy outcome with MCO for prostate cancer treatment.¹⁷⁰ The MCO model generates the set of pareto optimal plans which are then integrated into a Bayesian network to model the probabilities of outcomes such as toxicity, recurrence, distant metastasis. To predict these probabilities, the model uses information from expert opinion and published data, and patient characteristics such as clinical staging, Gleason score and PSA. The final step is use these probabilities in a Markov model then to predict Quality Adjusted Life Expectancy which is then the final basis for ranking and selecting the best plan. This approach appears to be promising; however, the authors point out that further work is required to validate the accuracy of the predictions of outcomes. A similar approach has also been reported for glioblastoma.¹³⁹ Valdes et al describe an AI approach which identifies previously approved treatment plans which are achievable for a prospective patient to aid decision made on a personalized level.⁴⁰

Automated planning for plan library in plan-of-the-day (PotD) adaptive radiotherapy

Due to various challenges in modern radiotherapy, such as limitations in accuracy in automated image segmentation and automated planning speed, daily online replanning based on daily reimaging has not been routinely applied. A simpler, but more feasible approach is PotD ART, which has been reported for bladder cancer,¹⁹⁴ cervical cancer¹⁹⁵ and rectum cancer.¹⁹⁶ In this approach, the patient is imaged daily and then treated with a plan selected from a pre-treatment established patient-specific plan library. The library contains plans for various patient anatomies. The PotD is the library plan that best fits with the anatomy-of-the-day. PotD ART involves generation of multiple plans for each patient, increasing the planning workload for a department. Heijkoop et al have avoided this problem by applying automated plan generation.¹⁹⁵ For each cervical patient, they created a library with up to three plans.

DISCUSSION

Over the last few years, innovations in automated treatment planning software have led to the potential to improve the efficiency and quality of radiotherapy treatment planning. The gain in plan quality and reduced interplanner variation may have clinical benefit for patients by removing the low-quality outliers and therefore potentially cure more patients, however, these need to be demonstrated with clinical evidence.

There has been a rapid increase in the number of papers published in this field and there were a variety of approaches and commercial implementations. In general, most papers in the literature showed improvements over manual planning across a variety of clinical sites. This review covered innovations in automated planning, but the patient's treatment planning pathway also involves the contouring of ROIs and the quality assurance procedures to ensure safe delivery of plans. Automation of these areas is also being addressed but have not been covered in this review.^{197–204}

Real-time interactive planning is a new paradigm for IMRT planning that is not based on using traditional optimisation algorithms. Instead, the goal is to be able to give the treatment planner the ability to perform real-time and interactive manipulation of the isodose or DVH lines and ultra-fast (automated) reoptimisation and dose estimation to update the user on the impact of interactive modifications. The main potential for this type of semi-automated planning is (along with advancements in imaging, auto-segmentation, and fast plan verification) to realise dose-guided, fast and intuitive adaptive replanning. This approach was still in the very early stages in the research domain, and as such there were limited proof-of-principle papers that were published,^{205–207} but it is anticipated that these systems may become commercially available within the next few years. Real-time interactive planning and other innovations in fast reoptimisation will play an essential role in the topic of on-table adaptation which was being made possible due to developments such as MR-radiotherapy machines.^{208,209} This is an area of active research, and some preliminary publications have been published, some of which use similar principles for automation of the plan reoptimisation as those discussed in this review.^{210–214} These developments were also in parallel with necessary developments in automation of contouring and QA which are also central to this concept.

There were a few research areas that require further study, which we highlight here. Some studies suggested that cross-institutional sharing of knowledge may improve the quality of the automated plan generation;^{66,68,76} however, further research on the impact of sharing of KBP models and *a priori* MCO wish-lists for more clinical sites was needed. Furthermore for KBP, the size and heterogeneity of the dataset required and the robustness of the resulting model needs further investigation. An area of automation where there were limited papers is in IMPT, and future developments will be needed as the number of centres that will be treating with IMPT is gradually increasing.^{64,193,215} Also, automated plan generation including (non-coplanar) beam angle optimization (or beam arc optimization for VMAT) was only explored with “bias-free” comparisons in a limited number of studies, however showing promising results.^{163,166,173,181,192} The use of automated planning in randomised clinical trials offers the opportunity to reduce the level of variability of plan quality which may affect the clinical outcomes, and the reduced variability may also lead to the possibility of reducing the required sample sizes meaning trials could potentially recruit quicker. This use of automation may also facilitate randomised trials testing two extreme treatment techniques which may be manually challenging to achieve at the desired plan quality. A possible example of this could be comparing treatment with a maximum focus on sparing of OAR1 with another treatment that maximally spares the competing OAR2. The endpoint of such a study could be the patients’ Quality of Life. Therefore, the role of automated planning within clinical trials is an important area of future investigations. Sharing of automated techniques between centres may lead to improved consistency between centres (perhaps even worldwide), particularly in rare diseases where adequate patient numbers to develop expertise may only be possible across a handful of centres. However, a significant barrier was that the

different types of automated planning implementations presents a challenge for sharing of different experiences between centres and commercial vendors. Furthermore, the equivalence between automated planning techniques was not well-known. Of potential interest is to investigate whether one automated planning technique could be ported to another. The role of automated planning for shared decision-making and personalised radiotherapy has been briefly discussed in this review but this research was in the early stages and it is anticipated that further studies will follow. For this to become more feasible, plan quality should ideally be linked to patient outcome and not DVH-like metrics. Finally, guidelines and recommendations on how to perform planning studies, agreed by planning experts and endorsed by professional bodies, are also highly needed. Planning studies can have a focus on development and evaluation of novel strategies/algorithms for automated planning, or automated planning may be used for investigating clinical questions, *e.g.* are protons better than photons. For both types of studies, statistical power and avoidance of bias (*e.g.* related to diversity in human planner experience) has to be considered. In the technical studies, clinical deliverability of generated plans may not always have the highest urgency, whereas this seems a must when using planning for comparing different treatment strategies in a clinical setting. Other factors that have to be considered in guidelines are quality of dose calculation algorithms, adequate coping with geometrical uncertainties (*e.g.* using margins, and appropriate metrics for comparing plans).

While there were some papers in the literature highlighting the technical feasibility of automated planning over manual planning, there were limited studies that describe the practicalities of implementing automated planning, including more detail on how the treatment planner’s role may evolve. One of the common elements that impact the successful adoption of any innovation is “interconnection” as described by Euchner,²¹⁶ where it is argued that “*innovation happens when ideas collide*”. This is true for automated planning techniques where there may be mixed viewpoints within a hospital organisation regarding its advantage and disadvantage amongst the multidisciplinary staff groups, which presents a challenge for the successful clinical adoption. Is it an efficiency saving meaning potentially less dedicated staff are required and therefore, a mean for cost saving or is it an essential tool to be used by a finite staff resource? Is it an evolution in treatment planning requiring modified treatment planning skills or a route to deskilling staff? As such there are the “human factors” that need to be considered for implementing automated planning. There are known questions and challenges that need to be rationalised, such as: what will happen to the role of a treatment planner? Can one fully automate a plan and trust the software? Is it possible to automatically generate a plan for every patient? How does one ensure people retain their skills and is this even necessary? Is there a balanced approach to implementing automated planning? For example, it may be best practice to prioritise implementing automated planning algorithms to handle the most routine cases that have a high workload (such as prostate cancer or breast cancer) to re-route expertise to focus more efforts on tackling complex cases (which may otherwise be referred elsewhere) and be able to spend more time on

innovative new treatment techniques and research. Potentially, this could mean more difficult cases could be treated closer to a patient's home, assuming adequate case numbers are available to maintain clinical expertise. Sharing of DVH prediction models (RapidPlan), Treatment Techniques (Pinnacle Auto-Planning) and wish-lists (Erasmus-iCycle) between centres may also address this issue, particularly in rare diseases where adequate patient numbers to develop a model may only be possible across several centres. Alternatively, should such techniques be initially focussed on those sites where significant time savings may be made? Experienced treatment planners, in particular, may not see the benefits in the use of more routine planning where time savings are modest. As such, it can be argued that there should always be a role for a multidisciplinary team consideration. There may be a clash of cultures amongst different staff groups concerning automated planning, and this may require addressing before the adoption of such techniques into clinical practice.

Furthermore, the actual benchmarking and validation of automated planning is not straightforward and requires expert physics resource, as highlighted in the literature it is not merely possible to use "out-of-the-box".⁷⁸ It is important to realize that all current implementations of automated planning require a high-level of manual planning knowledge for configuration. It is questionable that this will change in the coming years as no mathematical formulas exist for balancing trade-offs, and dose-response relationships have significant uncertainties. The impact of a suboptimal configuration of an automated planning algorithm may be similar to an overall geometric error in the treatment preparation process: it will introduce an overall systematic error in the treatments, *i.e.* for all patients (of a particular tumour type), the plan quality is lower than feasible. Treatment planning can be a complicated process, and achieving the best radiotherapy treatment with the optimal trade-offs between good tumour coverage and healthy tissue sparing requires expert knowledge; replicating this with automated software is a difficult achievement and there will often be scepticism that the software will do as good a job as a human, regardless of how good the published results are.

Guidance by professional bodies on implementation of automated planning and possible redefining of treatment planner

roles could provide rationalisation. Algorithms for (semi-)automation of the configuration of automated planning software could also be highly beneficial in facilitating better implementation, and the development of simple metrics in plan comparison may facilitate this.

CONCLUSION

Recent innovations in automated treatment planning software have given rise to the potential to broadly improve the efficiency of radiotherapy planning and to enhance the overall quality of generated treatment plans, which is expected to result in higher quality treatments. Automated planning can also facilitate better access to high-quality advanced treatments, and harmonize radiotherapy treatments between treatment centres. Automated planning is now available in different commercial packages, each with a different technical approach, and the initial clinical reports are promising. The field is currently still rapidly developing and in a steep upward trend, and there are various areas of future research required that have been highlighted in this review. Much work is still needed to explore practical issues related to clinical implementation, including staffing, and their changing roles. In a resource-limited world, disruptive innovative technology is essential to meet future healthcare needs, and the rapid adoption of automated planning is one area that should be embraced and also possibly supported by professional body recommendations.

ACKNOWLEDGEMENTS

MH would like to acknowledge funding from the UK National Measurement System.

DISCLOSURES

MH has received speaker's honoraria from Varian Medical Systems and AN has received travel and speaker's honoraria from Varian Medical Systems. The Royal Surrey County Hospital has a research agreement with Varian Medical Systems. MH is funded by the UK National Measurement System. BJMH, and Erasmus MC Cancer Institute, has received research grants from Elekta AB, Stockholm, Sweden and also has a research collaboration with Accuray Inc, Sunnyvale, USA.

REFERENCES

1. Cancer Research, U. K., and ., and NHS England. *Vision for radiotherapy 2014-2024*. London Available at:https://www.cancerresearchuk.org/sites/default/files/policy_feb2014_radiotherapy_vision2014-2024_final.pdf (Accessed 02/07/2018). 2014.
2. Slotman BJ, Cottier B, Bentzen SM, Heeren G, Lievens Y, van den Bogaert W. Overview of national guidelines for infrastructure and staffing of radiotherapy. ESTRO-QUARTS: Work package 1. *Radiother Oncol* 2005; 75: 349.E1-349.E6. doi: <https://doi.org/10.1016/j.radonc.2004.12.005>
3. Sullivan R, Peppercorn J, Sikora K, Zalcborg J, Meropol NJ, Amir E, et al. Delivering affordable cancer care in high-income countries. *Lancet Oncol* 2011; 12: 933-80. doi: [https://doi.org/10.1016/S1470-2045\(11\)70141-3](https://doi.org/10.1016/S1470-2045(11)70141-3)
4. Teoh M, Clark CH, Wood K, Whitaker S, Nisbet A. Volumetric modulated arc therapy: a review of current literature and clinical use in practice. *Br J Radiol* 2011; 84: 967-96. doi: <https://doi.org/10.1259/bjr/22373346>
5. Ghilezan M, Yan D, Martinez A. Adaptive Radiation Therapy for Prostate Cancer. *Semin Radiat Oncol* 2010; 20: 130-7. doi: <https://doi.org/10.1016/j.semradonc.2009.11.007>
6. Castadot P, Lee JA, Geets X, Grégoire V. Adaptive radiotherapy of head and neck cancer. *Semin Radiat Oncol* 2010; 20: 84-93. doi: <https://doi.org/10.1016/j.semradonc.2009.11.002>

7. Schwartz DL, Garden AS, Shah SJ, Chronowski G, Sejal S, Rosenthal DI, et al. Adaptive radiotherapy for head and neck cancer—dosimetric results from a prospective clinical trial. *Radiother Oncol* 2013; **106**: 80–4. doi: <https://doi.org/10.1016/j.radonc.2012.10.010>
8. Caudell JJ, Torres-Roca JF, Gillies RJ, Enderling H, Kim S, Rishi A, et al. The future of personalised radiotherapy for head and neck cancer. *Lancet Oncol* 2017; **18**: e266–e273. doi: [https://doi.org/10.1016/S1470-2045\(17\)30252-8](https://doi.org/10.1016/S1470-2045(17)30252-8)
9. Hirst DG, Robson T. Molecular biology: the key to personalised treatment in radiation oncology? *Br J Radiol* 2010; **83**: 723–8. doi: <https://doi.org/10.1259/bjr/91488645>
10. Georg D, Thwaites D. Medical physics in radiation Oncology: New challenges, needs and roles. *Radiother Oncol* 2017; **125**: 375–8. doi: <https://doi.org/10.1016/j.radonc.2017.10.035>
11. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *J Clin Oncol* 2010; **28**: 2996–3001. doi: <https://doi.org/10.1200/JCO.2009.27.4498>
12. Rogers EM. *Diffusion of innovations*. 5th Edition. New York: The Free Press; 2003.
13. Petrovic S, Khussainova G, Jagannathan R. Knowledge-light adaptation approaches in case-based reasoning for radiotherapy treatment planning. *Artif Intell Med* 2016; **68**: 17–28. doi: <https://doi.org/10.1016/j.artmed.2016.01.006>
14. McIntosh C, Purdie TG. Contextual Atlas Regression Forests: Multiple-Atlas-Based Automated Dose Prediction in Radiation Therapy. *IEEE Trans Med Imaging* 2016; **35**: 1000–12. doi: <https://doi.org/10.1109/TMI.2015.2505188>
15. Skarpman Munter J, Sjölund J. Dose-volume histogram prediction using density estimation. *Phys Med Biol* 2015; **60**: 6923–36. doi: <https://doi.org/10.1088/0031-9155/60/17/6923>
16. Yang Y, Xing L. Clinical knowledge-based inverse treatment planning. *Phys Med Biol* 2004; **49**: 5101–17. doi: <https://doi.org/10.1088/0031-9155/49/22/006>
17. Yang Y, Ford EC, Wu B, Pinkawa M, van Triest B, Campbell P, et al. An overlap-volume-histogram based method for rectal dose prediction and automated treatment planning in the external beam prostate radiotherapy following hydrogel injection. *Med Phys* 2013; **40**: 011709. doi: <https://doi.org/10.1118/1.4769424>
18. Petit SF, van Elmpt W. Accurate prediction of target dose-escalation and organ-at-risk dose levels for non-small cell lung cancer patients. *Radiother Oncol* 2015; **117**: 453–8. doi: <https://doi.org/10.1016/j.radonc.2015.07.040>
19. Ahmed S, Nelms B, Gintz D, Caudell J, Zhang G, Moros EG, et al. A method for a priori estimation of best feasible DVH for organs-at-risk: Validation for head and neck VMAT planning. *Med Phys* 2017; **44**: 5486–97. doi: <https://doi.org/10.1002/mp.12500>
20. Chau RM, Leung SF, Kam MK, Cheung KY, Kwan WH, Yu KH, et al. A broadly adaptive array of dose-constraint templates for planning of intensity-modulated radiation therapy for advanced T-stage nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys* 2009; **74**: 21–. doi: <https://doi.org/10.1016/j.ijrobp.2008.07.041>
21. Sheng Y, Ge Y, Yuan L, Li T, Yin F-F, Wu QJ, . Outlier identification in radiation therapy knowledge-based planning: A study of pelvic cases. *Med Phys* 2017; **44**: 5617–26. doi: <https://doi.org/10.1002/mp.12556>
22. Schreiber E, Fox T. Prior-knowledge treatment planning for volumetric arc therapy using feature-based database mining. *J Appl Clin Med Phys* 2014; **15**: 19–27. doi: <https://doi.org/10.1120/jacmp.v15i2.4596>
23. Mayo CS, Yao J, Eisbruch A, Balter JM, Litzenberg DW, Matuszak MM, et al. Incorporating big data into treatment plan evaluation: Development of statistical DVH metrics and visualization dashboards. *Adv Radiat Oncol* 2017; **2**: 503–14. doi: <https://doi.org/10.1016/j.adro.2017.04.005>
24. Tran A, Woods K, Nguyen D, Yu VY, Niu T, Cao M, et al. Predicting liver SBRT eligibility and plan quality for VMAT and 4 π plans. *Radiat Oncol* 2017; **12**. doi: <https://doi.org/10.1186/s13014-017-0806-z>
25. Sheng Y, Li T, Zhang Y, Lee WR, Yin F-F, Ge Y, et al. Atlas-guided prostate intensity modulated radiation therapy (IMRT) planning. *Phys Med Biol* 2015; **60**: 7277–91. doi: <https://doi.org/10.1088/0031-9155/60/18/7277>
26. Zhang HH, Meyer RR, Shi L, D'Souza WD. The minimum knowledge base for predicting organ-at-risk dose–volume levels and plan-related complications in IMRT planning. *Phys Med Biol* 2010; **55**: 1935–47. doi: <https://doi.org/10.1088/0031-9155/55/7/010>
27. Zawadzka A, Nesteruk M, Brzozowska B, Kukołowicz PF. Method of predicting the mean lung dose based on a patient's anatomy and dose-volume histograms. *Med Dosim* 2017; **42**: 57–62. doi: <https://doi.org/10.1016/j.meddos.2016.12.001>
28. Zarepisheh M, Long T, Li N, Tian Z, Romeijn HE, Jia X, et al. A DVH-guided IMRT optimization algorithm for automatic treatment planning and adaptive radiotherapy replanning. *Med Phys* 2014; **41**(6Part1): 061711. doi: <https://doi.org/10.1118/1.4875700>
29. Zhu X, Ge Y, Li T, Thongphiew D, Yin F-F, Wu QJ, . A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. *Med Phys* 2011; **38**: 719–. doi: <https://doi.org/10.1118/1.3539749>
30. Good D, Lo J, Lee WR, Wu QJ, Yin F-F, Das SK. A knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers: an example application to prostate cancer planning. *Int J Radiat Oncol Biol Phys* 2013; **87**: 176–81. doi: <https://doi.org/10.1016/j.ijrobp.2013.03.015>
31. Lee T, Hammad M, Chan TC, Craig T, Sharpe MB. Predicting objective function weights from patient anatomy in prostate IMRT treatment planning. *Med Phys* 2013; **40**: 121706. doi: <https://doi.org/10.1118/1.4828841>
32. Nwankwo O, Sihono DSK, Schneider F, Wenz F. A global quality assurance system for personalized radiation therapy treatment planning for the prostate (or other sites). *Phys Med Biol* 2014; **59**: 5575–91. doi: <https://doi.org/10.1088/0031-9155/59/18/5575>
33. Lu R, Radke RJ, Hong L, Chui CS, Xiong J, Yorke E, et al. Learning the relationship between patient geometry and beam intensity in breast intensity-modulated radiotherapy. *IEEE Trans Biomed Eng* 2006; **53**: 908–20. doi: <https://doi.org/10.1109/TBME.2005.863987>
34. Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. *Med Phys* 2016; **43**: 378–. doi: <https://doi.org/10.1118/1.4938583>
35. Liu J, Wu QJ, Kirkpatrick JP, Yin F-F, Yuan L, Ge Y. From active shape model to active optical flow model: a shape-based approach to predicting voxel-level dose distributions in spine SBRT. *Phys Med Biol* 2015; **60**: N83–N92. doi: <https://doi.org/10.1088/0031-9155/60/5/N83>
36. Lin K-M, Simpson J, Sasso G, Raith A, Ehrhott M. Quality assessment for VMAT prostate radiotherapy planning based on data envelopment analysis. *Phys Med Biol* 2013; **58**: 5753–69. doi: <https://doi.org/10.1088/0031-9155/58/16/5753>

37. Campbell WG, Miften M, Olsen L, Stumpf P, Scheffer T, Goodman KA, et al. Neural network dose models for knowledge-based planning in pancreatic SBRT. *Med Phys* 2017; **44**: 6148–58. doi: <https://doi.org/10.1002/mp.12621>
38. McIntosh C, Purdie TG. Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning. *Phys Med Biol* 2017; **62**: 415–31. doi: <https://doi.org/10.1088/1361-6560/62/2/415>
39. McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. *Phys Med Biol* 2017; **62**: 5926–44. doi: <https://doi.org/10.1088/1361-6560/aa71f8>
40. Valdes G, Simone CB, Chen J, Lin A, Yom SS, Pattison AJ, et al. Clinical decision support of radiotherapy treatment planning: a data-driven machine learning strategy for patient-specific dosimetric decision making. *Radiother Oncol* 2017; **125**: 392–7. doi: <https://doi.org/10.1016/j.radonc.2017.10.014>
41. Wang H, Dong P, Liu H, Xing L. Development of an autonomous treatment planning strategy for radiation therapy with effective use of population-based prior data. *Med Phys* 2017; **44**: 389–96. doi: <https://doi.org/10.1002/mp.12058>
42. Zhang Y, Li T, Xiao H, Ji W, Guo M, Zeng Z, et al. A knowledge-based approach to automated planning for hepatocellular carcinoma. *J Appl Clin Med Phys* 2018; **19**: 50–59. doi: <https://doi.org/10.1002/acm2.12219>
43. Ueda Y, Fukunaga JI, Kamima T, Adachi Y, Nakamatsu K, Monzen H. Evaluation of multiple institutions' models for knowledge-based planning of volumetric modulated arc therapy (VMAT) for prostate cancer. *Radiat Oncol* 2018; **13**: 46. doi: <https://doi.org/10.1186/s13014-018-0994-1>
44. Yu G, Li Y, Feng Z, Tao C, Yu Z, Li B, et al. Knowledge-based IMRT planning for individual liver cancer patients using a novel specific model. *Radiat Oncol* 2018; **13**: 52. doi: <https://doi.org/10.1186/s13014-018-0996-z>
45. Stanhope C, Wu QJ, Yuan L, Liu J, Hood R, Yin FF, et al. Utilizing knowledge from prior plans in the evaluation of quality assurance. *Phys Med Biol* 2015; **60**: 4873–. doi: <https://doi.org/10.1088/0031-9155/60/12/4873>
46. Wu B, Pang D, Simari P, Taylor R, Sanguineti G, McNutt T. Using overlap volume histogram and IMRT plan data to guide and automate VMAT planning: a head-and-neck case study. *Med Phys* 2013; **40**: 021714. doi: <https://doi.org/10.1118/1.4788671>
47. Liu ESF, Wu VWC, Harris B, Lehman M, Pryor D, Chan LWC. Vector-model-supported approach in prostate plan optimization. *Med Dosim* 2017; **42**: 79–84. doi: <https://doi.org/10.1016/j.meddos.2017.01.001>
48. Caine H, Whalley D, Kneebone A, McCloud P, Eade T. Using individual patient anatomy to predict protocol compliance for prostate intensity-modulated radiotherapy. *Med Dosim* 2016; **41**: 70–4. doi: <https://doi.org/10.1016/j.meddos.2015.08.005>
49. Moore KL, Schmidt R, Moiseenko V, Olsen LA, Tan J, Xiao Y, et al. Quantifying unnecessary normal tissue complication risks due to suboptimal planning: A secondary study of RTOG 0126. *Int J Radiat Oncol Biol Phys* 2015; **92**: 228–. doi: <https://doi.org/10.1016/j.ijrobp.2015.01.046>
50. Hall DC, Trofimov AV, Winey BA, Liebsch NJ, Paganetti H. Predicting patient-specific dosimetric benefits of proton therapy for skull-base tumors using a geometric knowledge-based method. *Int J Radiat Oncol Biol Phys* 2017; **97**: 1087–. doi: <https://doi.org/10.1016/j.ijrobp.2017.01.236>
51. Jiang F, Wu H, Yue H, Jia F, Zhang Y. Photon Optimizer (PO) prevails over Progressive Resolution Optimizer (PRO) for VMAT planning with or without knowledge-based solution. *J Appl Clin Med Phys* 2017; **18**: 9–14. doi: <https://doi.org/10.1002/acm2.12038>
52. Chatterjee A, Serban M, Abdulkarim B, Panet-Raymond V, Souhami L, Shenouda G, et al. Performance of knowledge-based radiation therapy planning for the glioblastoma disease site. *Int J Radiat Oncol Biol Phys* 2017; **99**: 1021–. doi: <https://doi.org/10.1016/j.ijrobp.2017.07.012>
53. Cagni E, Botti A, Micera R, Galeandro M, Sghedoni R, Orlandi M, et al. Knowledge-based treatment planning: an inter-technique and inter-system feasibility study for prostate cancer. *Phys Medica* 2017; **36**: 38–45. doi: <https://doi.org/10.1016/j.ejmp.2017.03.002>
54. Li N, Carmona R, Sirak I, Kasaova L, Followill D, Michalski J, et al. Highly efficient training, refinement, and validation of a knowledge-based planning quality-control system for radiation therapy clinical trials. *Int J Radiat Oncol Biol Phys* 2017; **97**: 164–. doi: <https://doi.org/10.1016/j.ijrobp.2016.10.005>
55. Kubo K, Monzen H, Ishii K, Tamura M, Kawamorita R, Sumida I, et al. Dosimetric comparison of RapidPlan and manually optimized plans in volumetric modulated arc therapy for prostate cancer. *Phys Medica* 2017; **44**: 199–204. doi: <https://doi.org/10.1016/j.ejmp.2017.06.026>
56. Wu H, Jiang F, Yue H, Li S, Zhang Y. A dosimetric evaluation of knowledge-based VMAT planning with simultaneous integrated boosting for rectal cancer patients. *J Appl Clin Med Phys* 2016; **17**: 78–85. doi: <https://doi.org/10.1120/jacmp.v17i6.6410>
57. Delaney AR, Tol JP, Dahele M, Cuijpers J, Slotman BJ, Verbakel WFAR. Effect of Dosimetric Outliers on the Performance of a Commercial Knowledge-Based Planning Solution. *Int J Radiat Oncol Biol Phys* 2016; **94**: 469–77. doi: <https://doi.org/10.1016/j.ijrobp.2015.11.011>
58. Chanyavanich V, Das SK, Lee WR, Lo JY, . Knowledge-based IMRT treatment planning for prostate cancer. *Med Phys* 2011; **38**: 2515–. doi: <https://doi.org/10.1118/1.3574874>
59. Tol JP, Dahele M, Delaney AR, Slotman BJ, Verbakel WF. Can knowledge-based DVH predictions be used for automated, individualized quality assurance of radiotherapy treatment plans? *Radiat Oncol* 2015; **10**: 234. doi: <https://doi.org/10.1186/s13014-015-0542-1>
60. ESF L, VWC W, Harris B, Foote M, Lehman M, Chan LWC. Vector-model-supported optimization in volumetric-modulated arc stereotactic radiotherapy planning for brain metastasis. *Med Dosim* 2017; **42**: 85–9.
61. Habraken SJM, Sharfo AWM, Buijssen J, Verbakel WFAR, Haasbeek CJA, Öllers MC, et al. The TRENDY multi-center randomized trial on hepatocellular carcinoma – Trial QA including automated treatment planning and benchmark-case results. *Radiother Oncol* 2017; **125**: 507–13. doi: <https://doi.org/10.1016/j.radonc.2017.09.007>
62. Yang Y, Li T, Yuan L, Ge Y, Yin F-F, Lee WR, et al. Quantitative comparison of automatic and manual IMRT optimization for prostate cancer: the benefits of DVH prediction. *J Appl Clin Med Phys* 2015; **16**: 241–50. doi: <https://doi.org/10.1120/jacmp.v16i2.5204>
63. Foy JJ, Marsh R, Ten Haken RK, Younge KC, Schipper M, Sun Y, et al. An analysis of knowledge-based planning for stereotactic body radiation therapy of the spine. *Pract Radiat Oncol* 2017; **7**: e355–e360. doi: <https://doi.org/10.1016/j.prro.2017.02.007>
64. Delaney AR, Dahele M, Tol JP, Cuijpers IT, Slotman BJ, Verbakel WFAR. Using a knowledge-based planning solution to select patients for proton therapy. *Radiother Oncol* 2017; **124**: 263–70. doi: <https://doi.org/10.1016/j.radonc.2017.03.020>

65. Wang J, Hu W, Yang Z, Chen X, Wu Z, Yu X, et al. Is it possible for knowledge-based planning to improve intensity modulated radiation therapy plan quality for planners with different planning experiences in left-sided breast cancer patients? *Radiat Oncol* 2017; **12**: 85. doi: <https://doi.org/10.1186/s13014-017-0822-z>
66. Schubert C, Waletzko O, Weiss C, Voelzke D, Toperim S, Roeser A, et al. Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium. *PLoS One* 2017; **12**: e0178034. doi: <https://doi.org/10.1371/journal.pone.0178034>
67. Ziemer BP, Shiraiishi S, Hattangadi-Gluth JA, Sanghvi P, Moore KL, automated F. Fully automated, comprehensive knowledge-based planning for stereotactic radiosurgery: Preclinical validation through blinded physician review. *Pract Radiat Oncol* 2017; **7**: e569–e578. doi: <https://doi.org/10.1016/j.prro.2017.04.011>
68. Wu B, Kusters M, Kunze-busch M, Dijkema T, McNutt T, Sanguineti G, et al. Cross-institutional knowledge-based planning (KBP) implementation and its performance comparison to Auto-Planning Engine (APE). *Radiother Oncol* 2017; **123**: 57–62. doi: <https://doi.org/10.1016/j.radonc.2017.01.012>
69. Long T, Chen M, Jiang S, Lu W. Threshold-driven optimization for reference-based auto-planning. *Phys Med Biol* 2018; **63**: 04NT01. doi: <https://doi.org/10.1088/1361-6560/aaa731>
70. Fogliata A, Wang P-M, Belosi F, Clivio A, Nicolini G, Vanetti E, et al. Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. *Radiat Oncol* 2014; **9**: 236. doi: <https://doi.org/10.1186/s13014-014-0236-0>
71. Wang Y, Heijmen BJM, Petit SF. Prospective clinical validation of independent DVH prediction for plan QA in automatic treatment planning for prostate cancer patients. *Radiother Oncol* 2017; **125**: 500–6. doi: <https://doi.org/10.1016/j.radonc.2017.09.021>
72. Fogliata A, Reggiori G, Stravato A, Lobefalo F, Franzese C, Franceschini D, et al. RapidPlan head and neck model: the objectives and possible clinical benefit. *Radiat Oncol* 2017; **12**. doi: <https://doi.org/10.1186/s13014-017-0808-x>
73. Boutillier JJ, Craig T, Sharpe MB, Chan TCY. Sample size requirements for knowledge-based treatment planning. *Med Phys* 2016; **43**: 1212–21. doi: <https://doi.org/10.1118/1.4941363>
74. Wu H, Jiang F, Yue H, Zhang H, Wang K, Zhang Y. Applying a RapidPlan model trained on a technique and orientation to another: a feasibility and dosimetric evaluation. *Radiat Oncol* 2016; **11**. doi: <https://doi.org/10.1186/s13014-016-0684-9>
75. Delaney AR, Dahele M, Tol JP, Slotman BJ, Verbakel WFAR. Knowledge-based planning for stereotactic radiotherapy of peripheral early-stage lung cancer. *Acta Oncol* 2017; **56**: 490–5. doi: <https://doi.org/10.1080/0284186X.2016.1273544>
76. Fogliata A, Nicolini G, Clivio A, Vanetti E, Laksar S, Tozzi A, et al. A broad scope knowledge based model for optimization of VMAT in esophageal cancer: validation and assessment of plan quality among different treatment centers. *Radiat Oncol* 2015; **10**. doi: <https://doi.org/10.1186/s13014-015-0530-5>
77. Berry SL, Ma R, Boczkowski A, Jackson A, Zhang P, Hunt M. Evaluating inter-campus plan consistency using a knowledge based planning model. *Radiother Oncol* 2016; **120**: 349–55. doi: <https://doi.org/10.1016/j.radonc.2016.06.010>
78. Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. *Radiother Oncol* 2016; **120**: 473–9. doi: <https://doi.org/10.1016/j.radonc.2016.06.022>
79. Kraysenbuehl J, Norton I, Studer G, Guckenberger M. Evaluation of an automated knowledge based treatment planning system for head and neck. *Radiat Oncol* 2015; **10**: 226. doi: <https://doi.org/10.1186/s13014-015-0533-2>
80. Wall PDH, Carver RL, Fontenot JD. An improved distance-to-dose correlation for predicting bladder and rectum dose-volumes in knowledge-based VMAT planning for prostate cancer. *Phys Med Biol* 2018; **63**: 015035. doi: <https://doi.org/10.1088/1361-6560/aa9a30>
81. Nwankwo O, Mekdash H, Sihono DSK, Wenz F, Glatting G. Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning. *Radiat Oncol* 2015; **10**: 111. doi: <https://doi.org/10.1186/s13014-015-0416-6>
82. Tol JP, Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Evaluation of a Knowledge-Based Planning Solution for Head and Neck Cancer. *Int J Radiat Oncol Biol Phys* 2015; **91**: 612–20. doi: <https://doi.org/10.1016/j.ijrobp.2014.11.014>
83. Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, et al. On the pre-clinical validation of a commercial model-based optimisation engine: Application to volumetric modulated arc therapy for patients with lung or prostate cancer. *Radiother Oncol* 2014; **113**: 385–91. doi: <https://doi.org/10.1016/j.radonc.2014.11.009>
84. Boutillier JJ, Lee T, Craig T, Sharpe MB, Chan TC. Models for predicting objective function weights in prostate cancer IMRT. *Med Phys* 2015; **42**: 1586–. doi: <https://doi.org/10.1118/1.4914140>
85. Kuo L, Yorke ED, Dumane VA, Foster A, Zhang Z, Mechalakos JG, et al. Geometric dose prediction model for hemithoracic intensity-modulated radiation therapy in mesothelioma patients with two intact lungs. *J Appl Clin Med Phys* 2016; **17**: 371–9. doi: <https://doi.org/10.1120/jacmp.v17i3.6199>
86. Boylan C, Rowbottom C. A bias-free, automated planning tool for technique comparison in radiotherapy - application to nasopharyngeal carcinoma treatments. *J Appl Clin Med Phys* 2014; **15**: 213–25. doi: <https://doi.org/10.1120/jacmp.v15i1.4530>
87. Kraysenbuehl J, Di Martino M, Guckenberger M, Andratschke N. Improved plan quality with automated radiotherapy planning for whole brain with hippocampus sparing: a comparison to the RTOG 0933 trial. *Radiat Oncol* 2017; **12**: 1–7.
88. Zhang X, Li X, Quan EM, Pan X, Li Y. A methodology for automatic intensity-modulated radiation treatment planning for lung cancer. *Phys Med Biol* 2011; **56**: 3873–93. doi: <https://doi.org/10.1088/0031-9155/56/13/009>
89. Hansen CR, Bertelsen A, Hazell I, Zukauskaite R, Gyldenkerne N, Johansen J, et al. Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans. *Clin Transl Radiat Oncol* 2016; **1**: 2–8. doi: <https://doi.org/10.1016/j.ctro.2016.08.001>
90. Wang W, Purdie TG, Rahman M, Marshall A, Liu FF, Fyles A. Rapid automated treatment planning process to select breast cancer patients for active breathing control to achieve cardiac dose reduction. *Int J Radiat Oncol Biol Phys* 2012; **82**: 386–. doi: <https://doi.org/10.1016/j.ijrobp.2010.09.026>
91. Nawa K, Haga A, Nomoto A, Sarmiento RA, Shiraiishi K, Yamashita H, et al. Evaluation of a commercial automatic treatment planning system for prostate cancers. *Med Dosim* 2017; **42**: 203–9. doi: <https://doi.org/10.1016/j.meddos.2017.03.004>
92. Li X, Wang L, Wang J, Han X, Xia B, Wu S, et al. Dosimetric benefits of automation in

- the treatment of lower thoracic esophageal cancer: Is manual planning still an alternative option? *Med Dosim* 2017; **42**: 289–95. doi: <https://doi.org/10.1016/j.meddos.2017.06.004>
93. Song Y, Wang Q, Jiang X, Liu S, Zhang Y, Bai S. Fully automatic volumetric modulated arc therapy plan generation for rectal cancer. *Radiother Oncol* 2016; **119**: 531–6. doi: <https://doi.org/10.1016/j.radonc.2016.04.010>
 94. Tol JP, Dahele M, Delaney AR, Doornaert P, Slotman BJ, Verbakel WFAR. Detailed evaluation of an automated approach to interactive optimization for volumetric modulated arc therapy plans. *Med Phys* 2016; **43**: 1818–28. doi: <https://doi.org/10.1118/1.4944063>
 95. Hansen CR, Nielsen M, Bertelsen AS, Hazell I, Holtved E, Zukauskaitė R, et al. Automatic treatment planning facilitates fast generation of high-quality treatment plans for esophageal cancer. *Acta Oncol* 2017; **56**: 1495–500. doi: <https://doi.org/10.1080/0284186X.2017.1349928>
 96. Purdie TG, Dinniwell RE, Fyles A, Sharpe MB. Automation and intensity modulated radiation therapy for individualized high-quality tangent breast treatment plans. *Int J Radiat Oncol Biol Phys* 2014; **90**: 688–95. doi: <https://doi.org/10.1016/j.ijrobp.2014.06.056>
 97. Gintz D, Latifi K, Caudell J, Nelms B, Zhang G, Moros E, et al. Initial evaluation of automated treatment planning software. *J Appl Clin Med Phys* 2016; **17**: 331–46. doi: <https://doi.org/10.1120/jacmp.v17i3.6167>
 98. Yan H, Yin F-F, Willett C. Evaluation of an artificial intelligence guided inverse planning system: Clinical case study. *Radiother Oncol* 2007; **83**: 76–85. doi: <https://doi.org/10.1016/j.radonc.2007.02.013>
 99. Wang S, Zheng D, Zhang C, Ma R, Bennion NR, Lei Y, et al. Automatic planning on hippocampal avoidance whole-brain radiotherapy. *Med Dosim* 2017; **42**: 63–8. doi: <https://doi.org/10.1016/j.meddos.2016.12.002>
 100. Speer S, Klein A, Kober L, Weiss A, Johannes I, Bert C. Automation of radiation treatment planning: evaluation of head and neck cancer patient plans created by the Pinnacle3 scripting and Auto-Planning functions. *Strahlentherapie Und Onkol* 2017; **193**: 656–65.
 101. Kusters JMAM, Bzdusek K, Kumar P, van Kollenburg PGM, Kunze-Busch MC, Wendling M, et al. Automated IMRT planning in Pinnacle. *Strahlentherapie Und Onkol* 2017; **193**: 1031–8. doi: <https://doi.org/10.1007/s00066-017-1187-9>
 102. Winkel D, Bol GH, van Asselen B, Hes J, Scholten V, Kerkmeijer LGW, et al. Development and clinical introduction of automated radiotherapy treatment planning for prostate cancer. *Phys Med Biol* 2016; **61**: 8587–95. doi: <https://doi.org/10.1088/1361-6560/61/24/8587>
 103. Hazell I, Bzdusek K, Kumar P, Hansen CR, Bertelsen A, Eriksen JG, et al. Automatic planning of head and neck treatment plans. *J Appl Clin Med Phys* 2016; **17**: 272–82. doi: <https://doi.org/10.1120/jacmp.v17i1.5901>
 104. Yan H, Yin F-F, Guan H, Kim JH. Fuzzy logic guided inverse treatment planning. *Med Phys* 2003; **30**: 2675–85. doi: <https://doi.org/10.1118/1.1600739>
 105. Yan H, Yin F-F, Guan H-qun, Kim JH. AI-guided parameter optimization in inverse treatment planning. *Phys Med Biol* 2003; **48**: 3565–80. doi: <https://doi.org/10.1088/0031-9155/48/21/008>
 106. Xhaferllari I, Wong E, Bzdusek K, Lock M, Chen JZ. Automated IMRT planning with regional optimization using planning scripts. *J Appl Clin Med Phys* 2013; **14**: 176–91. doi: <https://doi.org/10.1120/jacmp.v14i1.4052>
 107. Tol JP, Dahele M, Peltola J, Nord J, Slotman BJ, Verbakel WFAR. Automatic interactive optimization for volumetric modulated arc therapy planning. *Radiat Oncol* 2015; **10**. doi: <https://doi.org/10.1186/s13014-015-0388-6>
 108. Wilkens JJ, Alaly JR, Zakarian K, Thorstad WL, Deasy JO. IMRT treatment planning based on prioritizing prescription goals. *Phys Med Biol* 2007; **52**: 1675–92. doi: <https://doi.org/10.1088/0031-9155/52/6/009>
 109. Stieler F, Yan H, Lohr F, Wenz F, Yin F-F. Development of a neuro-fuzzy technique for automated parameter optimization of inverse treatment planning. *Radiat Oncol* 2009; **4**: 39. doi: <https://doi.org/10.1186/1748-717X-4-39>
 110. Wang H, Xing L. Application programming in C# environment with recorded user software interactions and its application in autopilot of VMAT/IMRT treatment planning. *J Appl Clin Med Phys* 2016; **17**: 189–203. doi: <https://doi.org/10.1120/jacmp.v17i6.6425>
 111. Bokrantz R, Miettinen K. Projections onto the Pareto surface in multicriteria radiation therapy optimization. *Med Phys* 2015; **42**: 5862–. doi: <https://doi.org/10.1118/1.4930252>
 112. Schlaefler A, Violet T, Muacevic A, Fürweger C. Multicriteria optimization of the spatial dose distribution. *Med Phys* 2013; **40**: 121720. doi: <https://doi.org/10.1118/1.4828840>
 113. Pardo-Montero J, Fenwick JD. An approach to multiobjective optimization of rotational therapy. II. Pareto optimal surfaces and linear combinations of modulated blocked arcs for a prostate geometry. *Med Phys* 2010; **37**: 2606–. doi: <https://doi.org/10.1118/1.3427410>
 114. Craft D, Halabi T, Shih HA, Bortfeld T. An approach for practical multiobjective IMRT treatment planning. *Int J Radiat Oncol Biol Phys* 2007; **69**: 1600–. doi: <https://doi.org/10.1016/j.ijrobp.2007.08.019>
 115. Zhang X, Wang X, Dong L, Liu H, Mohan R. A sensitivity-guided algorithm for automated determination of IMRT objective function parameters. *Med Phys* 2006; **33**: 2935–. doi: <https://doi.org/10.1118/1.2214171>
 116. Zarepisheh M, Uribe-Sanchez AF, Li N, Jia X, Jiang SB. A multicriteria framework with voxel-dependent parameters for radiotherapy treatment plan optimization. *Med Phys* 2014; **41**: 041705. doi: <https://doi.org/10.1118/1.4866886>
 117. Holdsworth C, Kim M, Liao J, Phillips MH. A hierarchical evolutionary algorithm for multiobjective optimization in IMRT. *Med Phys* 2010; **37**: 4986–. doi: <https://doi.org/10.1118/1.3478276>
 118. Chen W, Craft D, Madden TM, Zhang K, Kooy HM, Herman GT. A fast optimization algorithm for multicriteria intensity modulated proton therapy planning. *Med Phys* 2010; **37**: 4938–. doi: <https://doi.org/10.1118/1.3481566>
 119. Salari E, Unkelbach J. A column-generation-based method for multi-criteria direct aperture optimization. *Phys Med Biol* 2013; **58**: 621–39. doi: <https://doi.org/10.1088/0031-9155/58/3/621>
 120. van de Water S, Kraan AC, Breedveld S, Schillemans W, Teguh DN, Kooy HM, et al. Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams. *Phys Med Biol* 2013; **58**: 6969–83. doi: <https://doi.org/10.1088/0031-9155/58/19/6969>
 121. Lahanas M, Schreiber E, Baltas D. Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. *Phys Med Biol* 2003; **48**: 2843–71. doi: <https://doi.org/10.1088/0031-9155/48/17/308>
 122. Salari E, Wala J, Craft D. Exploring trade-offs between VMAT dose quality and delivery efficiency using a network optimization approach. *Phys Med Biol* 2012; **57**: 5587–600. doi: <https://doi.org/10.1088/0031-9155/57/17/5587>

123. Bokrantz R. Multicriteria optimization for volumetric-modulated arc therapy by decomposition into a fluence-based relaxation and a segment weight-based restriction. *Med Phys* 2012; **39**: 6712–. doi: <https://doi.org/10.1118/1.4754652>
124. Monshouwer R, Hoffmann AL, Kunze-Busch M, Bussink J, Kaanders JHAM, Huizenga H. A practical approach to assess clinical planning tradeoffs in the design of individualized IMRT treatment plans. *Radiother Oncol* 2010; **97**: 561–6. doi: <https://doi.org/10.1016/j.radonc.2010.10.019>
125. Redondo JL, Fernández J, Ortigosa PM. FEMOEA: a fast and efficient multi-objective evolutionary algorithm. *Math Methods Oper Res* 2017; **85**: 113–35. doi: <https://doi.org/10.1007/s00186-016-0560-2>
126. Clark VH, Chen Y, Wilkens J, Alaly JR, Zakaryan K, Deasy JO. IMRT treatment planning for prostate cancer using prioritized prescription optimization and mean-tail-dose functions. *Linear Algebra Appl* 2008; **428**: 1345–64. doi: <https://doi.org/10.1016/j.laa.2007.07.026>
127. Teichert K, Süß P, Serna JI, Monz M, Küfer KH, Thieke C. Comparative analysis of Pareto surfaces in multi-criteria IMRT planning. *Phys Med Biol* 2011; **56**: 3669–84. doi: <https://doi.org/10.1088/0031-9155/56/12/014>
128. Craft D, Bortfeld T. How many plans are needed in an IMRT multi-objective plan database? *Phys Med Biol* 2008; **53**: 2785–96. doi: <https://doi.org/10.1088/0031-9155/53/11/002>
129. Voet PWJ, Breedveld S, Dirckx MLP, Levendag PC, Heijmen BJM. Integrated multicriterial optimization of beam angles and intensity profiles for coplanar and noncoplanar head and neck IMRT and implications for VMAT. *Med Phys* 2012; **39**: 4858–65. doi: <https://doi.org/10.1118/1.4736803>
130. Potrebko PS, Fiege J, Biagioli M, Poleszczuk J. Investigating multi-objective fluence and beam orientation IMRT optimization. *Phys Med Biol* 2017; **62**: 5228–44. doi: <https://doi.org/10.1088/1361-6560/aa7298>
131. Monz M, Küfer KH, Bortfeld TR, Thieke C. Pareto navigation—algorithmic foundation of interactive multi-criteria IMRT planning. *Phys Med Biol* 2008; **53**: 985–98. doi: <https://doi.org/10.1088/0031-9155/53/4/011>
132. van Haveren R, Ogryczak W, Verduijn GM, Keijzer M, Heijmen BJM, Breedveld S. Fast and fuzzy multi-objective radiotherapy treatment plan generation for head and neck cancer patients with the lexicographic reference point method (LRPM). *Phys Med Biol* 2017; **62**: 4318–32. doi: <https://doi.org/10.1088/1361-6560/62/11/4318>
133. Fredriksson A, Bokrantz R. Deliverable navigation for multicriteria IMRT treatment planning by combining shared and individual apertures. *Phys Med Biol* 2013; **58**: 7683–97. doi: <https://doi.org/10.1088/0031-9155/58/21/7683>
134. Jee KW, McShan DL, Fraass BA. Lexicographic ordering: intuitive multicriteria optimization for IMRT. *Phys Med Biol* 2007; **52**: 1845–. doi: <https://doi.org/10.1088/0031-9155/52/7/006>
135. Breedveld S, Storchi PRM, Heijmen BJM. The equivalence of multi-criteria methods for radiotherapy plan optimization. *Phys Med Biol* 2009; **54**: 7199–209. doi: <https://doi.org/10.1088/0031-9155/54/23/011>
136. Wang Y, Breedveld S, Heijmen B, Petit SF. Evaluation of plan quality assurance models for prostate cancer patients based on fully automatically generated Pareto-optimal treatment plans. *Phys Med Biol* 2016; **61**: 4268–82. doi: <https://doi.org/10.1088/0031-9155/61/11/4268>
137. Bokrantz R. Distributed approximation of Pareto surfaces in multicriteria radiation therapy treatment planning. *Phys Med Biol* 2013; **58**: 3501–16. doi: <https://doi.org/10.1088/0031-9155/58/11/3501>
138. Halabi T, Craft D, Bortfeld T. Dose–volume objectives in multi-criteria optimization. *Phys Med Biol* 2006; **51**: 3809–18. doi: <https://doi.org/10.1088/0031-9155/51/15/014>
139. Holdsworth CH, Corwin D, Stewart RD, Rockne R, Trister AD, Swanson KR, et al. Adaptive IMRT using a multiobjective evolutionary algorithm integrated with a diffusion–invasion model of glioblastoma. *Phys Med Biol* 2012; **57**: 8271–83. doi: <https://doi.org/10.1088/0031-9155/57/24/8271>
140. Thieke C, Küfer K-H, Monz M, Scherrer A, Alonso F, Oelfke U, et al. A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation. *Radiother Oncol* 2007; **85**: 292–8. doi: <https://doi.org/10.1016/j.radonc.2007.06.020>
141. Romeijn HE, Dempsey JF, Li JG, . A unifying framework for multi-criteria fluence map optimization models. *Phys Med Biol* 2004; **49**: 1991–2013. doi: <https://doi.org/10.1088/0031-9155/49/10/011>
142. Chen W, Unkelbach J, Trofimov A, Madden T, Kooy H, Bortfeld T, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Phys Med Biol* 2012; **57**: 591–608. doi: <https://doi.org/10.1088/0031-9155/57/3/591>
143. Breedveld S, Storchi PRM, Keijzer M, Heemink AW, Heijmen BJM. A novel approach to multi-criteria inverse planning for IMRT. *Phys Med Biol* 2007; **52**: 6339–53. doi: <https://doi.org/10.1088/0031-9155/52/20/016>
144. Serna JI, Monz M, Küfer KH, Thieke C. Trade-off bounds for the Pareto surface approximation in multi-criteria IMRT planning. *Phys Med Biol* 2009; **54**: 6299–311. doi: <https://doi.org/10.1088/0031-9155/54/20/018>
145. Hu W, Wang J, Li G, Peng J, Lu S, Zhang Z. Investigation of plan quality between RapidArc and IMRT for gastric cancer based on a novel beam angle and multicriteria optimization technique. *Radiother Oncol* 2014; **111**: 144–7. doi: <https://doi.org/10.1016/j.radonc.2014.01.024>
146. Breedveld S, Storchi PRM, Voet PWJ, Heijmen BJM. iCycle: integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. *Med Phys* 2012; **39**: 951–63. doi: <https://doi.org/10.1118/1.3676689>
147. Craft D, McQuaid D, Wala J, Chen W, Salari E, Bortfeld T. Multicriteria VMAT optimization. *Med Phys* 2012; **39**: 686–96. doi: <https://doi.org/10.1118/1.3675601>
148. Craft D, Monz M. Simultaneous navigation of multiple Pareto surfaces, with an application to multicriteria IMRT planning with multiple beam angle configurations. *Med Phys* 2010; **37**: 736–. doi: <https://doi.org/10.1118/1.3292636>
149. Hoffmann AL, Siem AY, den Hertog D, Kaanders JH, Huizenga H. Derivative-free generation and interpolation of convex Pareto optimal IMRT plans. *Phys Med Biol* 2006; **51**: 6349–. doi: <https://doi.org/10.1088/0031-9155/51/24/005>
150. Buschmann M, Sharfo AWM, Penninkhof J, Seppenwoolde Y, Goldner G, Georg D, et al. Automated volumetric modulated arc therapy planning for whole pelvic prostate radiotherapy. *Strahlenther Onkol* 2018; **194**: 333–342. doi: <https://doi.org/10.1007/s00066-017-1246-2>
151. Sharfo AWM, Stieler F, Kupfer O, Heijmen BJM, Dirckx MLP, Breedveld S, et al. Automated VMAT planning for postoperative adjuvant treatment of advanced gastric cancer. *Radiat Oncol* 2018; **13**: 74. doi: <https://doi.org/10.1186/s13014-018-1032-z>
152. Janssen T, van Kesteren Z, Franssen G, Damen E, van Vliet C. Pareto fronts in clinical practice for pinnacle. *Int J Radiat Oncol Biol Phys* 2013; **85**: 873–. doi: <https://doi.org/10.1016/j.ijrobp.2012.05.045>

153. Ghandour S, Cosinschi A, Mazouni Z, Pachoud M, Matzinger O. Optimization of stereotactic body radiotherapy treatment planning using a multicriteria optimization algorithm. *Z Med Phys* 2016; **26**: 362–70. doi: <https://doi.org/10.1016/j.zemedi.2016.04.001>
154. Hong TS, Craft DL, Carlsson F, Bortfeld TR. Multicriteria optimization in intensity-modulated radiation therapy treatment planning for locally advanced cancer of the pancreatic head. *Int J Radiat Oncol Biol Phys* 2008; **72**: 1208–. doi: <https://doi.org/10.1016/j.ijrobp.2008.07.015>
155. Holdsworth C, Stewart RD, Kim M, Liao J, Phillips MH. Investigation of effective decision criteria for multiobjective optimization in IMRT. *Med Phys* 2011; **38**: 2964–. doi: <https://doi.org/10.1118/1.3589128>
156. Kierkels RG, Korevaar EW, Steenbakkers RJ, Janssen T, van't Veld AA, Langendijk JA, et al. Direct use of multivariable normal tissue complication probability models in treatment plan optimisation for individualised head and neck cancer radiotherapy produces clinically acceptable treatment plans. *Radiation Oncol* 2014; **11**: 430–6. doi: <https://doi.org/10.1016/j.radonc.2014.08.020>
157. Chen H, Winey BA, Daartz J, Oh KS, Shin JH, Gierga DP. Efficiency gains for spinal radiosurgery using multicriteria optimization intensity modulated radiation therapy guided volumetric modulated arc therapy planning. *Pract Radiat Oncol* 2015; **5**: 49–. doi: <https://doi.org/10.1016/j.prro.2014.04.003>
158. Song T, Li N, Zarepisheh M, Li Y, Gautier Q, Zhou L, et al. An automated treatment plan quality control tool for intensity-modulated radiation therapy using a voxel-weighting factor-based re-optimization algorithm. *PLoS One* 2016; **11**: e0149273. doi: <https://doi.org/10.1371/journal.pone.0149273>
159. Buschmann M, Seppenwoolde Y, Wiezorek T, Weibert K, Georg D. Advanced optimization methods for whole pelvic and local prostate external beam therapy. *Phys Medica* 2016; **32**: 465–73. doi: <https://doi.org/10.1016/j.ejmp.2016.03.002>
160. Leinders SM, Breedveld S, Méndez Romero A, Schaart D, Seppenwoolde Y, Heijmen BJ. Adaptive liver stereotactic body radiation therapy: automated daily plan reoptimization prevents dose delivery degradation caused by anatomy deformations. *Int J Radiat Oncol Biol Phys* 2013; **87**: 1016–. doi: <https://doi.org/10.1016/j.ijrobp.2013.08.009>
161. Sharfo AWM, Dirkx MLP, Breedveld S, Romero AM, Heijmen BJM. VMAT plus a few computer-optimized non-coplanar IMRT beams (VMAT+) tested for liver SBRT. *Radiation Oncol* 2017; **12**: 49–56. doi: <https://doi.org/10.1016/j.radonc.2017.02.018>
162. Voet PWJ, Dirkx MLP, Breedveld S, Franssen D, Levendag PC, Heijmen BJM. Toward Fully Automated Multicriterial Plan Generation: A Prospective Clinical Study. *Int J Radiat Oncol Biol Phys* 2013; **85**: 866–72. doi: <https://doi.org/10.1016/j.ijrobp.2012.04.015>
163. Lechner W, Kragl G, Georg D. Evaluation of treatment plan quality of IMRT and VMAT with and without flattening filter using Pareto optimal fronts. *Radiation Oncol* 2013; **10**: 437–41. doi: <https://doi.org/10.1016/j.radonc.2013.09.020>
164. Kierkels RGJ, Visser R, Bijl HP, Langendijk JA, van't Veld AA, Steenbakkers RJHM, et al. Multicriteria optimization enables less experienced planners to efficiently produce high quality treatment plans in head and neck cancer radiotherapy. *Radiation Oncol* 2015; **10**. doi: <https://doi.org/10.1186/s13014-015-0385-9>
165. Kyroudi A, Petersson K, Ghandour S, Pachoud M, Matzinger O, Ozsahin M, et al. Discrepancies between selected Pareto optimal plans and final deliverable plans in radiotherapy multi-criteria optimization. *Radiation Oncol* 2016; **12**: 346–8. doi: <https://doi.org/10.1016/j.radonc.2016.05.018>
166. Ottosson RO, Engström PE, Sjöström D, Behrens CF, Karlsson A, Knöös T, et al. The feasibility of using Pareto fronts for comparison of treatment planning systems and delivery techniques. *Acta Oncol* 2009; **48**: 233–7. doi: <https://doi.org/10.1080/02841860802251559>
167. Petersson K, Ceberg C, Engström P, Benedek H, Nilsson P, Knöös T. Conversion of helical tomotherapy plans to step-and-shoot IMRT plans-Pareto front evaluation of plans from a new treatment planning system. *Med Phys* 2011; **38**: 3130–. doi: <https://doi.org/10.1118/1.3592934>
168. Müller BS, Shih HA, Efstathiou JA, Bortfeld T, Craft D. Multicriteria plan optimization in the hands of physicians: a pilot study in prostate cancer and brain tumors. *Radiation Oncol* 2017; **12**. doi: <https://doi.org/10.1186/s13014-017-0903-z>
169. Wala J, Craft D, Paly J, Zietman A, Efstathiou J. Maximizing dosimetric benefits of IMRT in the treatment of localized prostate cancer through multicriteria optimization planning. *Med Dosim* 2013; **38**: 298–303. doi: <https://doi.org/10.1016/j.meddos.2013.02.012>
170. Smith WP, Kim M, Holdsworth C, Liao J, Phillips MH. Personalized treatment planning with a model of radiation therapy outcomes for use in multiobjective optimization of IMRT plans for prostate cancer. *Radiation Oncol* 2016; **11**. doi: <https://doi.org/10.1186/s13014-016-0609-7>
171. Chen H, Craft DL, Gierga DP. Multicriteria optimization informed VMAT planning. *Med Dosim* 2014; **39**: 64–73. doi: <https://doi.org/10.1016/j.meddos.2013.10.001>
172. Kamran SC, Mueller BS, Paetzold P, Dunlap J, Niemierko A, Bortfeld T, et al. Multi-criteria optimization achieves superior normal tissue sparing in a planning study of intensity-modulated radiation therapy for RTOG 1308-eligible non-small cell lung cancer patients. *Radiation Oncol* 2016; **11**: 515–20. doi: <https://doi.org/10.1016/j.radonc.2015.12.028>
173. Voet PWJ, Dirkx MLP, Breedveld S, Heijmen BJM. Automated generation of IMRT treatment plans for prostate cancer patients with metal hip prostheses: Comparison of different planning strategies. *Med Phys* 2013; **40**: 071704. doi: <https://doi.org/10.1118/1.4808117>
174. Sharfo AWM, Breedveld S, Voet PWJ, Heijkoop ST, Mens J-WM, Hoogeman MS, et al. Validation of fully automated VMAT plan generation for library-based plan-of-the-day cervical cancer radiotherapy. *PLoS One* 2016; **11**: e0169202. doi: <https://doi.org/10.1371/journal.pone.0169202>
175. Buergy D, Sharfo AWM, Heijmen BJM, Voet PWJ, Breedveld S, Wenz F, et al. Fully automated treatment planning of spinal metastases – A comparison to manual planning of Volumetric Modulated Arc Therapy for conventionally fractionated irradiation. *Radiation Oncol* 2017; **12**. doi: <https://doi.org/10.1186/s13014-017-0767-2>
176. Ghandour S, Matzinger O, Pachoud M. Volumetric-modulated arc therapy planning using multicriteria optimization for localized prostate cancer. *J Appl Clin Med Phys* 2015; **16**: 258–69. doi: <https://doi.org/10.1120/jacmp.v16i3.5410>
177. Rønne HS, Wee L, Pløen J, Appelt AL. Feasibility of preference-driven radiotherapy dose treatment planning to support shared decision making in anal cancer. *Acta Oncol* 2017; **56**: 1277–85. doi: <https://doi.org/10.1080/0284186X.2017.1315174>
178. Craft D. Calculating and controlling the error of discrete representations of Pareto surfaces in convex multi-criteria optimization. *Phys Medica* 2010; **26**: 184–91. doi: <https://doi.org/10.1016/j.ejmp.2009.11.005>

179. McGarry CK, Bokrantz R, O'Sullivan JM, Hounsell AR. Advantages and limitations of navigation-based multicriteria optimization (MCO) for localized prostate cancer IMRT planning. *Med Dosim* 2014; **39**: 205–11. doi: <https://doi.org/10.1016/j.meddos.2014.02.002>
180. Della Gala G, Dirx MLP, Hoekstra N, Franssen D, Lanconelli N, van de Pol M, et al. Fully automated VMAT treatment planning for advanced-stage NSCLC patients. *Strahlentherapie Und Onkol* 2017; **193**: 402–9. doi: <https://doi.org/10.1007/s00066-017-1121-1>
181. Sharfo AWM, Voet PWJ, Breedveld S, Mens JWM, Hoogeman MS, Heijmen BJM. Comparison of VMAT and IMRT strategies for cervical cancer patients using automated planning. *Radiother Oncol* 2015; **114**: 395–401. doi: <https://doi.org/10.1016/j.radonc.2015.02.006>
182. van de Schoot AJ, Visser J, van Kesteren Z, Janssen TM, Rasch CR, Bel A. Beam configuration selection for robust intensity-modulated proton therapy in cervical cancer using Pareto front comparison. *Phys Med Biol* 2016; **61**: 1780–. doi: <https://doi.org/10.1088/0031-9155/61/4/1780>
183. Craft DL, Halabi TF, Shih HA, Bortfeld TR. Approximating convex pareto surfaces in multiobjective radiotherapy planning. *Med Phys* 2006; **33**: 3399–. doi: <https://doi.org/10.1118/1.2335486>
184. Craft DL, Halabi TF, Shih HA, Bortfeld TR. Approximating convex Pareto surfaces in multiobjective radiotherapy planning. *Med Phys* 2006; **33**: 3399–407. doi: <https://doi.org/10.1118/1.2335486>
185. Voet PWJ, Dirx MLP, Breedveld S, Al-Mamgani A, Incrocci L, Heijmen BJM. Fully Automated Volumetric Modulated Arc Therapy Plan Generation for Prostate Cancer Patients. *Int J Radiat Oncol Biol Phys* 2014; **88**: 1175–9. doi: <https://doi.org/10.1016/j.ijrobp.2013.12.046>
186. Rossi L, Breedveld S, Heijmen BJM, Voet PWJ, Lanconelli N, Aluwini S. On the beam direction search space in computerized non-coplanar beam angle optimization for IMRT—prostate SBRT. *Phys Med Biol* 2012; **57**: 5441–58. doi: <https://doi.org/10.1088/0031-9155/57/17/5441>
187. van de Water S, Kraan AC, Breedveld S, Schillemans W, Teguh DN, Kooy HM, et al. Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams. *Phys Med Biol* 2013; **58**: 6969–83. doi: <https://doi.org/10.1088/0031-9155/58/19/6969>
188. Powis R, Bird A, Brennan M, Hinks S, Newman H, Reed K, et al. Clinical implementation of a knowledge based planning tool for prostate VMAT. *Radiat Oncol* 2017; **12**. doi: <https://doi.org/10.1186/s13014-017-0814-z>
189. Chin Snyder K, Kim J, Reding A, Fraser C, Gordon J, Ajlouni M, et al. Development and evaluation of a clinical model for lung cancer patients using stereotactic body radiotherapy (SBRT) within a knowledge-based algorithm for treatment planning. *J Appl Clin Med Phys* 2016; **17**: 263–75. doi: <https://doi.org/10.1120/jacmp.v17i6.6429>
190. Wu H, Jiang F, Yue H, Li S, Zhang Y. A dosimetric evaluation of knowledge-based VMAT planning with simultaneous integrated boosting for rectal cancer patients. *J Appl Clin Med Phys* 2016; **17**: 78–85. doi: <https://doi.org/10.1120/jacmp.v17i6.6410>
191. Stanhope C, Wu QJ, Yuan L, Liu J, Hood R, Yin F-F, et al. Utilizing knowledge from prior plans in the evaluation of quality assurance. *Phys Med Biol* 2015; **60**: 4873–91. doi: <https://doi.org/10.1088/0031-9155/60/12/4873>
192. Boylan C, Rowbottom C. A bias-free, automated planning tool for technique comparison in radiotherapy - application to nasopharyngeal carcinoma treatments. *J Appl Clin Med Phys* 2014; **15**: 213–25. doi: <https://doi.org/10.1120/jacmp.v15i1.4530>
193. Bijman RG, Breedveld S, Arts T, Astreindou E, de Jong MA, Granton PV, et al. Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy. *Acta Oncol* 2017; **56**: 1444–50. doi: <https://doi.org/10.1080/0284186X.2017.1355113>
194. Meijer GJ, van der Toorn P-P, Bal M, Schuring D, Weterings J, de Wildt M. High precision bladder cancer irradiation by integrating a library planning procedure of 6 prospectively generated SIB IMRT plans with image guidance using lipidol markers. *Radiother Oncol* 2012; **105**: 174–9. doi: <https://doi.org/10.1016/j.radonc.2012.08.011>
195. Heijkoop ST, Langerak TR, Quint S, Bondar L, Mens JWM, Heijmen BJM, et al. Clinical Implementation of an Online Adaptive Plan-of-the-Day Protocol for Nonrigid Motion Management in Locally Advanced Cervical Cancer IMRT. *Int J Radiat Oncol Biol Phys* 2014; **90**: 673–9. doi: <https://doi.org/10.1016/j.ijrobp.2014.06.046>
196. van de Schoot AJAJ, Visser J, van Kesteren Z, Janssen TM, Rasch CRN, Bel A. Beam configuration selection for robust intensity-modulated proton therapy in cervical cancer using Pareto front comparison. *Phys Med Biol* 2016; **61**: 1780–94. doi: <https://doi.org/10.1088/0031-9155/61/4/1780>
197. Ciller C, De Zanet SI, Rügsegger MB, Pica A, Sznitman R, Thiran J-P, et al. Automatic segmentation of the eye in 3D magnetic resonance imaging: a novel statistical shape model for treatment planning of retinoblastoma. *Int J Radiat Oncol Biol Phys* 2015; **92**: 794–802. doi: <https://doi.org/10.1016/j.ijrobp.2015.02.056>
198. Fortunati V, Verhaart RF, van der Lijn F, Niessen WJ, Veenland JF, Paulides MM, et al. Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling. *Med Phys* 2013; **40**: 071905. doi: <https://doi.org/10.1118/1.4810971>
199. Wong WKH, Leung LHT, Kwong DLW. Evaluation and optimization of the parameters used in multiple-atlas-based segmentation of prostate cancers in radiation therapy. *Br J Radiol* 2016; **89**: 20140732. doi: <https://doi.org/10.1259/bjr.20140732>
200. Martínez F, Romero E, Dréan G, Simon A, Haigron P, de Crevoisier R, et al. Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multi-scale edge detector. *Phys Med Biol* 2014; **59**: 1471–84. doi: <https://doi.org/10.1088/0031-9155/59/6/1471>
201. Li D, Liu L, Kapp DS, Xing L. Automatic liver contouring for radiotherapy treatment planning. *Phys Med Biol* 2015; **60**: 7461–83. doi: <https://doi.org/10.1088/0031-9155/60/19/7461>
202. Chen A, Niermann KJ, Deeley MA, Dawant BM. Evaluation of multiple-atlas-based strategies for segmentation of the thyroid gland in head and neck CT images for IMRT. *Phys Med Biol* 2012; **57**: 93–111. doi: <https://doi.org/10.1088/0031-9155/57/1/93>
203. Chandra SS, Dowling JA, Greer PB, Martin J, Wratten C, Pichler P, et al. Fast automated segmentation of multiple objects via spatially weighted shape learning. *Phys Med Biol* 2016; **61**: 8070–84. doi: <https://doi.org/10.1088/0031-9155/61/22/8070>
204. Altman MB, Kavanaugh JA, Wooten HO, Green OL, DeWees TA, Gay H, et al. A framework for automated contour quality assurance in radiation therapy including adaptive techniques. *Phys Med Biol* 2015; **60**: 5199–209. doi: <https://doi.org/10.1088/0031-9155/60/13/5199>
205. Otto K. Real-time interactive treatment planning. *Phys Med Biol* 2014; **59**: 4845–59. doi: <https://doi.org/10.1088/0031-9155/59/17/4845>

206. Ziegenhein P, Ph Kamerling C, Oelfke U. Interactive dose shaping part 1: a new paradigm for IMRT treatment planning. *Phys Med Biol* 2016; **61**: 2457–70. doi: <https://doi.org/10.1088/0031-9155/61/6/2457>
207. Ph Kamerling C, Ziegenhein P, Sterzing F, Oelfke U. Interactive dose shaping part 2: proof of concept study for six prostate patients. *Phys Med Biol* 2016; **61**: 2471–84. doi: <https://doi.org/10.1088/0031-9155/61/6/2471>
208. Raaymakers BW, Lagendijk JJ, Overweg J, Kok JG, Raaijmakers AJ, Kerkhof EM, et al. Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol* 2009; **54**: N229–N237. doi: <https://doi.org/10.1088/0031-9155/54/12/N01>
209. Mutic S, Dempsey JF. The ViewRay System: Magnetic Resonance–Guided and Controlled Radiotherapy. *Semin Radiat Oncol* 2014; **24**: 196–9. doi: <https://doi.org/10.1016/j.semradonc.2014.02.008>
210. Li N, Zarepisheh M, Uribe-Sanchez A, Moore K, Tian Z, Zhen X, et al. Automatic treatment plan re-optimization for adaptive radiotherapy guided with the initial plan DVHs. *Phys Med Biol* 2013; **58**: 8725–. doi: <https://doi.org/10.1088/0031-9155/58/24/8725>
211. Amit G, Purdie TG. Automated planning of breast radiotherapy using cone beam CT imaging. *Med Phys* 2015; **42**: 770–. doi: <https://doi.org/10.1118/1.4905111>
212. Court LE, Tishler RB, Petit J, Cormack R, Chin L. Automatic online adaptive radiation therapy techniques for targets with significant shape change: a feasibility study. *Phys Med Biol* 2006; **51**: 2493–. doi: <https://doi.org/10.1088/0031-9155/51/10/009>
213. Li X, Quan EM, Li Y, Pan X, Zhou Y, Wang X, et al. A fully automated method for CT-on-rails-guided online adaptive planning for prostate cancer intensity modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 2013; **86**: 835–. doi: <https://doi.org/10.1016/j.ijrobp.2013.04.014>
214. Acharya S, Fischer-Valuck BW, Kashani R, Parikh P, Yang D, Zhao T, et al. Onlinemagnetic resonance image guided adaptive radiation therapy: first clinical applications. *Int J Radiat Oncol Biol Phys* 2016; **94**: 394–403. doi: <https://doi.org/10.1016/j.ijrobp.2015.10.015>
215. van de Sande MA, Creutzberg CL, van de Water S, Sharfo AW, Hoogeman MS. Which cervical and endometrial cancer patients will benefit most from intensity-modulated proton therapy? *Radiother Oncol* 2016; **120**: 397–403. doi: <https://doi.org/10.1016/j.radonc.2016.06.016>
216. Euchner J. Building aculture of innovation. *Res Manag* 2016; **59**: 10–11. doi: <https://doi.org/10.1080/08956308.2016.1232131>



Detectability of pulmonary nodules by deep learning: results from a phantom study

Qiong Li¹ · Qing-chu Li¹ · Rui-ting Cao² · Xiang Wang¹ · Ru-tan Chen¹ · Kai Liu¹ · Li Fan¹ · Yi Xiao¹ · Zi-tian Zhang² · Chi-Cheng Fu² · Qiong Song² · Weiping Liu² · Qu Fang² · Shi-yuan Liu¹

Received: 17 June 2019 / Revised: 25 September 2019 / Accepted: 10 October 2019
© Springer Nature Singapore Pte Ltd. 2019

Abstract

Purpose To investigate how nodule size, nodule density, scan dose, slice thickness and reconstruction methods affect the performance of a deep learning (DL) model for detection of pulmonary nodules in phantom CT scans.

Materials and methods Spherical lung nodule phantoms of two different densities (−630 HU and +100 HU) and five different sizes (3, 5, 8, 10, and 12 mm) were inserted into an anthropomorphic chest phantom. CT data were scanned and reconstructed using three different tube current (10, 50, 200 mAs), two different slice thickness of 1 and 2 mm, four reconstruction methods (FBP-standard (FBP-STD), FBP-Y sharp kernel (FBP-YA), iDose4-standard kernel (iDose4-STD) and iDose4-Y sharp kernel (iDose4-YA)). Evaluation of deep learning model focused on detection sensitivity and precision.

Results According to the statistical results from the study, we found that the sensitivity and precision performance depends on the nodule sizes, nodule type, tube current, reconstruction methods and image thickness. Comparing the solid (100 HU) and ground-glass (GGO, −630 HU) nodule phantoms, solid nodule phantom predictions are rarely affected by tube current, reconstruction methods and nodule sizes. Both sensitivity and precision are close to 100% in all solid nodule phantom prediction cases. While the sensitivity and precision metrics of GGO nodule phantoms change in a wide range from 42.9 to 100%. Larger nodule size and higher tube current gives a better sensitivity and precision for GGO nodule phantoms in most cases. We also analyze the relationships between the image thickness and the reconstruction methods. For 1-mm thickness images, iDose4-STD and FBP-STD shows a better result in both sensitivity and precision metrics. As for 2-mm thickness images, iDose-YA and FBP-YA gain a better performance.

Conclusion The results of this phantom study demonstrated that high stability and flexibility of deep learning model can be used in daily clinical and screening practice.

Keywords Lung nodule · Deep learning · Detectability · Computed tomography · Phantom

Introduction

The early identification of pulmonary nodules is an important task for the management of lung cancer. However, reading of CT images by radiologists for detecting the presence of pulmonary nodules is a tedious and time-consuming work. The increased clinical demand on radiologists and heavier workload have resulted in less time for interpretation

of images together with high risk potential for more detection or interpretation errors to occur [1, 2]. Studies have reported that the double reading by radiologist showed a detection rate of 59.1% and a missed rate of 40.9% [3]. Missed lung cancers are an important diagnostic concern. Hence, computer-aided detection (CAD) of lung nodules would be valuable for lung cancer screening.

Although many traditional academic and commercial computer-aided detection (CAD) systems have been developed to improve the nodule detection rate [4–7], they are still not commonly used in our routine work. Based on conventional image-processing techniques, these algorithms showed a wide range of sensitivity and high false-positive rates (FPR), ranges from 73 to 96.7% with FP rates of 0.55–8.2 per scan on an average, and may not be robust

Qiong Li, Qing-chu Li and Rui-ting Cao contributed equally to this work.

✉ Shi-yuan Liu
liushiyuan@smmu.edu.cn

Extended author information available on the last page of the article

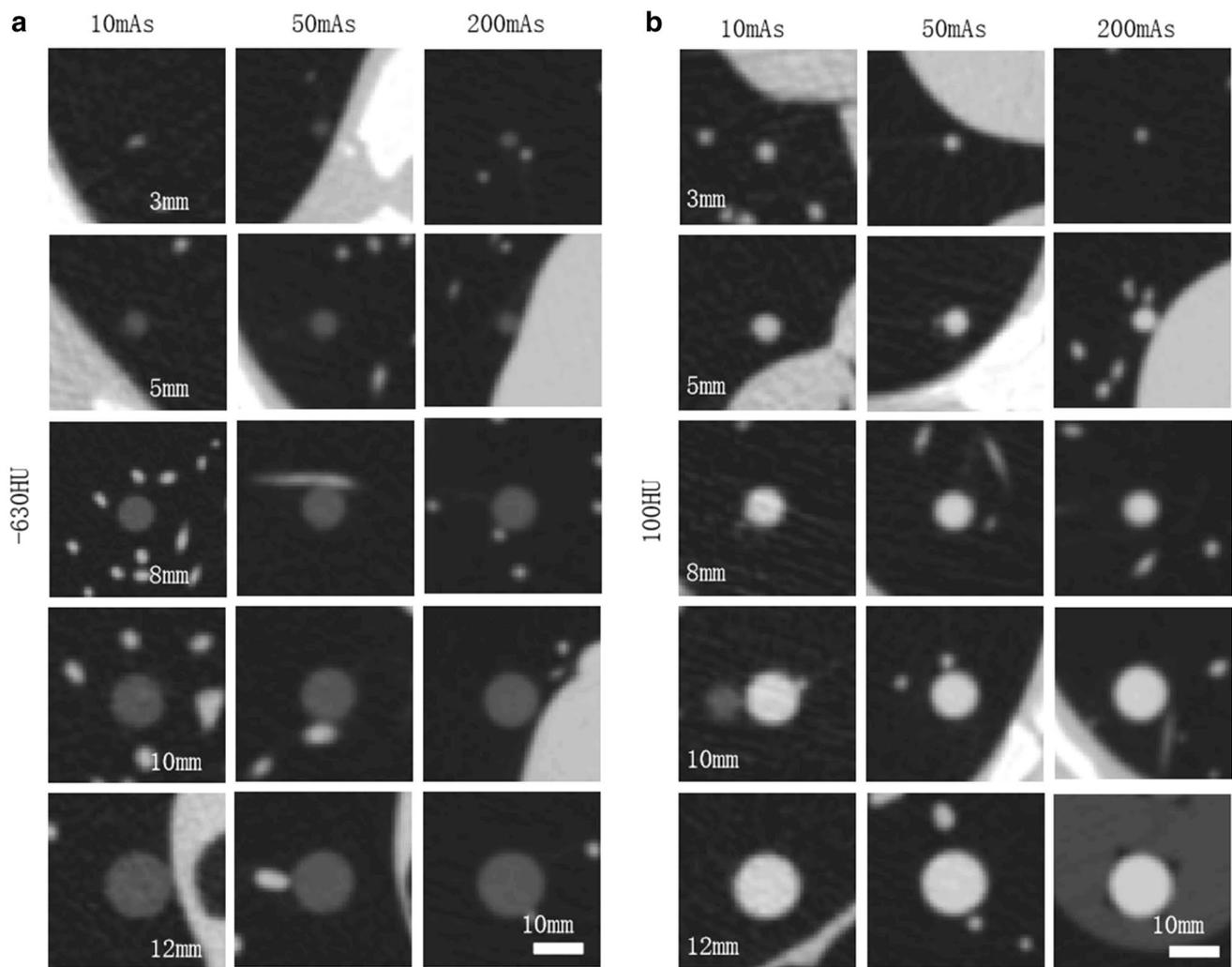


Fig. 1 Example of CT images reconstructed with FBP-STD methods at slice thickness of 1 mm

across various data sources [8]. Recently, deep learning has been widely used in many real fields, particularly it has been used for the detection and diagnosis of the lesions in medical images, improving the accuracy with efficiency. Deep learning (DL) allows computational models that are composed of a large number of hidden layers to learn representation of data with multiple levels of abstraction. Now more and more studies have explored their use for detection and diagnosis of pulmonary nodules [9–11].

Data are the core of data mining required by deep learning algorithm. It is impossible to obtain better results by only mastering algorithm and lacking data. However, imaging data from clinical studies or lung cancer screening programs are usually varied by slice thickness, reconstruction algorithm and scan parameters, may have implications for the robustness of the DL model. In our previous clinical study, no significant dependence regarding radiation dose was observed, and the DL model showed elevated overall

sensitivity compared with manual review of lung nodules [12]. For accurate analysis, images acquired with different parameters should be normalized to the same conditions. Employing the lung nodule phantom allowed us to scan the identical anatomical and lesion conditions repetitively at multiple settings from standard to ultra low dose, which is clearly not possible in patients. Therefore, we tried to use the diversified phantom data to investigate the influence of radiation dose settings, reconstruction algorithm and slice thickness on the performance of the DL model.

Materials and methods

Phantom and synthetic nodules

An anthropomorphic thoracic phantom (Lungman, Kyoto Kagaku, Tokyo, Japan) was used in this study. The phantom

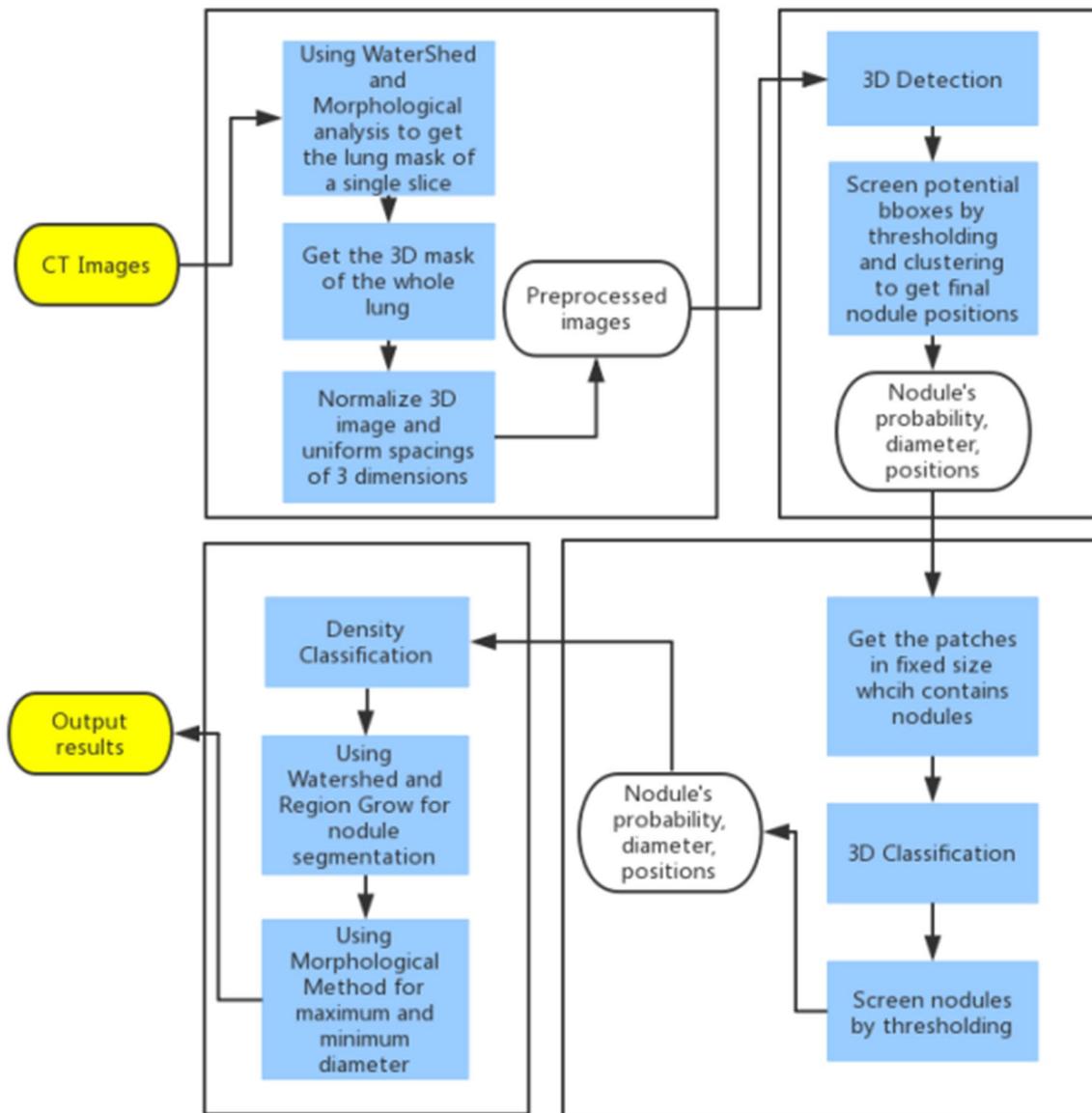


Fig. 2 Flowchart of image analysis and nodule prediction

has artificial thoracic wall, heart, mediastinum, diaphragm and pulmonary vessels. The simulated spherical nodules included five different sizes (3, 5, 8, 10, and 12 mm), two different densities: -630 HU (ground-glass nodule), 100 HU (solid nodule). In each scan, The 10 synthetic nodules were randomly placed at different sites of the phantom, such as the lung apex, tracheal bifurcation, subpleural, attached to the vasculature and so on.

Scan protocols

All CT data were obtained with a 256-slice MDCT scanner (Brilliance iCT, Philips Healthcare, Cleveland, OH, USA). The following data acquisition parameters were kept

constant: detector configuration: 128×0.625 mm, beam pitch: 0.99; rotation time: 0.5 s; and FOV: 350 mm. The effective tube current time products (mAs) were 10, 50, and 200 mAs (Fig. 1). Thin-section CT images at all tube currents were reconstructed as a contiguous section thickness of 1 mm and 2 mm with the FBP-standard (FBP-STD), FBP-Y sharp kernel (FBP-YA), iDose4-standard kernel (iDose4-STD) and iDose4-Y sharp kernel (iDose4-YA). iDose4 (Philips Healthcare) was one of the iterative reconstruction algorithms. At each tube current, the chest phantoms were scanned 3 times. Thus, yielding a total of 72 datasets.

Table 1 Detection sensitivity of the nodule phantoms in slice thickness of 1 mm

Reconstruction method	Size	Density					
		− 630 HU			100 HU		
		Tube current					
		10 mAs (%)		50 mAs (%)		200 mAs (%)	
10 mAs (%)		50 mAs (%)		200 mAs (%)			
iDose4-STD	3 mm	58.3	75.0	71.4	95.8	100.0	100.0
	5 mm	73.9	78.3	78.3	100.0	100.0	100.0
	8 mm	90.9	85.0	81.8	95.8	100.0	100.0
	10 mm	100.0	95.8	95.0	100.0	100.0	100.0
	12 mm	100.0	100.0	100.0	100.0	100.0	100.0
iDose4-YA	3 mm	50.0	46.7	84.6	95.2	100.0	100.0
	5 mm	77.3	81.8	82.6	100.0	100.0	100.0
	8 mm	81.8	86.4	85.7	95.7	100.0	100.0
	10 mm	100.0	95.5	96.0	100.0	95.2	100.0
	12 mm	100.0	100.0	100.0	100.0	100.0	100.0
FBP-STD	3 mm	55.6	56.2	90.9	100.0	100.0	100.0
	5 mm	77.3	72.7	78.3	100.0	100.0	100.0
	8 mm	91.3	86.4	85.0	100.0	100.0	100.0
	10 mm	90.0	92.0	91.7	100.0	100.0	100.0
	12 mm	100.0	95.5	100.0	100.0	100.0	100.0
FBP-YA	3 mm	42.9	57.1	66.7	100.0	100.0	100.0
	5 mm	88.2	82.6	90.9	100.0	100.0	100.0
	8 mm	94.4	90.5	85.7	100.0	100.0	100.0
	10 mm	77.8	95.5	95.5	100.0	95.5	100.0
	12 mm	100.0	100.0	100.0	100.0	100.0	100.0

Data analysis

Aitrox (Shanghai, China) provided software and hardware support. The authors who were not affiliated with Aitrox had control of data and information submitted for publication.

As shown in Fig. 2, we analyzed the CT images and do nodule prediction by four main stages: preprocess, detection, classification and postprocess. CT images are the input of the whole prediction process. Before predicted, they first go through the preprocess stage to reduce the unrelated region in further process so as to promote the computation efficiency. Specifically, we used watershed and morphological analysis to extract the lung regions in CT images. After segmenting the lung region through thresholding, the intact lung region mask was generated, while the unrelated objects were removed by size and location analysis. Then the masked image will be normalized between zero to one and its spacings will be uniformed to one.

The second stage is the detection stage. To extract possible pulmonary nodule from the lung region and ensure high sensitivity, we developed a nodule detector with the framework of RPN (Region Proposal Network), where using U-net, a commonly used network in medical image segmentation, as its backbone. Different from classic U-net [13], we use dilated convolution layers as well as self-designed block which combine dense block and residual block. This progress may generate more various features which can provide more choices for optimizing the model behavior. With the encoder–decoder workflow of U-net, the features of nodule are enhanced. After obtaining the feature map by U-net, RPN is presented to achieve the precise location and the probability of each objects. In our implementation, the RPN generated 3 potential objects with different window size as input and computed classification score for each object proposals during bounding box regression. After that,

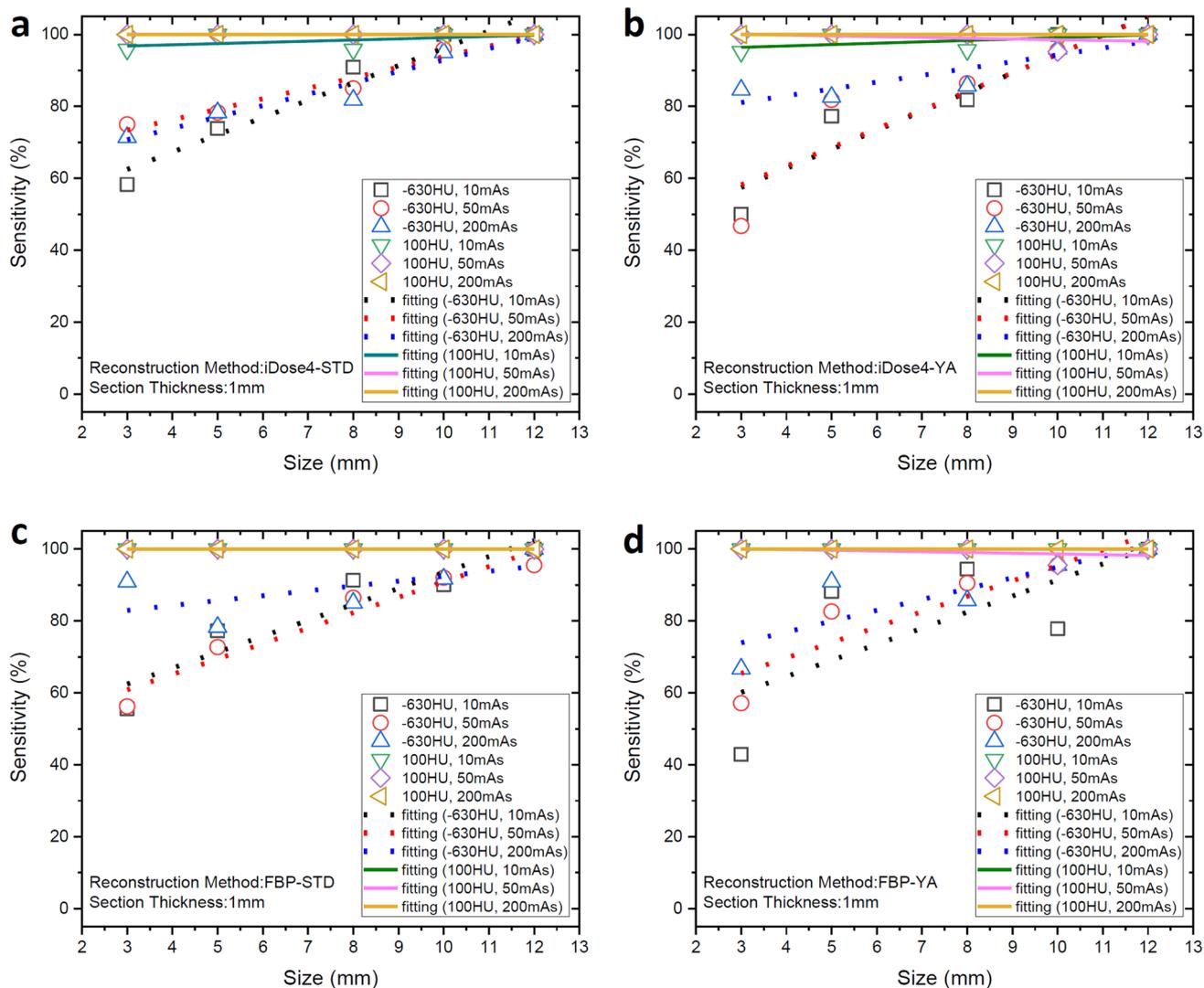


Fig. 3 The sensitivity change according to tumor size change with section thickness of 1 mm and the fitted linear of each case were plotted

the precise centroid was computed by weighting the location and probability of cluster of bounding boxes.

Since the detection results still contain false-positive prediction, we added a classification stage after the detection to eliminate them. In the classification stage, we implemented a WRN (Wide Residual Network) based on a classical structure—ResNet [14], where residual learning allows deeper network to classify whether each patch contains a nodule. However, as the model goes deeper, the model will encounter overfitting problem even though residual block is applied. As the result, we put the inception module into the residual block and Global Average Pooling is also applied in substitute for Fully Connected Layer. We use these techniques

to reach the balance between width and depth of deep networks. After the classification stage, most of the false-positive nodules can be removed by filtering low-probability patches.

The final stage is the postprocess stage which is to calculate the features of predicted nodules. For each nodule candidate, we form another classification model to predict its density. Then, the nodule is segmented along its edge by watershed and region grow algorithms. Finally, we apply morphological method to the segmentation result to calculate the nodule’s maximum and minimum diameter.

Table 2 Detection sensitivity of the nodule phantoms in slice thickness of 2 mm

Reconstruction model	Size	Density					
		- 630 HU			100 HU		
		Tube current					
		10 mAs (%)		50 mAs (%)		200 mAs (%)	
10 mAs (%)		50 mAs (%)		200 mAs (%)			
iDose4-STD	3 mm	42.90	75.00	66.70	86.40	100.00	100.00
	5 mm	73.90	82.60	79.20	100.00	100.00	100.00
	8 mm	83.30	86.40	78.30	96.00	100.00	100.00
	10 mm	86.40	92.30	100.00	96.00	100.00	96.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00
iDose4-YA	3 mm	63.60	72.70	86.70	95.50	100.00	100.00
	5 mm	72.70	80.00	73.90	100.00	100.00	100.00
	8 mm	87.00	83.30	87.50	95.50	100.00	100.00
	10 mm	82.60	91.30	91.30	96.20	100.00	100.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00
FBP-STD	3 mm	42.90	70.00	75.00	90.90	100.00	100.00
	5 mm	73.90	82.60	78.30	100.00	100.00	100.00
	8 mm	82.60	82.60	84.00	95.70	100.00	100.00
	10 mm	91.30	90.50	100.00	96.30	100.00	96.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00
FBP-YA	3 mm	62.50	81.80	66.70	89.50	100.00	100.00
	5 mm	86.40	77.30	77.30	100.00	100.00	100.00
	8 mm	85.70	83.30	87.00	100.00	95.50	100.00
	10 mm	87.00	86.40	100.00	100.00	100.00	100.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00

We get the final output from the postprocess stage. The output is a list of predicted nodules in the CT image, and each nodule contains information of its position, probability of being a nodule, diameter, density and its segmentation results.

The models mentioned above were trained based on real patients' CT data from hospitals. We randomly crop a patch from processed image, whose size is 128*128*128 and with at least a nodule in it. Then, we use a data generator to send the patch into the training process which will give the coordinates(X, Y, Z) and size as well as the probability of all potential nodules. Then, based on the label (coordinates and size of true nodules), we get the difference between our predictions and labels, and use back propagation techniques which is mainly an optimizer to get the update of weights to lower the difference. With thousands of iterations, the deep network learned to imitate the observations of doctors' and achieve good metrics of the test set.

To give the detection metrics, we refer to the rule of LUNA competition to determine whether the prediction matches the annotation. The hit criterium is that, a predicted nodule should be in the range of the radius of the annotated nodule center. If hit, the predicted nodule is considered to be a true positive (TP). Predicted nodules that do not hit any annotated nodule are considered false positives (FP), and annotated nodules not matched to any predicted nodule are counted as false negatives (FN). We further calculate the sensitive and precision based on TP, FP, and FN using the two equations below:

$$\text{Sensitive} = \text{TP}/(\text{TP} + \text{FN}),$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}).$$

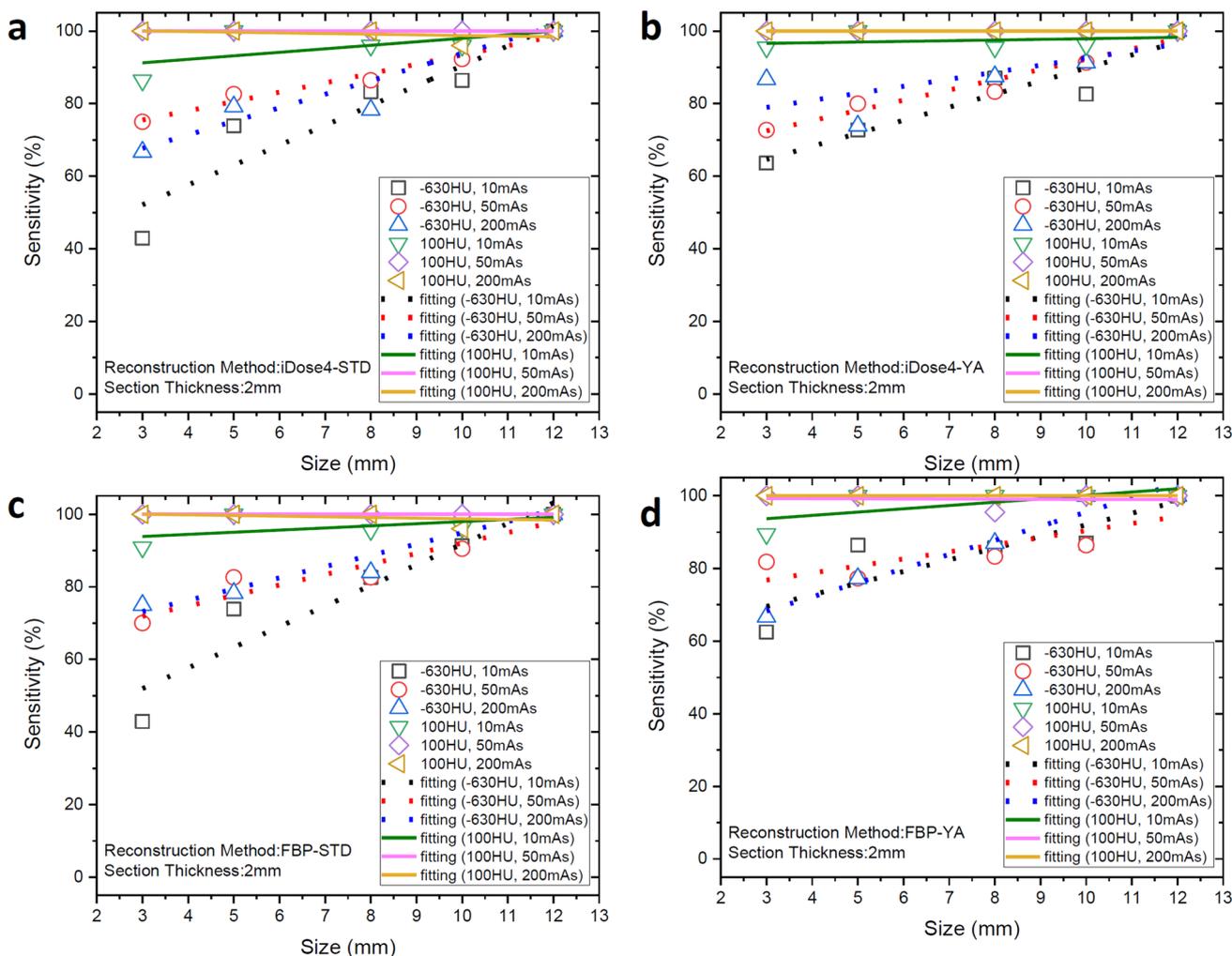


Fig. 4 The sensitivity change according to tumor size with section thickness of 2 mm, change together with the fitted line of each case

Results

The sensitivity performance of nodule phantom for images with thickness of 1 mm is shown in Table 1 and Fig. 3. For all - 630 HU nodule phantoms, the sensitivity ranges from 42.9 to 100.0% in different nodule sizes, reconstruction methods and mAs values. For small GGO nodule phantoms of - 630 HU and 3 mm, best sensitivity performances for 10 mAs, 50 mAs, 200 mAs are 58.3% with iDose4-STD, 75.0% with iDose4-STD, and 90.9% with

FBP-STD, respectively. As for 100 HU nodule phantoms, we got sensitivity ranging from 95.2 to 100.0%. FBP-STD gains the best average sensitivity performance among all four reconstruction methods for solid nodule phantoms with all sizes and mAs values.

Table 2 and Fig. 4 show the sensitivity performance for images with thickness of 2 mm. The sensitivity of - 630 HU nodule phantoms still ranges from 42.9 to 100.0% with a larger thickness, but the best sensitivity results for small 3 mm GGO phantoms are obtained by different reconstruction methods. We got 63.6% for 10 mAs with

Table 3 Detection precision of the nodule phantoms in section thickness of 1 mm

Reconstruction method	Size	Density										
		– 630 HU			100 HU							
		Tube current										
		10 mAs (%)		50 mAs (%)		200 mAs (%)		10 mAs (%)		50 mAs (%)		200 mAs (%)
iDose4-STD	3 mm	50.0	66.7	71.4	91.7	100.0	100.0					
	5 mm	69.6	73.9	73.9	100.0	100.0	100.0					
	8 mm	86.4	85.0	81.8	87.5	90.5	91.3					
	10 mm	83.3	87.5	85.0	100.0	100.0	100.0					
	12 mm	100.0	100.0	100.0	95.2	100.0	100.0					
iDose4-YA	3 mm	50.0	46.7	84.6	95.2	100.0	100.0					
	5 mm	77.3	72.7	82.6	100.0	100.0	100.0					
	8 mm	81.8	86.4	85.7	95.7	91.3	95.7					
	10 mm	85.0	81.8	88.0	94.7	95.2	100.0					
	12 mm	100.0	100.0	100.0	100.0	100.0	100.0					
FBP-STD	3 mm	55.6	56.2	81.8	95.5	100.0	100.0					
	5 mm	72.7	72.7	73.9	100.0	100.0	100.0					
	8 mm	87.0	86.4	85.0	87.0	91.3	90.9					
	10 mm	80.0	84.0	83.3	100.0	100.0	100.0					
	12 mm	100.0	95.5	100.0	95.2	100.0	100.0					
FBP-YA	3 mm	42.9	50.0	66.7	100.0	100.0	100.0					
	5 mm	88.2	78.3	81.8	100.0	100.0	100.0					
	8 mm	94.4	90.5	81.0	100.0	90.9	95.5					
	10 mm	77.8	81.8	86.4	100.0	95.5	100.0					
	12 mm	100.0	100.0	100.0	100.0	100.0	100.0					

iDose4-YA, 81.8% for 50 mAs with FBP-YA, and 86.7% for 200 mAs with iDose4-YA. In 2-mm CT images, the sensitivity of 100 HU nodule phantoms goes a bit lower than 1 mm images, which ranges from 86.4 to 100.0%. Reconstruction method of iDose4-YA gains the best average sensitivity results.

In addition to the sensitivity, we also calculated the statistical data for precision performance. Table 3 and Fig. 5 give the precision for images with thickness of 1 mm. For all – 630 HU nodule phantoms, the precision ranges from 42.9 to 100.0%. Best small 3 mm GGO nodule phantoms' precision result for 10 mAs is 55.6% with FBP-STD, for 50 mAs is 66.7% with iDose4-STD, and for 200 mAs is 84.6% with iDose4-YA. Precision of 100 HU nodule phantoms,

the precision is from 95.5 to 100.0%. FBP-YA gets the best average precision performance in overall mAs values and solid nodule sizes.

Precision performance for images with thickness of 2 mm is demonstrated in Table 4 and Fig. 6. For – 630 phantom nodules, the precision ranges from 42.9% to 100% which is the same as images with 1-mm thickness. Among small 3 mm GGO nodule phantoms, we got highest precision 50.0% for 10 mAs with FBP-YA, 81.8% for 50 mAs still with FBP-YA, and 66.7% for 200 mAs with both iDose4-YA and FBP-STD. As for solid nodule phantoms with 100 HU, the precision ranges from 90.9 to 100.0%. FBP-YA gains the best average precision among all reconstruction methods.

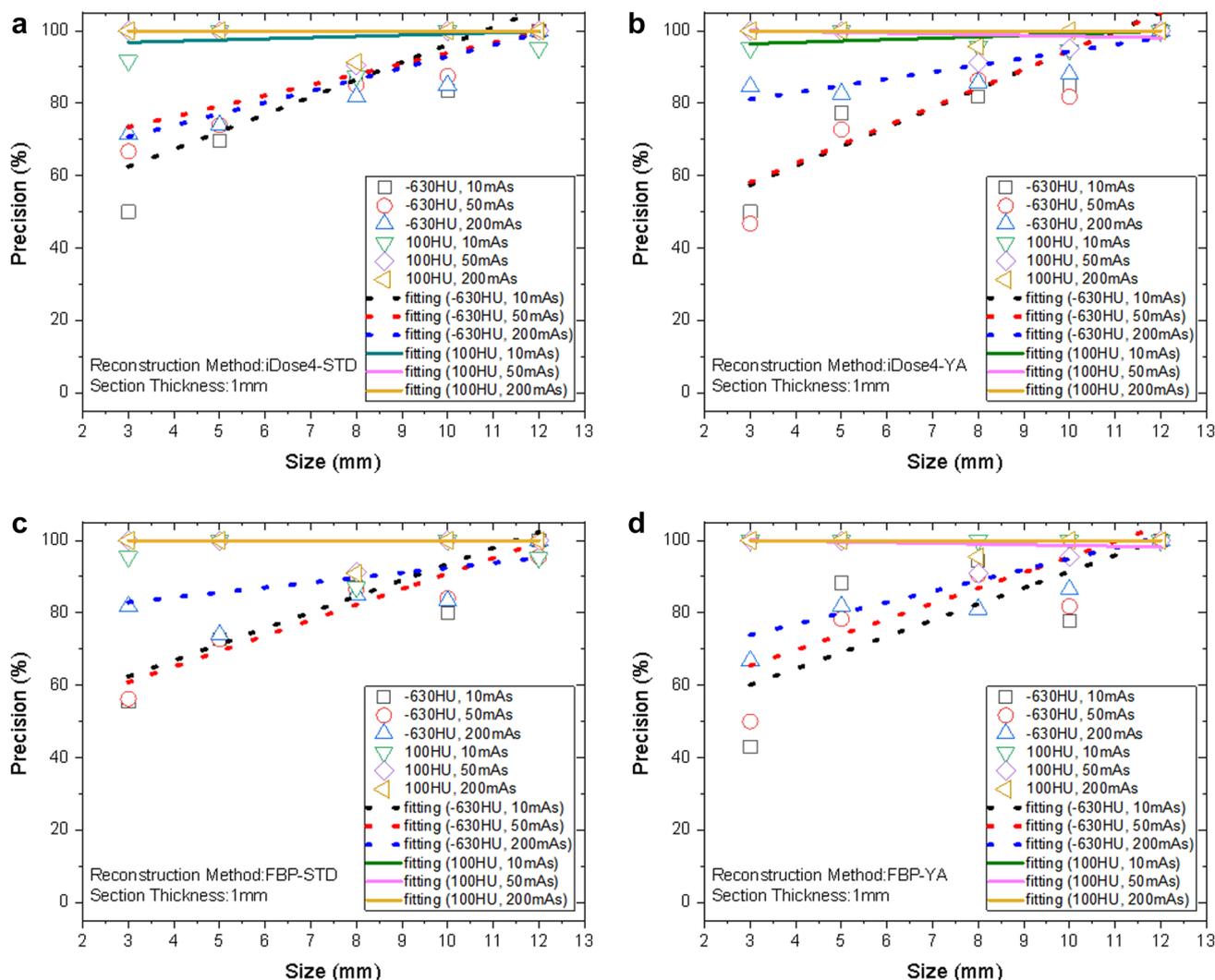


Fig. 5 Detection precision of the nodule phantoms at slice thickness of 1 mm

Discussion

Previous studies have demonstrated that several factors, such as tube current, detector collimation, beam pitch, slice thickness, reconstruction kernels as well as iterative reconstruction algorithms will affect nodule detection by radiologists [15, 16], these scan parameters can affect the image quality or noise. High spatial frequency reconstruction algorithms are good for showing fine structures within lung tissue. The use of this kernel will result in high image noise, also provides a peripheral edge enhancement effect [17]. In this study, we compared two different kernels combined with iterative reconstruction algorithm (iDose) or not: one with higher image noise, and the other with lower image noise.

Wielpütz et al. employed an ex vivo lung phantom and prepared with 162 artificial nodules (20 ± 20 HU) of a clinically relevant volume and maximum diameter (46–1063 μ l, and 6.2–21.5 mm), showed that the sensitivity of a commercially available CAD system on low-dose MDCT scans with a CTDI between 0.25 and 8.07 mGy is 88.9–91.4% for FBP and 88.3–90.1% for IR [18]. Our results also showed that decreasing tube current had no effect on the detection sensitivity and precision of 100 HU phantom nodules, even for small one. DL model is robust over a wide range of exposure settings for solid nodule. However, we found that the change of tube current can affect the performance of deep learning model for the small size of –630 HU nodule phantom. An increase in the image noise may impair the detection

Table 4 Detection precision of the nodule phantoms in section thickness of 2 mm

Reconstruction model	Size	Density					
		−630 HU			100 HU		
		Tube current					
		10 mAs (%)		50 mAs (%)		200 mAs (%)	
iDose4-STD	3 mm	42.90	58.30	58.30	90.90	100.00	100.00
	5 mm	73.90	73.90	75.00	100.00	100.00	100.00
	8 mm	83.30	95.50	78.30	88.00	86.40	90.90
	10 mm	86.40	84.60	87.00	96.00	100.00	96.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00
iDose4-YA	3 mm	45.50	72.70	66.70	95.50	100.00	100.00
	5 mm	72.70	72.00	82.60	100.00	100.00	100.00
	8 mm	87.00	87.50	87.50	90.90	91.30	91.70
	10 mm	82.60	87.00	91.30	96.20	100.00	100.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00
FBP-STD	3 mm	42.90	70.00	66.70	90.90	100.00	100.00
	5 mm	73.90	73.90	73.90	100.00	100.00	100.00
	8 mm	82.60	87.00	84.00	87.00	86.40	90.90
	10 mm	87.00	85.70	87.00	96.30	100.00	96.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00
FBP-YA	3 mm	50.00	81.80	58.30	94.70	100.00	100.00
	5 mm	77.30	72.70	81.80	100.00	100.00	100.00
	8 mm	85.70	91.70	82.60	95.70	90.90	95.70
	10 mm	87.00	86.40	90.50	95.70	96.30	100.00
	12 mm	100.00	100.00	100.00	100.00	100.00	100.00

of ground glass opacity (GGO) nodules at the minimal tube current second product (10 mAs). The difference of CT images between lung nodules and lung parenchyma is relatively small and the increase in image noise at the level of minimal tube current may affect the detection rate. Our study also indicates that DL model's performance at 2 mm is comparable with that at 1 mm.

Importantly, iterative reconstruction (IR) algorithm is not detrimental to DL model's sensitivity and accuracy, appears that IR can be applied alongside DL model for the comprehensive management of low-dose CT screening.

A conventional CAD framework used for nodule detection requires several steps including organ of interest segmentation, lesion candidate detection, feature extraction, selection, and integration via image processing and pattern recognition [19]. In such a pipeline, every step depends heavily and easily on the performance of the previous step. Besides, because medical image combined of many non-linear transformations always has high complexity, the performance of conventional CAD scheme is affected seriously by volume effect, intensity inhomogeneities, artifacts, and the similarity of intensity in different soft tissues. However,

deep learning architectures, which can capture more abstract information, are an effective method to avoid some above problems in the conventional CAD scheme. Unlike traditional studies, deep learning architectures have the advantage of automatic exploitation feature by multi-scale convolution regardless of the limitation of the feature extraction method, and tuning of performance in a seamless fashion. In this study, we proposed a CAD system with deep learning framework. The empirical results of our studies indicate the deep learning method performs well on nodule detection. Otherwise, as a classification model of supervised learning, deep learning presented here shows detection of high stability and flexibility in various conditions including different X-ray tube current, reconstruction method, phantoms diameters, and densities. Moreover, deep learning has the capability to combine information from various medical image types for further diagnosis and analysis.

Our study has several limitations. First, we adopted a commercially available chest CT phantom which included several perfectly round simulated GGOs and solid nodules, their frequency was quite different from that observed in our daily clinical work. Real pulmonary nodules also have

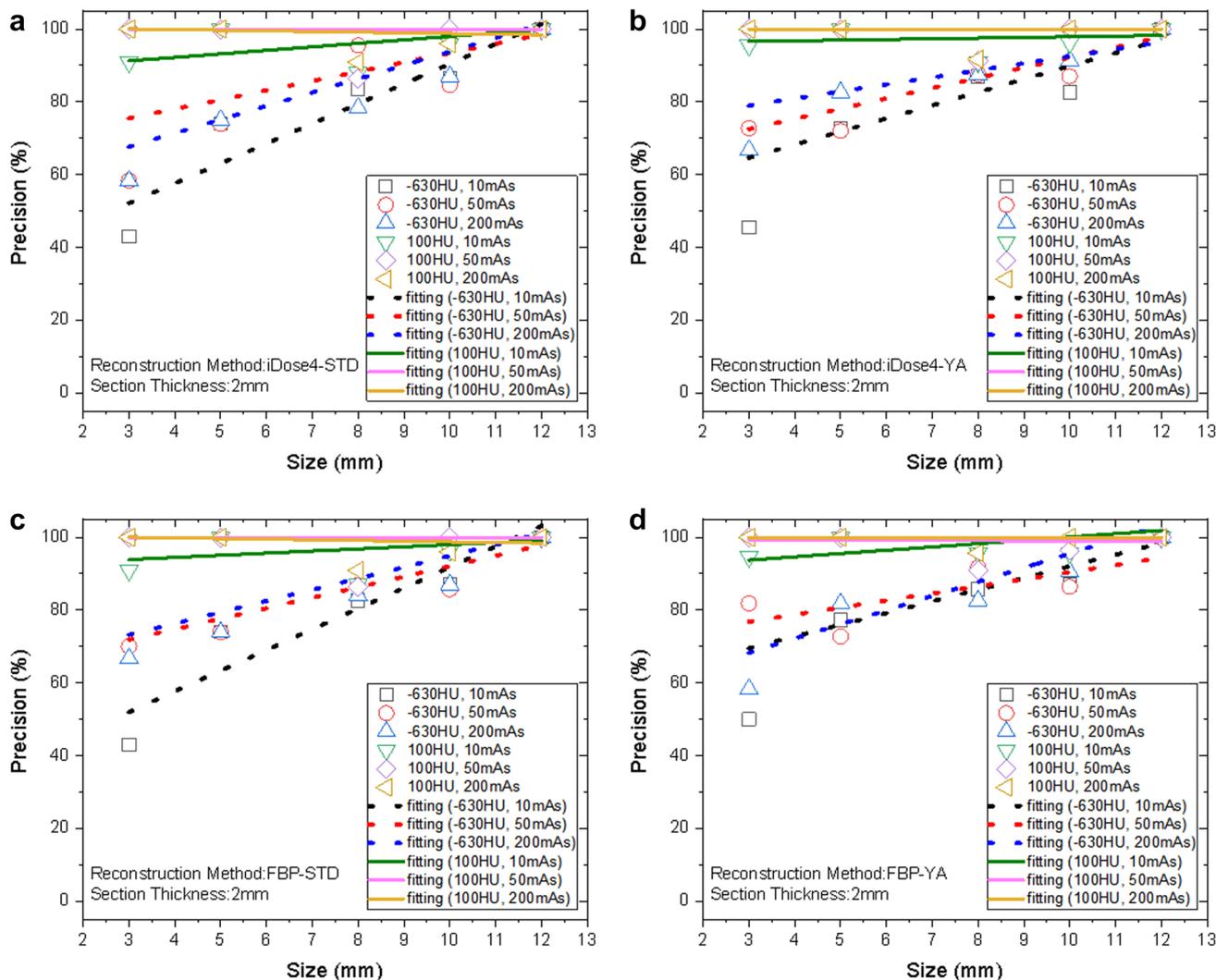


Fig. 6 Detection precision of the nodule phantoms at slice thickness of 2 mm

several factors affect their detection including shape, margin and density. Second, only several scan and reconstruction parameters were evaluated in this study. Another limitation is the relative small sample size of nodules in our study, and this made us refrain from analyzing even smaller subdivisions.

Conclusion

This chest phantom study demonstrated that the change of tube current can affect the performance of deep learning model for the small size of ground glass opacity nodule phantom, whereas for phantom density of 100 HU, deep learning presented here shows detection of high stability and flexibility in various conditions including different tube

current and reconstruction methods. Going forward, this makes deep learning model appropriate to be used in daily clinical and screening practice, provide evidence for future studies with real patients.

Funding The funding has been received from Shanghai Technology Committee Research Program (Grant no. 17411952400), Shanghai Hygiene Committee Intelligence Medical Research Program (Grant no. 2018ZHYL0101), Youth Medical Talents-Medical Imaging Practitioner Program (Grant no. 201972) and Shanghai Municipal Commission of Health and Family planning Program (Grant no. 20184Y0037).

References

1. Naidich DP, Rusinek H, McGuinness G, et al. Variables affecting pulmonary nodule detection with computed tomography:

- evaluation with three-dimensional computer simulation. *J Thorac Imaging*. 1993;8:291–9.
2. Del CA, Paola F, Andrea C, et al. Missed lung cancer: when, where, and why? *Diagn Interv Radiol*. 2017;23(2):118–26.
 3. Tang W, Wang JW, Wu N, et al. Computer-aided detection of nodule in low-dose CT screening for lung cancer. *Chin J Radiol*. 2012;46(7):619–23.
 4. Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology*. 2004;230:347–52.
 5. Beigelman-Aubry C, Raffy P, Yang W, et al. Computer-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. *AJR Am J Roentgenol*. 2007;189:948–55.
 6. Jeon KN, Goo JM, Lee CH, et al. Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening computed tomography. *Invest Radiol*. 2012;47:457–61.
 7. Sahiner B, Chan HP, Hadjiiski LM, et al. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: analysis of an observer performance study by nodule size. *Acad Radiol*. 2009;16:1518–30.
 8. Cai JL, Xu DM, Liu SY, Cham MD. The added value of computer-aided detection of small pulmonary nodules and missed lung cancers. *J Thorac Imaging*. 2018;33:390–5.
 9. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. *Inf Process Med Imaging*. 2015;24:588–99.
 10. Ciompi F, Chung K, Van Riel SJ, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep*. 2017;7:464–79.
 11. Causey JL, Zhang J, Ma S, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans. *Sci Rep*. 2018;8(1):9286.
 12. Liu K, Li Q, Ma JC, et al. Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance. *Radiol Artif Intell*. 2019;1(3):e180084.
 13. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. *Medical image computing and computer-assisted intervention – MICCAI 2015, Lecture notes in computer science*, vol. 9351. Springer, Cham (2015)
 14. He, K, Zhang, X, Ren, et al. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR) 2016, 770–778.
 15. Funama Y, Awai KD, Oda S, et al. Detection of nodules showing ground-glass opacity in the lungs at low-dose multidetector computed tomography: phantom and clinical study. *J Comput Assist Tomogr*. 2009;33(1):49–53.
 16. Ohno Y, Yaguchi A, Okazaki T, et al. Comparative evaluation of newly developed model-based and commercially available hybrid-type iterative reconstruction methods and filter back projection method in terms of accuracy of computer-aided volumetry (CADv) for low-dose CT protocols in phantom study. *Eur J Radiol*. 2016;85:1375–82.
 17. McNitt-Gray MF. AAPM/RSNA physics tutorial for residents: topics in CT-radiation dose in CT. *Radiographics*. 2002;22(6):1541–53.
 18. Wielpütz Mark O, Wroblewski J, et al. Computer-aided detection of artificial pulmonary nodules using an ex vivo lung phantom: influence of exposure parameters and iterative reconstruction. *Eur J Radiol*. 2015;84(5):1005–11.
 19. Suzuki K. A review of computer-aided diagnosis in thoracic and colonic imaging. *Quant Imaging Med Surg*. 2012;2:163–76.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Qiong Li¹  · Qing-chu Li¹ · Rui-ting Cao² · Xiang Wang¹ · Ru-tan Chen¹ · Kai Liu¹ · Li Fan¹ · Yi Xiao¹ · Zi-tian Zhang² · Chi-Cheng Fu² · Qiong Song² · Weiping Liu² · Qu Fang² · Shi-yuan Liu¹

Qiong Li
liqiongsmmu2008@qq.com

Qing-chu Li
40450409@qq.com

Rui-ting Cao
caort@fosun.com

Xiang Wang
7098487@qq.com

Ru-tan Chen
78169184@qq.com

Kai Liu
469884318@qq.com

Li Fan
fanli0930@163.com

Yi Xiao
xiaoyi@188.com

Zi-tian Zhang
zhangzt@fosun.com

Chi-Cheng Fu
fuqz@fosun.com

Qiong Song
songqiong20110806@fosun.com

Weiping Liu
liuwp@fosun.com

Qu Fang
fangqu@fosun.com

¹ Department of Radiology, Changzheng Hospital, Second Military Medical University, NO.415, Fengyang Road, Shanghai 200003, China

² Shanghai Aitrox Technology Corporation Limited, Shanghai, China

Received:
15 August 2018

Revised:
29 October 2018

Accepted:
01 November 2018

<https://doi.org/10.1259/bjr.20180726>

Cite this article as:

Pickhardt PJ, Lee SJ, Liu J, Yao J, Lay N, Graffy PM, et al. Population-based opportunistic osteoporosis screening: Validation of a fully automated CT tool for assessing longitudinal BMD changes. *Br J Radiol* 2019; **91**: 20180726.

FULL PAPER

Population-based opportunistic osteoporosis screening: Validation of a fully automated CT tool for assessing longitudinal BMD changes

¹PERRY J. PICKHARDT, MD, ¹SCOTT J. LEE, MD, ²JIAMIN LIU, PhD, ²JIANHUA YAO, PhD, ²NATHAN LAY, PhD, ¹PETER M GRAFFY, BA and ²RONALD M SUMMERS, MD, PhD

¹Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

²Department of Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD, USA

Address correspondence to: Perry J. Pickhardt
E-mail: ppickhardt2@uwhealth.org

Objective: To validate a fully automated CT-based spinal trabecular bone mineral density (BMD) tool and apply it to a longitudinal screening cohort.

Methods: The automated BMD tool was retrospectively applied to non-contrast abdominal CT scans in 1603 consecutive asymptomatic adults (mean age, 55.9 years; 770 M/833 F) undergoing longitudinal screening (mean interval, 5.7 years; range, 1.0–12.3 years). The spinal column was automatically segmented, with standardized L1 and L2 anterior trabecular ROI placement. Automated and manual L1 HU values were compared, as were automated supine-prone measures. L1–L2 CT attenuation values were converted to BMD values through a linear regression model. BMD values and changes were assessed according to age and gender.

Results: Success rate of the automated BMD tool was 99.8 % (four failed cases). Both automated supine vs prone and manual vs automated L1 attenuation measurements showed good agreement. Overall mean annual rate of bone loss was greater in females than males (–2.0% vs –1.0%), but the age-specific rate declined

faster in females from age 50 (–2.1%) to age 65 (–0.3%) compared with males (–0.9% to –0.5%). Mean BMD was higher in females than males at age 50 (143.6 vs 135.1 mg cm^{–3}), but post-menopausal bone loss in females reversed this relationship beyond age 60. By age 70, mean BMD in females and males was 100.8 and 107.7 mg cm^{–3}, respectively.

Conclusion: This robust, fully automated CT BMD tool allows for both individualized and population-based assessment. Mean BMD was lower in men than women aged 50–60, but accelerated post-menopausal bone loss in women resulted in lower values beyond age 60.

Advances in knowledge: This fully automated tool can be applied to routine abdominal CT scans for prospective or retrospective opportunistic BMD assessment, including change over time. Mean BMD was lower in men compared with women aged 50–60 years, but accelerated bone loss in women during this early post-menopausal period resulted in lower BMD values for women beyond age 60.

INTRODUCTION

Osteoporosis is an important yet underdiagnosed and undertreated bone disease whose direct and indirect costs are estimated to substantially increase in the coming decades.^{1–4} As such, efforts to increase osteoporosis screening and treatment are important for reducing the impending global public health burden of fragility fractures. Dual-energy X-ray absorptiometry (DXA) is the most widely used screening tool for low bone mineral density (BMD), but its planar nature precludes direct trabecular assessment of the spine. An additional opportunity for osteoporosis screening exists using bone data obtained from body CT scans, which are frequently performed in

older adults for a wide variety of indications.^{5–10} In particular, manual CT-based L1 trabecular attenuation measurements have been shown to correlate with DXA-based BMD categories, is predictive of future fragility fractures, and this level is imaged on both thoracic and abdominal CT scans.^{11,12}

Advances in radiology image processing have the potential to provide fully automated measurements of CT-based images that are more objective than manual human measures, and can be applied to larger patient cohorts.^{13–15} In this study, we implemented an image-processing algorithm that performs automatic volumetric segmentation of the lumbar trabecular space on CT scans and then estimates

BMD. Utilizing a large, unique CT-based adult screening cohort undergoing longitudinal colonography assessment,¹⁶ we sought to validate this BMD tool, including both measurement fidelity and objective population-based assessment according to age and gender over time.

METHODS AND MATERIALS

Patient population

This was a HIPAA-compliant retrospective cohort study performed at a single academic medical center. The need for patient consent was waived by the IRB. Asymptomatic, generally healthy outpatient adults aged 50–70 years at the time of their initial screening CT colonography (CTC) study, and who had at least two CTC studies over time at the University of Wisconsin between January 2004 and March 2017, were eligible for inclusion. For patients with more than one follow-up CTC, the initial and last scans were utilized for primary analysis. Patients whose scans were performed less than 1 year apart were excluded as relevant decreases in BMD would be less likely over such a short follow-up period. A total of 1607 patients were initially eligible for inclusion. We also searched the electronic medical records of included patients for evidence of osteoporosis screening with central DXA (spine and/or hips), either before or after their initial CT examination used in this study.

Age at initial CT, length of time between CT studies, and gender were collected as exposure variables. After excluding four failed cases, the final cohort consisted of 1603 adults (833 females and 770 males), with a mean age at initial CT of 55.9 ± 5.0 years. The mean interval to follow-up CT was 5.7 ± 1.9 years (range, 1.0–12.3 years). Outcome variables were the automated CT-derived L1 trabecular attenuation values [in Hounsfield units (HU)] and the calculated BMD values, which were based on the average trabecular HU at L1 and L2.

CT image acquisition

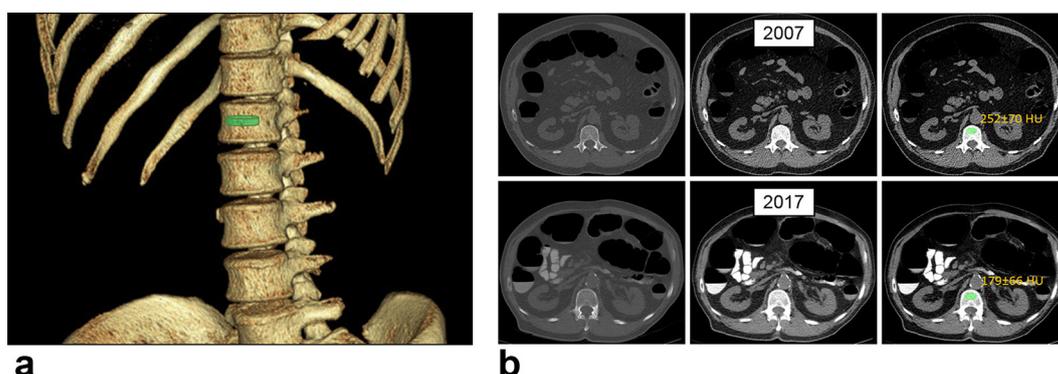
Per standard CTC screening technique,^{17,18} breath-hold supine and prone acquisitions were performed without i.v. contrast on a variety of 8–64 channel multidetector CT scanners (GE Healthcare, Waukesha, WI). These CT scanners undergo daily QA, including HU calibration testing. Scan parameters included 1.25 mm collimation and slice thickness, 120 kV, variable low-dose mAs settings, and 1 mm reconstruction interval. The 1.25 mm supine series was utilized for primary BMD analysis, include rate of change over time; prone series BMD measurements were also obtained for intra patient comparison.

Automated trabecular spine segmentation and BMD estimation

A fully automated method was developed to segment the trabecular spine from CT images, which was developed and tested on a separate patient cohort.¹⁹ The CT images were sent offline from PACS for anonymization prior to accessing the BMD tool. The algorithm first extracts the spinal column based on thresholding, region growing, watershed, and directed graph search. Second, the segmented spinal column is then partitioned into individual vertebrae using curved planar reformation and dynamic programming. For each vertebral body, the cortical bone was automatically eroded to isolate the trabecular space.

An oval region of interest (ROI) was automatically placed in the middle transverse (axial) cross-sections of extracted L1 and L2 vertebrae (Figure 1). The location of the ROI was set at the anterior one-third of the anteroposterior center axis of the vertebral body, to simulate the established manual method.^{11,20,21} The size of the ROI was standardized at one-half of the vertebral body lateral width by one-fourth of the vertebral body anteroposterior height, using a single slice at the vertebral midline. Mean HU values were calculated within the ROIs for L1 and L2 and converted to BMD estimation through a linear regression model trained by a set of QCT images with calibration phantoms.

Figure 1. Automated CT BMD tool in asymptomatic 59-year-old male at initial evaluation, evaluated over a 10-year interval. (A) 3D volume-rendered image from the initial CT scan performed in 2007 when the patient was 59 years old shows the relative placement of the automated ROI (green cylinder) within the anterior trabecular space of the L1 vertebral body. (B) Collage of 2D transverse images at the L1 level from CT in 2007 (top row) and 2017 (bottom row). Bone windows (left images) and soft tissue windows (middle images) are shown, as well as placement of the automated ROI (green ovals) with resulting mean HU values (right images). The HU values correspond to an estimated BMD of 186.9 mg cm^{-3} in 2006 and 131.6 mg cm^{-3} in 2016, reflecting a 30% loss, or annual mean rate of BMD change of $-3\%/year$. Note also the significant interval weight loss. 3D, three-dimensional; BMD, bone mineral density; ROI, region of interest.



To estimate BMD from HU, we used a previously derived calibration curve obtained from phantoms used in dedicated QCT scans, which demonstrated excellent correlation ($r^2 = 0.98$).¹⁹ The curve maps CT attenuation in HUs to BMD in milligrams per milliliter.

Comparison of manual and automated trabecular attenuation values

To further assess the automated L1 HU measurement, we compared manual L1 attenuation measurements in a random subset of 588 patients, using a previously validated method, with placement of a single ROI in the anterior L1 trabecular space, just off the mid-vertebral level.^{8,11,21}

Statistical analysis

Annual rate of change in BMD was assessed in overall and percentage-based terms. The rate of annual BMD change was estimated by calculating the difference between initial and follow-up BMD and dividing it by the scan interval in years (Figure 1). The full patient range at initial CT (50–70 years) was utilized for age-specific BMD values; age-specific rates of BMD loss were considered for the 50–65 year-old range, as there were insufficient data points for rates of change beyond 65 years.

Inpatient comparison between automated L1 HU values on the supine and prone CT series were compared as an internal quality assurance measure in a subset of 2851 CT scans. For all other purposes, BMD values for the supine CT series were utilized. Multivariate linear regression was performed to estimate the effects of age and gender on annual change in BMD and to model expected annual change in BMD. Interaction of the effects of age and gender on outcome variables were included in the regression analyses. Bland-Altman analysis was utilized for comparing manual and automated L1 attenuation measurement, and for comparing automated supine and prone BMD measurements. In addition, the square of the Pearson correlation coefficient (r^2 , the correlation of determination) was derived for these

comparisons. All statistical analyses were performed using R software (R Core Development Team, v3.3.2).

RESULTS

QA performance of the BMD algorithm

The automated BMD tool successfully derived an L1 HU and BMD value in 99.8% (1603/1607) of cases. There was good agreement ($r^2 = 0.90$) between the automated supine and prone BMD measurements (Figure 2), with a mean difference of -0.15 mg cm^{-3} , without bias. Bland-Altman 95% limits of agreement were -21.7 to 21.4 mg cm^{-3} . In the subset of 588 patients, supine manual and automated L1 attenuation measurements also showed good linear agreement (Figure 3; $r^2 = 0.80$). The Bland-Altman 95% limits of agreement for the attenuation difference were -32.8 to 22.4 HU . Automated L1 attenuation measurements were slightly increased relative to manual measurements by 5.2 HU on average, without bias.

BMD change over time at CT

Overall results for the entire cohort, not stratified by age, are shown on Table 1. The overall mean annual change in BMD was -2.0% per year in females and -1.0% per year in males (-1.5% per year for the entire cohort). However, a deeper understanding is gained when the results are stratified by age and gender (Table 2). As expected, mean BMD decreased with increasing age from 50 to 70 years for both males and females (Figure 4). Mean BMD in females was higher than males at 50 years of age (143.6 vs 135.1 mg cm^{-3}), but accelerated post-menopausal bone loss in females reversed this relationship beyond 60 years of age. However, the rate of bone loss progressively declined more rapidly with increasing age in females compared with males (Figure 5). By 65 years of age, the mean rate of annual bone loss was slightly greater in males compared with females ($-0.5\%/ \text{year}$ vs $-0.3\%/ \text{year}$). By 70 years of age, mean BMD in females and males was 100.8 and 107.7 mg cm^{-3} , respectively. Age at initial CT, sex, and their interaction term were all significantly associated ($p < 0.001$) with mean BMD at initial CT, annual mean rate of change

Figure 2. Automated supine vs prone measurements (A), Plot of automated BMD measurement of the supine vs prone CT series ($r^2 = 0.90$) (B), Corresponding Bland-Altman plot with 95% limits of agreement. No bias is noted. BMD, bone mineral density.

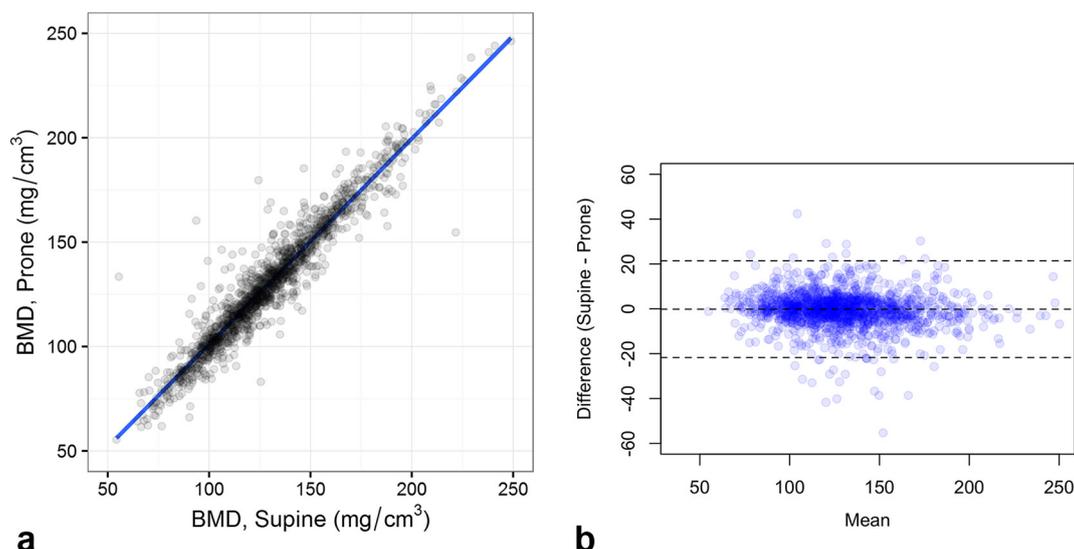
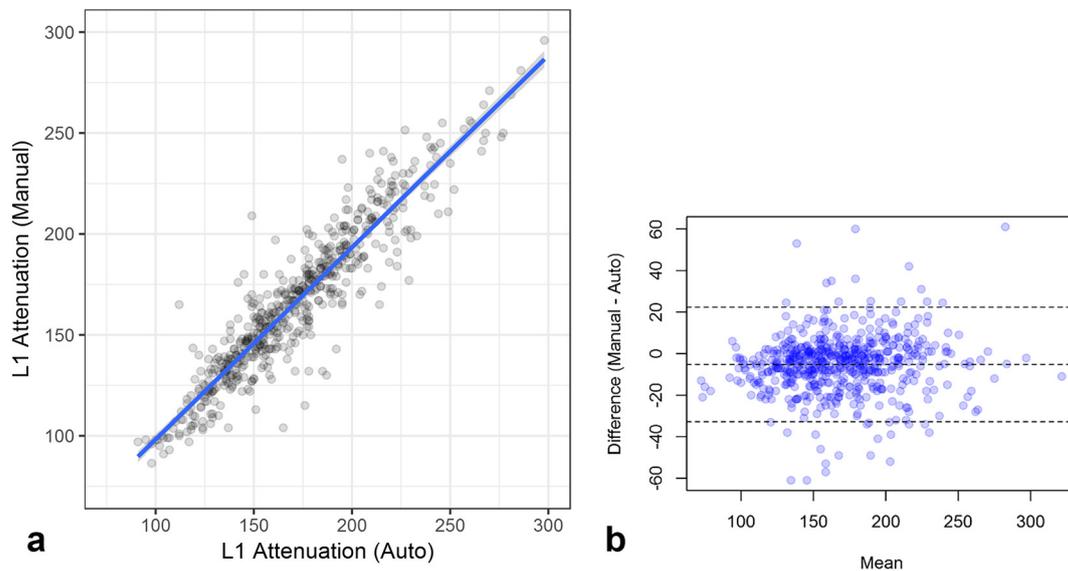


Figure 3. Manual vs automated L1 trabecular attenuation measurements in 588 subjects. (A) Plot of manual single-slice ROI and automated volumetric measurements for L1 trabecular attenuation (in HU). ($r^2 = 0.80$). (B) Corresponding Bland-Altman plot with 95% limits of agreement. Note that the automated measures are approximately 5 HU higher on average, without bias according to variation of the mean. HU, Hounsfield unit; ROI, region of interest.



in BMD, and annual percent change in BMD at multiple linear regression modeling.

Clinical osteoporosis screening with DXA

A total of 377 patients underwent DXA for osteoporosis screening prior to their initial CT screening study, corresponding to 23.5% (377/1603) of the cohort. Another 17.3% (278/1603) of patients underwent DXA study after the date of the initial CT study. Nearly 90% of patients screened with DXA were females. To date, 59.1% (948/1603) of patients from this study have not been screened by DXA.

DISCUSSION

In this study, we determined annual age- and gender-specific rates of trabecular spine BMD change in an adult screening cohort using fully automated CT-based software. We found that annual rates of BMD loss peaked in females ages 50–55 years, corresponding to the early post-menopausal period (mean age of menopause in females is 51 years). This is in agreement with prior DXA²² and qCT^{23–25} data that demonstrated accelerated bone loss in females during late perimenopause and early post-menopause. Prior to this accelerated post-menopausal BMD loss in females, we found that age-matched males actually

Table 1. Characteristics for overall patient cohort

		Female	Male	All
	N=	833	770	1603
Age at initial CT (years)	Mean (SD)	56.0 (5.0)	55.8 (5.0)	55.9 (5.0)
CT Interval (years)	Mean (SD)	5.8 (1.8)	5.7 (1.9)	5.7 (1.9)
L1 Attenuation at Initial CT (HU)	Mean (SD)	177.4 (40.8)	171.1 (39.5)	174.3 (40.3)
BMD at Initial CT (mg cm ⁻³)	Mean (SD)	130.7 (30.5)	127.5 (28.5)	129.1 (29.6)
	Min	55.4	54.3	54.3
	Max	253.9	253.1	253.9
BMD at follow-up CT (mg cm ⁻³)	Mean (SD)	119.1 (29.0)	121.6 (28.7)	120.3 (28.8)
	Min	39.5	58.1	39.5
	Max	237.22	244.8	244.8
Change in BMD (mg cm ⁻³)	Mean (SD)	-11.6 (17.2)	-5.6 (12.5)	-8.7 (15.5)
% Annual Change in BMD	Mean	-2.0%	-1.0%	-1.5%

BMD, bone mineral density; HU, Hounsfield Unit; SD, standard deviation.

Table 2. Mean BMD values and rates of change according to gender and age

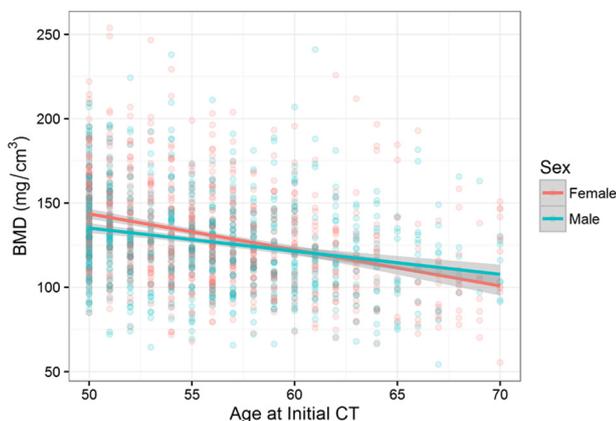
Age at initial CT	Sex	Mean BMD (mg cm ⁻³)	Mean rate of annual change in BMD (mg cm ⁻³ /year)	% Rate of annual change in BMD
50	Female	143.6 (140.6, 146.5)	-3.1 (-3.4, -2.8)	-2.1% (-2.3, -1.9)
55	Female	132.9 (130.9, 134.8)	-2.2 (-2.4, -2.0)	-1.5% (-1.5, -1.7)
60	Female	122.2 (119.8, 124.6)	-1.3 (-1.6, -1.1)	-0.9% (-1.1, -0.8)
65	Female	111.5 (107.7, 115.4)	-0.4 (-0.8, -0.1)	-0.3% (-0.7, -0.09)
70	Female	100.8 (95.3, 106.4)	-	-
50	Male	135.1 (132.1, 138.1)	-1.2 (-1.5, -0.9)	-0.9% (-1.1, -0.6)
55	Male	128.3 (126.3, 130.3)	-1.0 (-1.2, -0.8)	-0.8% (-0.2, -0.6)
60	Male	121.4 (118.9, 124.0)	-0.8 (-1.0, -0.5)	-0.6% (-0.8, -0.4)
65	Male	114.6 (110.5, 118.7)	-0.6 (-1.0, -0.2)	-0.5% (-0.8, -0.2)
70	Male	107.7 (101.8, 113.6)	-	-

BMD, bone mineral density.

have lower lumbar trabecular BMD than females, on average. Although not widely appreciated, this lower BMD level in males has been previously shown for adults in the third decade of life.²⁶ We also found that rates of BMD loss progressively slowed after menopause in females, which is also in accordance with previously published results. In combination, these data stress the potential importance and possible preventive opportunity for this accelerated BMD loss seen in the early post-menopausal period in females. Because these changes occur before the age of recommended osteoporosis screening, other interventions might be considered.

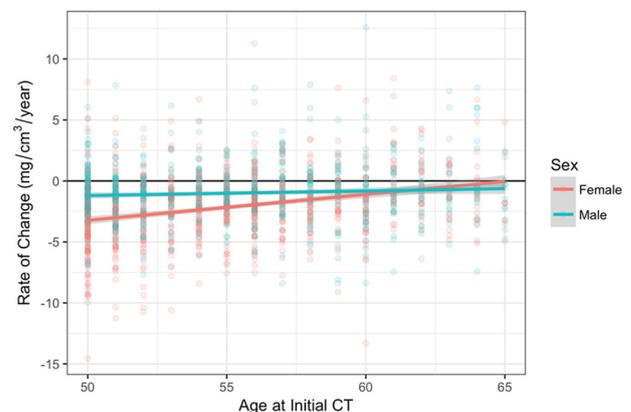
Perhaps more important than redemonstrating normative BMD levels and rates of BMD loss according to age and gender is the fact that all our CT-based measurements were obtained automatically without reader input. This robust automated tool, representing an improvement over an earlier version,¹⁹ demonstrated a very low failure rate, and correlated well with

Figure 4. Automated mean BMD values at initial CT according to age and gender. Plots using lines of best fit show that males have lower BMD values on average in the 50-60 age range, whereas females have lower values on average at older ages. The shaded areas around lines correspond to the 95% CIs for the mean BMD values. BMD, bone mineral density.



the manual L1 trabecular ROI measurements. Manual L1 HU measures have been previously validated for clinical important outcomes,^{6,8,11,12,20} and are now being used every day in our routine clinical practice to provide opportunistic BMD information. Use of an automated tool could greatly expand utilization of this opportunistic screening approach, and enhance the value of our CT interpretations. In addition, this tool could be applied retrospectively within a group or practice to provide population-based screening. As we have shown, there is a substantial subset of patients who may otherwise not be screened by DXA, and opportunistic screening with CT could therefore have a meaningful impact. Although one could argue that many of these patients were not yet of the recommended age for osteoporosis screening, many osteoporotic fractures occur prior to initial DXA screening. Furthermore, the majority of patients undergoing DXA who have a fragility fracture will have a non-osteoporotic T-score (*i.e.* greater than -1.0),¹¹ which may relate in part to the planar nature of DXA. As a volumetric cross-sectional imaging technique, CT is able to directly assess the trabecular bone without the issues of overlying cortical bone and degenerative changes.

Figure 5. Annual rates of BMD change according to age at initial CT and gender. BMD, bone mineral density.



In general, rapid advances in artificial intelligence and machine learning applied to cross-sectional imaging could eventually lead to fully automated measurements that provide objective assessment to complement a radiologist's interpretation.^{13–15} Beyond this BMD tool, we are currently investigating a number of abdominal CT-based machine learning algorithms to automatically measure abdominal aortic calcification, hepatic attenuation, muscle mass for sarcopenia, and visceral and subcutaneous fat. While potentially providing useful information in isolation, the combination of these tools could conceivably synergize to provide meaningful cardiometabolic risk assessment and stratification. In addition, a companion machine learning algorithm for automatically detecting vertebral compression fractures was not applied in this study but could be used in concert with automated BMD assessment.²⁷

One strength of our study design was the relatively unique longitudinal CT series in an asymptomatic screening cohort, which was ideal for applying this automated tool to assess for age- and gender-based differences in BMD values and rates of BMD loss. In addition, CT has intrinsic advantages over the clinical standard of DXA, which is a planar technique that cannot directly assess the spinal trabecular space. However, we also acknowledge limitations of the current study. The small but measurable increased offset in automated L1 attenuation over 5 HU over the manual technique likely relates to placement of the automated ROI in the central aspect of the anterior trabecular space, where a stripe of increased linear density can be seen on sagittal CT reconstructions and at gross anatomy.²⁸ In addition, one would expect some offset in HU measurements between supine and prone measurements given the known effects related to positional changes within the scanner.²⁹ The phantomless technique with calibration on one scanner in a limited data set could potentially lead to inaccuracies or variability in the conversions from HU to mg/cc. However, HU values are calibrated as part of daily quality control, and a recent study found that phantomless

BMD results based on HU values represent good estimates across different scanners.³⁰ We did not investigate the impact of other risk factors, such as those included in the FRAX tool,³¹ nor did we search for prevalent or incident fractures, all of which would require a detailed patient-by-patient search. We plan to investigate the predictive ability of this automated BMD tool by applying it to a large retrospective cohort both with and without future fragility fractures.

In conclusion, this fully automated BMD tool can be applied opportunistically to routine abdominal CT scans for prospective clinical or retrospective population-based assessments, including monitoring changes over time.

Numbers in parentheses represent 95% confidence intervals. Accelerated bone loss in the early post-menopausal period in females leads to overall lower BMD values beyond age 60. Note, however, that the rate of BMD loss progressively decreases faster in females than in males, resulting a lower rate of BMD loss by age 65. Data for rates of change beyond age 65 were too sparse for accurate reporting.

Plots using lines of best fit show that females demonstrate accelerated bone loss over the post-menopausal 50–60 year age range, but that this rate of loss progressively declines faster than males. Shaded areas around lines correspond to 95% CIs for the mean annual rate of BMD change.

ACKNOWLEDGMENT

This research was supported in part by the Intramural Research Program of the National Institutes of Health Clinical Center.

DISCLOSURE

Dr Pickhardt serves as an advisor to Bracco and is a shareholder in SHINE and Elucen.

REFERENCES

- Graffy PM, Lee SJ. A crisis in the treatment of osteoporosis. *J Bone Miner Res* 2016; **31**: 1485–7. doi: <https://doi.org/10.1002/jbmr.2888>
- Black DM, Rosen CJ, Osteoporosis P. Clinical practice. postmenopausal osteoporosis. *N Engl J Med* 2016; **374**: 254–62. doi: <https://doi.org/10.1056/NEJMcp1513724>
- Wright NC, Looker AC, Saag KG, Curtis JR, Delzell ES, Randall S, et al. The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine. *J Bone Miner Res* 2014; **29**: 2520–6. doi: <https://doi.org/10.1002/jbmr.2269>
- Siris ES, Adler R, Bilezikian J, Bolognese M, Dawson-Hughes B, Favus MJ, et al. The clinical diagnosis of osteoporosis: a position statement from the National Bone Health Alliance Working Group. *Osteoporos Int* 2014; **25**: 1439–43. doi: <https://doi.org/10.1007/s00198-014-2655-z>
- Pooler BD, Kim DH, Pickhardt PJ. Extracolonic findings at screening CT colonography: Prevalence, benefits, challenges, and opportunities. *AJR Am J Roentgenol* 2017; **209**: 94–102. doi: <https://doi.org/10.2214/AJR.17.17864>
- Lee SJ, Pickhardt PJ. Opportunistic screening for osteoporosis using body CT scans obtained for other Indications: the UW experience. *Clin Rev Bone Miner Metab* 2017; **15**: 128–37. doi: <https://doi.org/10.1007/s12018-017-9235-7>
- Lee SJ, Anderson PA, Pickhardt PJ. Predicting future hip fractures on routine abdominal CT using opportunistic osteoporosis screening measures: A matched case-control study. *AJR Am J Roentgenol* 2017; **209**: 395–402. doi: <https://doi.org/10.2214/AJR.17.17820>
- Graffy PM, Lee SJ, Ziemlewicz TJ, Pickhardt PJ. Prevalence of vertebral compressor fractures on routine CT scans according to L1 trabecular attenuation: determining Relevant thresholds for opportunistic osteoporosis screening. *AJR Am J Roentgenol* 2017; **209**: 491–6. doi: <https://doi.org/10.2214/AJR.17.17853>

9. Ziemlewicz TJ, Maciejewski A, Binkley N, Brett AD, Brown JK, Pickhardt PJ. Opportunistic quantitative CT bone mineral density measurement at the proximal femur using routine contrast-enhanced scans: direct comparison with DXA in 355 adults. *J Bone Miner Res* 2016; **31**: 1835–40. doi: <https://doi.org/10.1002/jbmr.2856>
10. Ziemlewicz TJ, Binkley N, Pickhardt PJ. Opportunistic osteoporosis screening: addition of quantitative CT bone mineral density evaluation to CT colonography. *J Am Coll Radiol* 2015; **12**: 1036–41. doi: <https://doi.org/10.1016/j.jacr.2015.04.018>
11. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med* 2013; **158**: 588–95. doi: <https://doi.org/10.7326/0003-4819-158-8-201304160-00003>
12. Lee SJ, Graffy PM, Zea RD, Ziemlewicz TJ, Pickhardt PJ. Future osteoporotic fracture risk related to lumbar vertebral trabecular attenuation measured at routine body CT. *J Bone Miner Res* 2018; **33**: 860–7. doi: <https://doi.org/10.1002/jbmr.3383>
13. Erickson BJ. Machine learning: Discovering the future of medical imaging. *J Digit Imaging* 2017; **30**: 391. doi: <https://doi.org/10.1007/s10278-017-9994-1>
14. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 2017; **208**: 754–60. doi: <https://doi.org/10.2214/AJR.16.17224>
15. Summers RM. Progress in fully automated abdominal CT interpretation. *AJR Am J Roentgenol* 2016; **207**: 67–79. doi: <https://doi.org/10.2214/AJR.15.15996>
16. Pickhardt PJ, Pooler BD, Mbah I, Weiss JM, Kim DH. Colorectal findings at repeat CT colonography screening after initial CT colonography screening negative for polyps larger than 5 mm. *Radiology* 2017; **282**: 139–48. doi: <https://doi.org/10.1148/radiol.2016160582>
17. Pickhardt PJ. Screening CT colonography: how I do it. *AJR Am J Roentgenol* 2007; **189**: 290–8. doi: <https://doi.org/10.2214/AJR.07.2136>
18. Pickhardt PJ. Imaging and screening for colorectal cancer with CT colonography. *Radiol Clin North Am* 2017; **55**: 1183–96. doi: <https://doi.org/10.1016/j.rcl.2017.06.009>
19. Summers RM, Baecher N, Yao J, Liu J, Pickhardt PJ, Choi JR, et al. Feasibility of simultaneous computed tomographic colonography and fully automated bone mineral densitometry in a single examination. *J Comput Assist Tomogr* 2011; **35**: 212–6. doi: <https://doi.org/10.1097/RCT.0b013e3182032537>
20. Lee SJ, Binkley N, Lubner MG, Bruce RJ, Ziemlewicz TJ, Pickhardt PJ. Opportunistic screening for osteoporosis using the sagittal reconstruction from routine abdominal CT for combined assessment of vertebral fractures and density. *Osteoporos Int* 2016; **27**: 1131–6. doi: <https://doi.org/10.1007/s00198-015-3318-4>
21. Pickhardt PJ, Lee LJ, del Rio AM, Lauder T, Bruce RJ, Summers RM, et al. Simultaneous screening for osteoporosis at CT colonography: bone mineral density assessment using MDCT attenuation techniques compared with the DXA reference standard. *J Bone Miner Res* 2011; **26**: 2194–203. doi: <https://doi.org/10.1002/jbmr.428>
22. Finkelstein JS, Brockwell SE, Mehta V, Greendale GA, Sowers MR, Ettinger B, et al. Bone mineral density changes during the menopause transition in a multiethnic cohort of women. *J Clin Endocrinol Metab* 2008; **93**: 861–8. doi: <https://doi.org/10.1210/jc.2007-1876>
23. Riggs BL, Melton LJ, Robb RA, Camp JJ, Atkinson EJ, McDaniel L, et al. A population-based assessment of rates of bone loss at multiple skeletal sites: evidence for substantial trabecular bone loss in young adult women and men. *J Bone Miner Res* 2008; **23**: 205–14. doi: <https://doi.org/10.1359/jbmr.071020>
24. Kalender WA, Felsenberg D, Louis O, Lopez P, Klotz E, Osteaux M, et al. Reference values for trabecular and cortical vertebral bone density in single and dual-energy quantitative computed tomography. *Eur J Radiol* 1989; **9**: 75–80.
25. Cann CE, Genant HK, Kolb FO, Ettinger B. Quantitative computed tomography for prediction of vertebral fracture risk. *Bone* 1985; **6**: 1–7. doi: [https://doi.org/10.1016/8756-3282\(85\)90399-0](https://doi.org/10.1016/8756-3282(85)90399-0)
26. Riggs BL, Melton Iii LJ, Robb RA, Camp JJ, Atkinson EJ, Peterson JM, et al. Population-based study of age and sex differences in bone volumetric density, size, geometry, and structure at different skeletal sites. *J Bone Miner Res* 2004; **19**: 1945–54. doi: <https://doi.org/10.1359/jbmr.040916>
27. Burns JE, Yao J, Summers RM. Vertebral body compression fractures and bone density: automated detection and classification on CT images. *Radiology* 2017; **284**: 788–97. doi: <https://doi.org/10.1148/radiol.2017162100>
28. Kennedy OD, Brennan O, Rackard SM, O'Brien FJ, Taylor D, Lee TC. Variation of trabecular microarchitectural parameters in cranial, caudal and mid-vertebral regions of the ovine L3 vertebra. *J Anat* 2009; **214**: 729–35. doi: <https://doi.org/10.1111/j.1469-7580.2009.01054.x>
29. Szczykutowicz TP, DuPlissis A, Pickhardt PJ. Variation in CT Number and Image Noise Uniformity According to Patient Positioning in MDCT. *AJR Am J Roentgenol* 2017; **208**: 1064–72. doi: <https://doi.org/10.2214/AJR.16.17215>
30. Lee DC, Hoffmann PF, Kopperdahl DL, Keaveny TM. Phantomless calibration of CT scans for measurement of BMD and bone strength-Inter-operator reanalysis precision. *Bone* 2017; **103**: 325–33. doi: <https://doi.org/10.1016/j.bone.2017.07.029>
31. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int* 2008; **19**: 385–97. doi: <https://doi.org/10.1007/s00198-007-0543-5>



Radiomic features based on MRI for prediction of lymphovascular invasion in rectal cancer

Yu Fu¹ · Xiangchun Liu¹ · Qi Yang¹ · Jianqing Sun² · Yunming Xie¹ · Yiyi Zhang¹ · Huimao Zhang¹

Received: 24 June 2019 / Revised: 23 August 2019 / Accepted: 11 October 2019 / Published online: 7 November 2019
© Springer Nature Singapore Pte Ltd. 2019

Abstract

Purpose To investigate the value of radiomics in predicting lymphovascular invasion (LVI) status of rectal cancer based on MRI.

Materials and methods The retrospective study included 188 patients based on MRI with histologically confirmed rectal cancer and evaluated LVI status. Clinical factors and image data were collected, and radiomics features were extracted from multi-region (tumor and mesorectum) and single region (tumor), respectively, on T2WI and DWI. Spearman correlation analysis and the LASSO algorithm were used for radiomic feature extraction and selection; preliminarily selection of an optimal classifier by the results of the fivefold cross-validation performance in the six preselected specific machine learning classifier. Multi-regional and single-regional predictive models were both built and evaluated by calculating the area under the ROC curve (AUC) and corresponding accuracy, specificity, sensitivity, etc.

Results A Ridge Classification model was constructed with 21 features (2 clinical features, 10 radiomics features from mesorectum region, and 9 radiomics features from tumor region) selected by Spearman correlation and LASSO analysis. The multi-regional model shows a good performance in the differentiation of the status of LVI in training data sets (AUC = 0.87, accuracy = 0.79). The model was further validated in the testing data sets, giving an AUC and an accuracy of 0.74 and 0.68, respectively. Furthermore, the performance of single-regional model (AUC = 0.72, accuracy = 0.67) is lower compared to the values given by the multi-regional model.

Conclusion The radiomics model which we developed demonstrates that multi-regional radiomics features based on multiparametric MRI are useful for preoperatively predicting lymphovascular invasion in patients with rectal cancer.

Keywords Rectal cancer · Machine learning · Radiomics · Lymphovascular · MRI

Abbreviations

AUC	Area under the ROC curve
EMVI	Extramural venous invasion
LASSO	Least absolute shrinkage and selection operator
LVI	Lymphovascular invasion
MRF	Mesorectal fascia
ROC	Receiver-operating characteristic

Introduction

Colorectal cancer (CRC) is the third most common malignant tumor in the world [1, 2], and about one-third to 44% of CRC are occurred in the rectum [3]. The National Comprehensive Cancer Network (NCCN) Guidelines consider lymphovascular invasion (LVI), which is defined as the presence of tumor cells in the lymphatic vessels or blood vessels or both, as a significant negative factor in treatment options and prognostication in rectal cancer [4, 5]. Several investigations have revealed that patients with LVI may associate with lymph-node metastasis and benefit from adjuvant systemic therapy [6–9]. Hence, it becomes increasingly important to evaluate LVI status preoperatively, so that patients with LVI might benefit from radical surgery and adjuvant treatments [4, 6, 10, 11].

Currently, the LVI status is evaluated by histopathology after resection, which provides no accuracy preoperative

Yu Fu and Xiangchun Liu have contributed equally to this work as the first author.

✉ Huimao Zhang
huimaozhanglinda@163.com

¹ Department of Radiology, The First Hospital of Jilin University, No. 71 Xinmin Street, Changchun, Jilin, China

² Clinical Science Team, Philips (China) Investment Co., Ltd., Shanghai, China

evaluation to a treatment option. The biopsy may provide the LVI status before surgery; however, the limited specimen fails to provide the information of the whole tumor [12, 13]. In addition to histopathology, rectal magnetic resonance imaging (MRI) is also an important means in tumor evaluation, which is noninvasively [3, 4, 14]. Previous studies on rectal MRI have demonstrated that the diagnostic performance of extramural venous invasion (EMVI) by MRI was good [15, 16]. However, the intramural blood vessels and lymphatic vessel invasion status, which are parts of the LVI, are failed to evaluate by MRI [17, 18]. Therefore, is there a way to accurately evaluate the LVI status in rectal cancer before treatment?

Radiomics is an emerging method for extracting quantitative features from medical imaging and assisting clinical decision to improve diagnostic, prognostic, and predictive accuracy [13, 19–21]. The central hypothesis that drives the development of radiomics is based on the tumor microenvironment description, which helps to assess the biological characteristics of the tumor [12, 19]. Rectal MRI is essential for pre- and post-treatment assessment of rectal cancer, as it provides anatomic structures and their relationship with the tumor with a high-spatial resolution [22, 23]. Rectal MRI-based radiomics have been used for treatment response [24], lymph-node metastasis [25], and prognostic evaluation [26]. However, MRI-based radiomics for LVI prediction remains underinvestigated in rectal cancer.

Therefore, the aim of this study was to develop a radiomics model for prediction of lymphovascular invasion in rectal cancer based on MRI.

Materials and methods

Patients

This study was approved by the ethics committee of The First Hospital of Jilin University, and the informed consent requirement was waived. We retrospectively evaluated patients with rectal cancer in our hospital between January 2016 and December 2018. Inclusion criteria were as follows (a) histologically confirmed rectal adenocarcinoma; (b) rectal MRI were performed before surgery within 2 weeks; (c) LVI were assessed by histopathology after resection. The exclusion criteria included a history of (a) preoperative chemoradiotherapy (CRT), radiotherapy, chemotherapy, or distant metastases, considering that the preoperative treatment maybe changed the LVI status; (b) poor MRI quality; (c) lack of clinic information, such as pretreatment carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9(CA19-9). Finally, we enrolled a total of 188 patients, 80 LVI+, and 108 LVI-. We randomly divide patients into training and testing cohorts in a 2:1 ratio.

Clinicopathologic data, including age, gender, the level of carcinoembryonic antigen (CEA), and carbohydrate antigen 19-9 (CA 19-9), were derived from medical records. Laboratory analysis of CEA and CA19-9 was tested within 1 week before surgery. The threshold value for the CEA level was ≤ 5 ng/mL and >5 ng/mL, and the threshold value for the CA19-9 level was ≤ 39 U/mL and >39 U/mL, according to the normal range used in clinics.

Measurement of conventional radiology evaluation indicators: including the location of primary tumor, lesion involvement length, tumor thickness (measured on oblique axis T2WI), extramural depth of invasion (measured on oblique axis T2WI), mesorectal fascia (MRF, > 1 mm diagnostic negative, ≤ 1 mm positive), and maximum lymph-node short diameter (measured on-axis T2WI).

The data enrolled flowchart of the study is shown in Fig. 1.

MRI data acquisition

The enrolled rectal MRIs were all performed on the same MR scanner (3.0T, Philips Ingenia, The Netherlands). And glycerine enema was required for rectal cleansing before scanning. To reduce abdomen motility, 20 mg of anisodamine was injected intramuscularly 30 min before MRI scanning. All patients underwent a rectal MRI protocol including sagittal, axial, oblique axial, and coronal T2-weighted images and DWI. High-resolution T2WI images were obtained using turbo spin-echo with a repetition time (TR) = 3500 ms, echo time (TE) = 100 ms, the field of view (FOV) = 180×180 mm, echo train length = 29, matrix = 288×256 , thickness = 3.0 mm, and gap = 0.3 mm. DWI images were obtained with 2 b factors (0 and 1000 s/mm²), and TR = 2800 ms, TE = 70 ms, FOV = 340×340 mm, matrix = 256×256 , thickness = 4.0 mm, and gap = 1.0 mm. All MRI images were retrieved from the picture archiving and communication system (PACS) for tumor masking and radiomic feature extraction.

Tumor masking

Two radiologists (Dr. Fu and Dr. Liu with 8 and 3 years of experience in rectal cancer radiology diagnosis, respectively) who blinded to the histopathology results segmented the volumes of interest (VOIs) on high-spatial-resolution T2WI and DWI via IntelliSpace Discovery (Philips, Best, The Netherlands). The volumes of interest (VOIs) were defined as follows: (a) the volumes of the whole primary tumor and excluding the intestinal lumen, which was manually drawn on each slice based on T2WI (slightly high signal) and DWI (high signal, b value of 1000 s/mm²), which were drawn along the contour of the

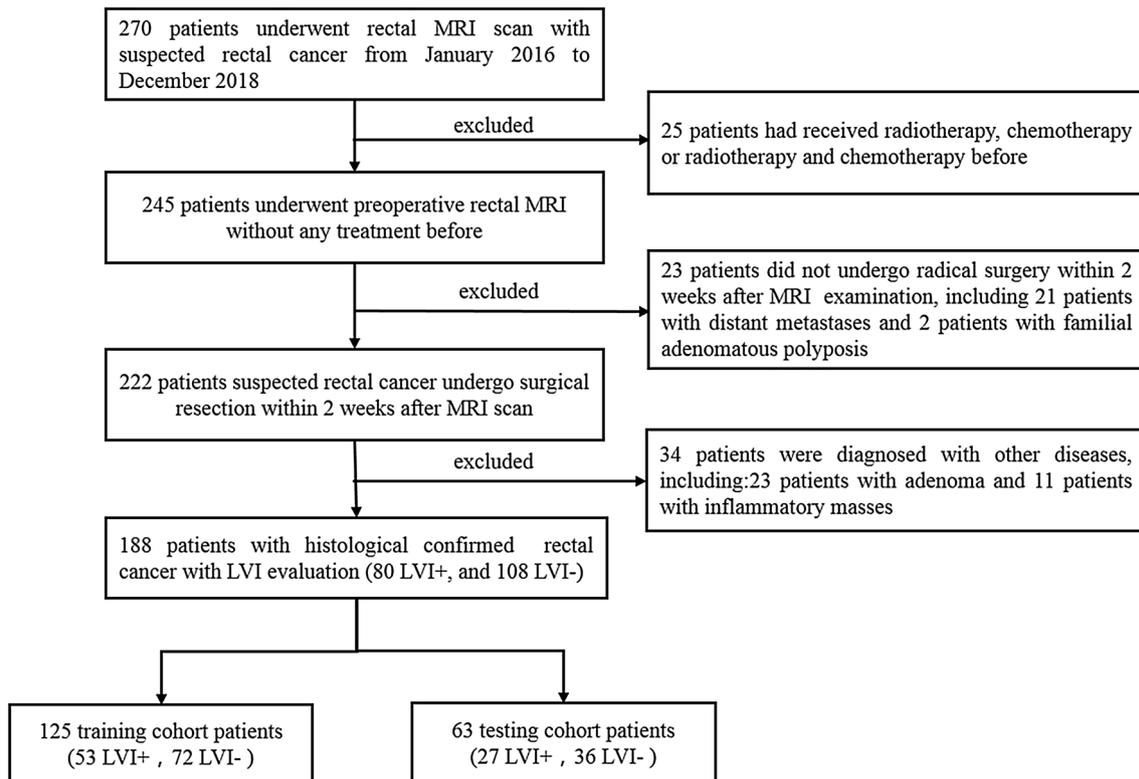


Fig. 1 Flowchart showing the numbers of the included and excluded patients in the study

tumor; (b) the volume of the mesorectal region on unfat suppressed T2WI, which was between the MRF (the thin-low-signal intensity surrounding the mesorectum) and the outer edge of the tumor and rectal wall. The schematic diagram of the tumor and mesorectal region segmentation is shown in Fig. 2.

Radiomic feature extraction and selection

We have used three VOIs in radiomic feature calculation. The radiomic feature was analyzed by Philips Radiomics Tool (Philips Healthcare, China, the core feature calculation is based on pyRadiomics [27]). The extracted features are shown in Table 1.

First, at the feature normalization step, we used the Min–Max scaling algorithm (Eq. 1):

$$X_{\text{normal}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \quad (1)$$

Next, a Spearman correlation analysis of radiomic feature and the label were done. Features with the coefficient lower than absolute value 0.2 or the *p* value greater than 0.05 were removed accordingly because of the low correlation between these radiomic features and pathological labels.

Finally, at the dimensionality reduction step, least absolute shrinkage and selection operator (LASSO) algorithm [28] were used.

Radiomic model construction and evaluation

At the model construction and evaluation step, six linear classification algorithms were investigated, including Passive Aggressive Classifier, Perceptron, Ridge Classifier, SGD Classifier, Logistic Regression, and Linear Support Vector Classifier for training and prediction. First, in the model training stage, we used fivefold cross validation to evaluate the performance of six specific machine learning classifiers in the training cohort with ‘accuracy’ as the optimization metric, preliminarily selected the prediction model with the best prediction performance, and then evaluated the model in the training and test cohorts with the area under the ROC curve (AUC), etc.

Statistical analysis

All statistical analyses were performed using SPSS 24.0 (IBM Corp). Chi-square test was used to compare the differences in categorical variables (gender, the location of primary tumor, the level of CEA and CA19-9, and MRF status), while an independent sample *t* test or

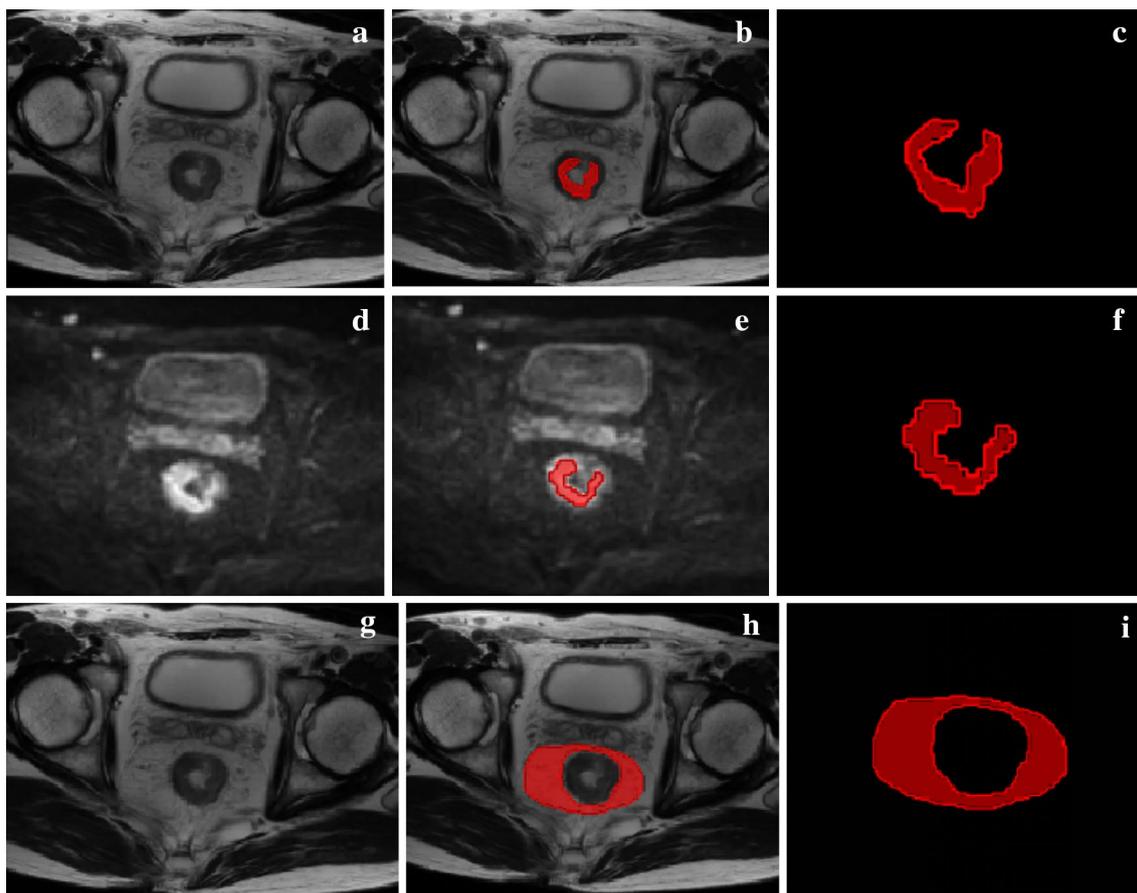


Fig. 2 Segmentation result of tumor and mesorectum on T2WI and DWI in an LVI-positive patient. Images in a 51-year-old male, LVI-positive rectal cancer. **a–c** VOIs of primary tumor on T2WI; **d–f** VOIs of primary tumor on DWI; **g–i** VOIs of mesorectum on T2WI

Table 1 Extracted features by Philips Radiomics software using pyRadiomics

Indexes	Introduction	Feature number
Direct features	Including first-order statistics features, shape-based features, gray-level co-occurrence matrix features, gray-level size zone matrix features, gray-level run-length matrix features, neighbouring gray tone difference matrix features, and gray-level dependence matrix features	105
Indirect features	Calculated based on direct features, through the algorithm of square, square root, logarithm, and exponential	368
Wavelet transform features	Information about the frequency of similar SIs and describes the wavelet transform of the pixels in the ROI	720
Laplacian of Gaussian filtered features	Description of texture based on the images filtered by Laplacian of Gaussian	460

Mann–Whitney U test, as appropriate, was used to compare the differences in continues variables (age, lesion involvement length, tumor thickness, and extramural depth of invasion).

Receiver-operating characteristic (ROC) curves were generated to assess the diagnostic performance of the radiomic models in predicting LVI status by calculating the area under

the ROC curve (AUC) and corresponding accuracy, specificity, sensitivity, and so on were calculated.

The reported statistical significance levels are all two-sided, with the statistical significance set at 0.05.

Results

Clinical characteristics

In total, 188 patients were identified and comprise the study cohort: 128 males (68%) and 60 females (32%), the age ranged from 24 to 89 years, with an average of 59.61 ± 11.75 years old. The demographic statistics characteristics of patients in the training and testing cohorts are shown in Table 1. As is shown in Table 2, there were significant statistical differences in gender and maximum lymph-node short diameter in the training cohort between the LVI positive and negative groups ($P < 0.05$).

No significant differences in LVI prevalence were found between the two cohorts ($P = 0.952$). Overall, 42.4% and

42.9% of cases were LVI positive in the training and testing cohorts, respectively.

Radiomic feature extraction and selection

For each VOI, a total of 1653 three-dimensional (3D)-based radiomic features were extracted. These radiomic features quantified tumor characteristics using tumor size and shape, intensity statistics, and texture. For each patient, we integrated all of the 4959 radiomic features from three VOIs together.

We extracted radiomic features from multi-region (tumor and mesorectum) and single region (tumor), respectively, to investigate whether multi-regional radiomic model could improve the predictive performance in LVI. After the Spearman correlation analysis and LASSO algorithm, 21 radiomic features were retained for

Table 2 Characteristic of patients in the training and testing cohort

Characteristics	Training cohort			Testing cohort		
	LVI (+) <i>n</i> = 53	LVI (–) <i>n</i> = 72	<i>P</i>	LVI (+) <i>n</i> = 27	LVI (–) <i>n</i> = 36	<i>P</i>
Gender, no. (%)			0.429 ^a			0.184 ^a
Male	34 (64.2)	51 (70.8)		16 (59.3)	27 (75.0)	
Female	19 (35.8)	21 (29.2)		11 (40.7)	9 (25.0)	
Age, years	57.0 (51.0, 62.5)	61.5 (52.2, 69.8)	0.036 ^{c*}	62.8 ± 10.5	59.4 ± 11.7	0.239 ^b
The location of the tumor, no. (%)			0.154 ^a			0.089 ^a
Upper	3 (5.7)	2 (2.8)		5 (18.5)	1 (2.8)	
Middle	31 (58.5)	32 (44.4)		10 (37.0)	19 (52.8)	
Lower	19 (35.8)	38 (52.8)		12 (44.4)	16 (44.4)	
The tumor involved length (CM)	5.2 ± 1.7	5.1 ± 2.1	0.770 ^b	5.3 ± 2.7	4.9 ± 2.1	0.568 ^b
Tumor thickness (CM)	1.4 (1.1, 1.7)	1.3 (1.1, 1.6)	0.661 ^c	1.2 ± 0.4	1.3 ± 0.5	0.377 ^b
Extramural depth of invasion (MM)	5.0 (3.0, 8.0)	4.0 (1.0, 8.0)	0.260 ^c	4.0 (3.0, 7.0)	4.5 (2.5, 8.0)	0.650 ^c
Maximum lymph node short diameter (MM)	6.0 (5.0, 9.0)	5.0 (3.0, 6.0)	< 0.001 ^{c*}	5.9 ± 2.8	4.6 ± 3.0	0.091 ^b
CEA level, no (%)			0.382 ^a			0.787 ^a
Normal	29 (54.7)	45 (62.5)		21 (77.8)	29 (80.6)	
Abnormal	24 (45.3)	27 (37.5)		6 (22.2)	7 (19.4)	
CA19-9 level, no (%)			0.081 ^a			0.572 ^a
Normal	43 (81.1)	66 (91.7)		25 (92.6)	35 (97.2)	
Abnormal	10 (18.9)	6 (8.3)		2 (7.4)	1 (2.8)	
MRF, no. (%)			0.139 ^a			0.578 ^a
Normal	30 (56.6)	50 (69.4)		21 (77.8)	30 (83.3)	
Abnormal	23 (43.4)	22 (30.6)		6 (22.2)	6 (16.7)	

The threshold value for CEA level was 5 ng/mL and > 5 ng/mL, and the threshold value for CA 19-9 level was 39 U/mL and > 39 U/mL, according to the normal range used in clinics

LVI – lymphovascular invasion negative, LVI+ lymphovascular invasion positive, CEA carcinoembryonic antigen, MRF mesorectal fascia, CA19-9 carbohydrate antigen 19-9

**P* value < 0.05

^aChi-square test, data are number of patients, with percentages in parentheses

^bIndependent sample *t* test, data are mean ± SD

^cMann–Whitney *U* test, data are median, with interquartile range in parentheses

constructing the multi-regional radiomic model, including 2 clinical features (the location of primary tumor and maximum lymph-node short diameter), 10 radiomic features from mesorectum region, and 9 radiomic features from tumor region. In single-regional radiomic models, 10 radiomic features were retained (Table 3).

Radiomic model construction and evaluation

We constructed multi-regional and single-regional radiomic models, and then compared their predictive performance. In the model training stage, we use the results of fivefold cross validation as the performance of a specific machine learning classifier. Ridge Classifier used in the feature extraction of

Table 3 Feature coefficients of trained model

Feature name	Coefficient	
	With mesorectum	Without mesorectum
Location	- 0.550270779	- 0.650599274
Maximum lymph-node short diameter (mm)	0.810826003	0.913718234
mesorectum-T2WI-ExponentialGLCM-exponential-Imc2	- 0.407578931	Non-available
mesorectum-T2WI-ExponentialGLDM-exponential-LargeDependenceLowGrayLevelEmphasis	0.537410184	Non-available
mesorectum -T2WI-LogarithmGLSZM-logarithm-GrayLevelNonUniformityNormalized	0.854873667	Non-available
mesorectum -T2WI-LogarithmGLSZM-logarithm-SizeZoneNonUniformityNormalized	1.169559029	Non-available
mesorectum -T2WI-ShapeBased-Flatness	0.344792771	Non-available
mesorectum -T2WI-SquareFirstOrder-square-Minimum	- 0.5321271	Non-available
mesorectum -T2WI-SquareGLCM-square-Imc2	- 0.095079393	Non-available
mesorectum -T2WI-SquareRootGLCM-squareroot-ClusterShade	- 0.81680237	Non-available
mesorectum -T2WI-WaveletFirstOrder-wavelet-HLH-Median	0.688704501	Non-available
mesorectum -T2WI-WaveletNGTDM-wavelet-LHL-Strength	- 0.394190462	Non-available
tumor-DWI-WaveletGLCM-wavelet-HLL-MCC	- 0.662314866	- 1.055273156
tumor-T2WI-ShapeBased-SphericalDisproportion	0.187089993	0.297837686
tumor-T2WI-WaveletFirstOrder-wavelet-HLL-Mean	0.688141186	0.941876692
tumor-T2WI-WaveletFirstOrder-wavelet-HLL-Median	0.54697935	0.828889871
tumor-T2WI-WaveletFirstOrder-wavelet-HLL-Minimum	- 0.370569768	- 0.801286924
tumor-T2WI-WaveletFirstOrder-wavelet-LLH-Kurtosis	0.420803299	0.649334246
tumor-T2WI-WaveletGLCM-wavelet-HHH-Correlation	0.687261356	1.013091721
tumor-T2WI-WaveletGLCM-wavelet-HLL-Idn	0.257655009	Non-available
tumor-T2WI-WaveletGLRLM-wavelet-HLL-ShortRunHighGrayLevelEmphasis	0.465780302	0.336917221
Intercept	- 1.484964	- 1.190307866

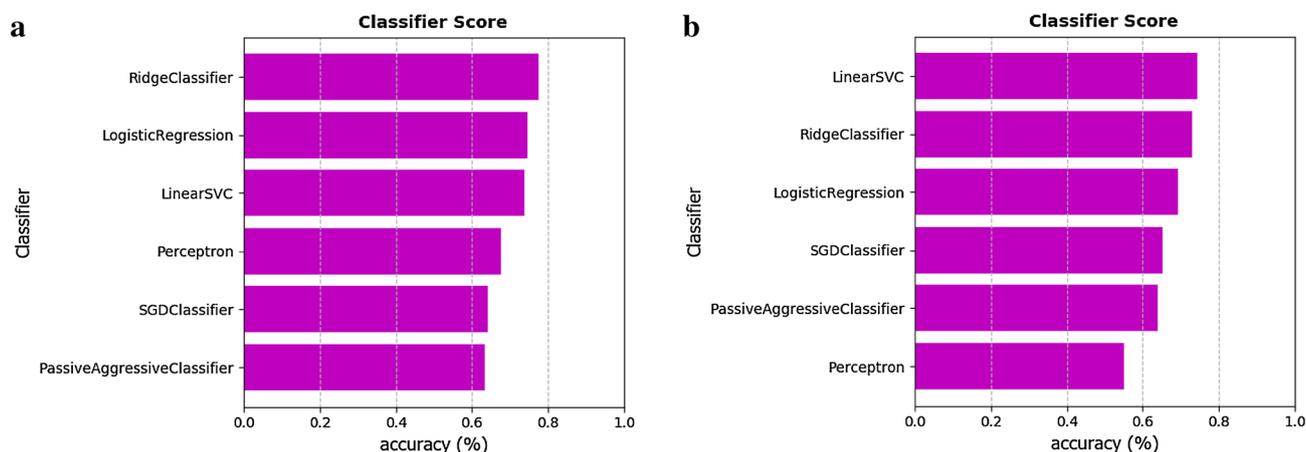


Fig. 3 Accuracy ranking from six models from multi-region (a) and single region (b) trained by different classifiers on fivefold cross-validation data set

multi-region and linear SVC used in the feature extraction of single region were found to produce the most accurate model on fivefold cross-validation data set, respectively (Fig. 3). In this linear models, the coefficients of each radiomic features are shown in Table 4.

As is shown in Fig. 4, in the predictive multi-regional radiomic model, the mean AUC of the ROC curves of five-fold validation is 0.82; the AUC of training data sets and testing data sets is 0.87 and 0.74, respectively. In the predictive single-regional radiomic model, the mean AUC of the ROC curves of fivefold validation is 0.79; the AUC of training data sets and testing data sets is 0.81 and 0.72, respectively. For more performance index, please see Table 4.

The AUC, accuracy, F1, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the multi-regional model is better than that of the single-regional model, except PPV in testing data sets.

Discussion

This retrospective study investigated a radiomic model based on MRI for the preoperative prediction of LVI status in patients with rectal cancer. We constructed and compared the single and multi-regional radiomic models for discriminate LVI invasion of rectal cancer patients. Our results suggest that MRI radiomic signature is significantly correlated with LVI. The multi-regional radiomic model was performed better than that of the single-regional model in preoperative prediction of LVI, with acceptable accuracy.

LVI is regarded as an important negative factor in treatment options and prognostication in rectal cancer from multiple statistically published trials. The previous study which was reported by Liu et al. found that the DCE-MRI-based radiomic features and LVI status were correlated in breast cancer, and indicated that the radiomic features were

effective in predicting the LVI status of patients with invasive breast cancer before surgery [29]. However, it is still a challenge for LVI preoperative prediction in rectal cancer. In the clinic, rectal MRI was suggested to reflect EMVI in rectal cancer, which is only one part of LVI [15–18]. Kim Y et al., who have investigated the visually assessed features, considered that the LVI presents when the mesorectal perivascular infiltrative signal was visible on pelvic MR imaging, and the sensitivity of MR-reported LVI status was 68.2% [30]. While the radiomic model in our study showed a better predictive performance than MRI-reported LVI status by Kim. The better performance in our research might be due to that the radiomic features which was derived from multi-regional VOIs in multiparametric MR images could provide comprehensive information on LVI status, including intramural, extramural blood vessels, and lymphatic vessels in rectal cancer.

Previous studies showed multi-regional MRI radiomics allowing for a more comprehensive characterization of the tumor heterogeneity. This may offer potential to improve the prediction performance [31, 32]. LVI, which is defined as the presence of tumor cells in the lymphatic vessels or blood vessels or both, include intramural, extramural blood vessels, and lymphatic vessels. In addition to the region of the tumor, the surrounding mesorectal tissues may also exhibit abnormal microscopic changes in the microvascular and lymphatic networks, extracellular matrix, and interstitial pressure, which cannot be ignored [33, 34]. Hence, we investigated whether multi-regional radiomics, including both tumor and mesorectum, could provide more features to discriminate LVI positive from LVI-negative lesions. When the current multi-regional radiomics signature was introduced into the prediction model, the performance improved than that of the single-regional model [34]. This suggests that the multi-regional radiomic signature could enhance the prediction of LVI in rectal cancer patients. In addition, our study used

Table 4 Performance of multi- and single-regional radiomic models on fivefold cross validation, training, and testing data sets

Models	Fivefold cross validation		Training data sets		Testing data sets	
	Multi-region	Single region	Multi-region	Single region	Multi-region	Single region
AUC	0.82	0.79	0.87	0.81	0.74	0.72
Accuracy	0.78	0.74	0.79	0.76	0.68	0.67
F1	0.73	0.69	0.82	0.80	0.70	0.70
Sensitivity	0.72	0.68	0.83	0.81	0.77	0.73
Specificity	0.82	0.79	0.75	0.70	0.61	0.60
PPV	0.75	0.71	0.81	0.76	0.64	0.67
NPV	0.80	0.77	0.77	0.76	0.74	0.67
FPR	0.18	0.21	0.26	0.30	0.39	0.40
FNR	0.28	0.32	0.17	0.19	0.23	0.27
FDR	0.25	0.29	0.19	0.24	0.36	0.33

PPV positive predictive value, NPV negative predictive value, FPR false-positive rate, FNR false-negative rate, FDR false discovery rate

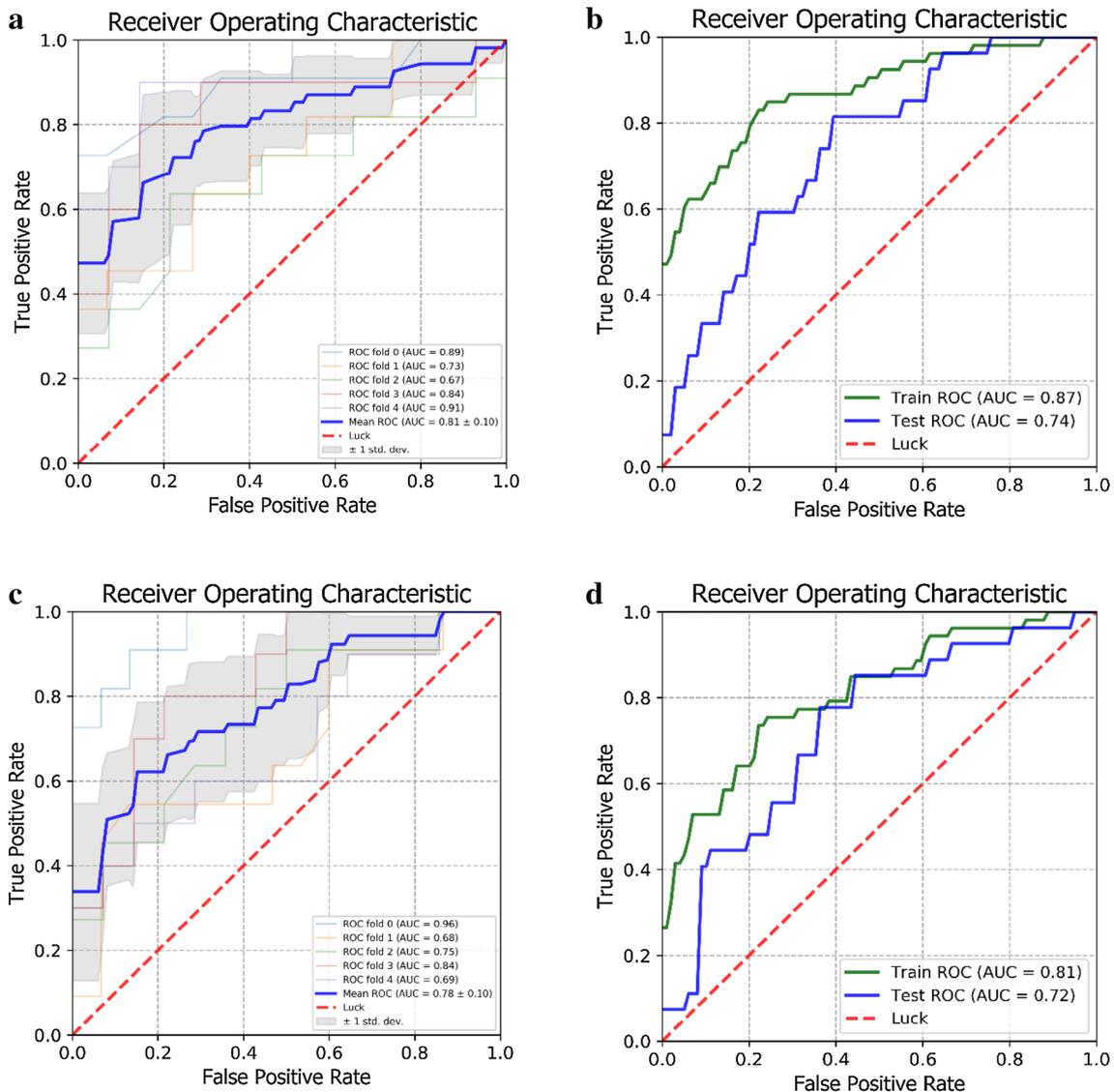


Fig. 4 ROC curves of radiomic models: **a** ROC of multi-regional model in fivefold cross validation; **b** ROCs of multi-regional model in training and testing data sets; **c** ROC of single-regional model in

fivefold cross validation; **d** ROCs of single-regional model in training and testing data sets

a 3 D VOI radiomic features by segmenting the tumor and mesorectum layer-by-layer. The 3D features provided more comprehensive information about lesions and improved the prediction accuracy of radiomic analysis compared with 2D features. Previous studies have shown that the 3D VOI provided more information about the heterogeneity of the whole lesion than the 2D region of interest (ROI) [35, 36].

How to select a modeling method is important for the performance of the radiomic model. Hence, a variety of machine learning methods should be used and the implementation should be fully documented [13], and then compare the performance of different algorithms. In our study, Ridge Classifier used in the features extraction of multi-region and linear SVC used in the feature extraction of single

region were found to produce the most accurate model on fivefold cross-validation data set, respectively, can predict the LVI and maybe assist clinical decision-making.

There are some limitations to this study. First, the sample was divided into training and testing cohorts, but lacked of external testing validation. It likely led to overfitting. And all the enrolled MRIs were performed on the same MR scanner, which may also reduce the robustness of the prediction models. In the future, the study cohort should mixed different MRI scanners' data sets to enhance robustness. Moreover, a multicenter study with a larger sample size and external validation is warranted. Second, the study did not evaluate T1 W and enhanced MR images, and only the VOIs of T2WI and DWI were calculated. However, in clinic, the T2WI and

DWI play vital role in tumor evaluation, which have a proven high diagnostic accuracy [23, 37]. Third, LVI status was only classified as positive or negative in this study. LVI status was further categorized into four grades based on the number of lymphovascular structures invaded, according to Jass classification (expanding vs infiltrative) [17, 18]. Further study should investigate the relationship between radiomic feature scores with grades of LVI. Finally, this research was a retrospective study. Therefore, there is an inevitable selectivity bias. In the future, we will design a prospective study of radiomic data related to rectal cancer.

In conclusion, the radiomic model which we developed demonstrates that multi-regional and multiparametric radiomic features based on MRI are useful tools for preoperatively predicting lymphovascular invasion in patients with rectal cancer.

Acknowledgements This study has received funding by Jilin Province Science and Technology Department Science and Technology Innovation Talents Cultivation Program (20180519008JH), Jilin Provincial Department of Finance (2018SCZWSZX-02, P8022740001, JLS-CZD2019-062), Jilin Province Development and Reform Commission (2017C020), and Prevention and Control of Major Diseases Science and Technology Action Plan of China [ZX07-c20160036].

References

- Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66(4):683–91. <https://doi.org/10.1136/gutjnl-2015-310912>.
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359–86. <https://doi.org/10.1002/ijc.29210>.
- Glynn-Jones R, Wyrwicz L, Tiret E, Brown G, Rodel C, Cervantes A et al. Rectal cancer: esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2017;28(suppl_4):iv22–iv40. <https://doi.org/10.1093/annonc/mdx224>.
- Benson AB, 3rd, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK et al. Rectal cancer, Version 2.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2018;16(7):874–901. <https://doi.org/10.6004/jnccn.2018.0061>.
- Ale Ali H, Kirsch R, Razaz S, Jhaveri A, Thipphavong S, Kennedy ED, et al. Extramural venous invasion in rectal cancer: overview of imaging, histopathology, and clinical implications. *Abdom Radiol (NY)*. 2019;44(1):1–10. <https://doi.org/10.1007/s00261-018-1673-2>.
- Al-Sukhni E, Attwood K, Gabriel EM, LeVeae CM, Kanehira K, Nurkin SJ. Lymphovascular and perineural invasion are associated with poor prognostic features and outcomes in colorectal cancer: a retrospective cohort study. *Int J Surg*. 2017;37:42–9. <https://doi.org/10.1016/j.ijsu.2016.08.528>.
- Huh JW, Kim HR, Kim YJ. Lymphovascular or perineural invasion may predict lymph node metastasis in patients with T1 and T2 colorectal cancer. *J Gastrointest Surg*. 2010;14(7):1074–80. <https://doi.org/10.1007/s11605-010-1206-y>.
- Barresi V, Reggiani Bonetti L, Vitarelli E, Di Gregorio C, Ponz de Leon M, Barresi G. Immunohistochemical assessment of lymphovascular invasion in stage I colorectal carcinoma: prognostic relevance and correlation with nodal micrometastases. *Am J Surg Pathol*. 2012;36(1):66–72. <https://doi.org/10.1097/pas.0b013e31822d3008>.
- Cienfuegos JA, Rotellar F, Baixauli J, Beorlegui C, Sola JJ, Arba L, et al. Impact of perineural and lymphovascular invasion on oncological outcomes in rectal cancer treated with neoadjuvant chemoradiotherapy and surgery. *Ann Surg Oncol*. 2015;22(3):916–23. <https://doi.org/10.1245/s10434-014-4051-5>.
- Lee JH, Jang HS, Kim JG, Cho HM, Shim BY, Oh ST, et al. Lymphovascular invasion is a significant prognosticator in rectal cancer patients who receive preoperative chemoradiotherapy followed by total mesorectal excision. *Ann Surg Oncol*. 2012;19(4):1213–21. <https://doi.org/10.1245/s10434-011-2062-z>.
- The Eighth Edition AJCC cancer staging manual. <https://www.springer.com/us/book/9783319406176>.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–6. <https://doi.org/10.1016/j.ejca.2011.11.036>.
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749–62. <https://doi.org/10.1038/nrclinonc.2017.141>.
- Beets-Tan RGH, Lambregts DMJ, Maas M, Bipat S, Barbaro B, Curvo-Semedo L, et al. Magnetic resonance imaging for clinical management of rectal cancer: updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol*. 2018;28(4):1465–75. <https://doi.org/10.1007/s00330-017-5026-2>.
- Smith NJ, Barbachano Y, Norman AR, Swift RI, Abulafi AM, Brown G. Prognostic significance of magnetic resonance imaging-detected extramural vascular invasion in rectal cancer. *Br J Surg*. 2008;95(2):229–36. <https://doi.org/10.1002/bjs.5917>.
- Bae JS, Kim SH, Hur BY, Chang W, Park J, Park HE, et al. Prognostic value of MRI in assessing extramural venous invasion in rectal cancer: multi-readers' diagnostic performance. *Eur Radiol*. 2019;29(8):4379–88. <https://doi.org/10.1007/s00330-018-5926-9>.
- Jass JR, Love SB, Northover JM. A new prognostic classification of rectal cancer. *Lancet (London, England)*. 1987;1(8545):1303–6. [https://doi.org/10.1016/s0140-6736\(87\)90552-6](https://doi.org/10.1016/s0140-6736(87)90552-6).
- Betje J, Pollheimer MJ, Lindtner RA, Kornprat P, Schlemmer A, Rehak P, et al. Intramural and extramural vascular invasion in colorectal cancer: prognostic significance and quality of pathology reporting. *Cancer*. 2012;118(3):628–38. <https://doi.org/10.1002/cncr.26310>.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures. They are data. *Radiology*. 2016;278(2):563–77. <https://doi.org/10.1148/radiol.2015151169>.
- Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol*. 2017;28(6):1191–206. <https://doi.org/10.1093/annonc/mdx034>.
- Verma V, Simone CB, 2nd, Krishnan S, Lin SH, Yang J, Hahn SM. The rise of radiomics and implications for oncologic management. *J Natl Cancer Inst*. 2017;109(7). <https://doi.org/10.1093/jnci/djx055>.
- Kaur H, Choi H, You YN, Rauch GM, Jensen CT, Hou P, et al. MR imaging for preoperative evaluation of primary rectal cancer: practical considerations. *Radiographics*. 2012;32(2):389–409. <https://doi.org/10.1148/rg.322115122>.
- Horvat N, Carlos Tavares Rocha C, Clemente Oliveira B, Petkovska I, Gollub MJ. MRI of rectal cancer: tumor staging, imaging

- techniques, and management. *Radiographics*. 2019;39(2):367–87. <https://doi.org/10.1148/rg.2019180114>.
24. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res*. 2017;23(23):7253–62. <https://doi.org/10.1158/1078-0432.CCR-17-1038>.
 25. Yang L, Liu D, Fang X, Wang Z, Xing Y, Ma L, et al. Rectal cancer: can T2WI histogram of the primary tumor help predict the existence of lymph node metastasis? *Eur Radiol*. 2019. <https://doi.org/10.1007/s00330-019-06328-z>.
 26. Jalil O, Afaq A, Ganeshan B, Patel UB, Boone D, Endozo R, et al. Magnetic resonance based texture parameters as potential imaging biomarkers for predicting long-term survival in locally advanced rectal cancer treated by chemoradiotherapy. *Colorectal Dis*. 2017;19(4):349–62. <https://doi.org/10.1111/codi.13496>.
 27. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Can Res*. 2017;77(21):e104–7. <https://doi.org/10.1158/0008-5472.can-17-0339>.
 28. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;26(30):5512–28. <https://doi.org/10.1002/sim.3148>.
 29. Liu Z, Feng B, Li C, Chen Y, Chen Q, Li X, et al. Preoperative prediction of lymphovascular invasion in invasive breast cancer with dynamic contrast-enhanced-MRI-based radiomics. *J Magn Reson Imaging*. 2019. <https://doi.org/10.1002/jmri.26688>.
 30. Kim Y, Chung JJ, Yu JS, Cho ES, Kim JH. Preoperative evaluation of lymphovascular invasion using high-resolution pelvic magnetic resonance in patients with rectal cancer: a 2-year follow-up study. *J Comput Assist Tomogr*. 2013;37(4):583–8. <https://doi.org/10.1097/RCT.0b013e31828d616a>.
 31. Li ZC, Bai H, Sun Q, Li Q, Liu L, Zou Y, et al. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. *Eur Radiol*. 2018;28(9):3640–50. <https://doi.org/10.1007/s00330-017-5302-1>.
 32. Kickingereder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology*. 2016;281(3):907–18. <https://doi.org/10.1148/radiol.2016161382>.
 33. Zhou Z, Lu ZR. Molecular imaging of the tumor microenvironment. *Adv Drug Deliv Rev*. 2017;113:24–48. <https://doi.org/10.1016/j.addr.2016.07.012>.
 34. Chen LD, Liang JY, Wu H, Wang Z, Li SR, Li W, et al. Multiparametric radiomics improve prediction of lymph node metastasis of rectal cancer compared with conventional radiomics. *Life Sci*. 2018;208:55–63. <https://doi.org/10.1016/j.lfs.2018.07.007>.
 35. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur J Radiol*. 2013;82(2):342–8. <https://doi.org/10.1016/j.ejrad.2012.10.023>.
 36. Blazic IM, Lilic GB, Gajic MM. Quantitative assessment of rectal cancer response to neoadjuvant combined chemotherapy and radiation therapy: comparison of three methods of positioning region of interest for ADC measurements at diffusion-weighted MR imaging. *Radiology*. 2017;282(2):418–28. <https://doi.org/10.1148/radiol.2016151908>.
 37. Li J, Liu H, Yin J, Liu S, Hu J, Du F, et al. Wait-and-see or radical surgery for rectal cancer patients with a clinical complete response after neoadjuvant chemoradiotherapy: a cohort study. *Oncotarget*. 2015;6(39):42354–61. <https://doi.org/10.18632/oncotarget.6093>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received:
06 August 2018

Revised:
03 October 2018

Accepted:
11 November 2018

Cite this article as:

Taylor J, Fenner J. The challenge of clinical adoption—the insurmountable obstacle that will stop machine learning?. *BJR Open* 2019; **1**: 20180017.

OPINION

The challenge of clinical adoption—the insurmountable obstacle that will stop machine learning?

¹JONATHAN TAYLOR, MEng, MSc and ²JOHN FENNER, PhD

¹Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

²University of Sheffield, Sheffield, UK

Address correspondence to: Mr Jonathan Taylor
E-mail: jonathan.taylor@sth.nhs.uk

ABSTRACT

Machine learning promises much in the field of radiology, both in terms of software that can directly analyse patient data and algorithms that can automatically perform other processes in the reporting pipeline. However, clinical practice remains largely untouched by such technology. This article highlights what we consider to be the major obstacles to widespread clinical adoption of machine learning software, namely: representative data and evidence, regulations, health economics, heterogeneity of the clinical environment and support and promotion. We argue that these issues are currently so substantial that machine learning will struggle to find acceptance beyond the narrow group of applications where the potential benefits are readily evident. In order that machine learning can fulfil its potential in radiology, a radical new approach is needed, where significant resources are directed at reducing impediments to translation rather than always being focused solely on development of the technology itself.

Machine learning is a topic of major interest in several areas of medicine. With a history of success in non-medical image analysis problems, it promises disruptive and transformative change in radiology.¹ In particular, it has been developed for computer-aided diagnosis and detection applications, and for data processing tasks such as automated tumour volume measurements. However, despite the media attention and impressive results generated in the lab, where, for instance, machines have been shown to outperform radiologists in specific disease recognition tasks,² uptake in the clinic remains vanishingly small. In the following sections, we highlight some of the substantial challenges of clinical translation which we believe have so far blocked progression towards widespread adoption, but which are rarely discussed in the literature. We argue that without substantial resources being focused on these issues, machine learning will continue to see limited application in clinical radiology.

REPRESENTATIVE DATA AND EVIDENCE

In order to make an informed decision on whether to invest in machine learning technology, representative clinical performance results are required. However, such data are often severely lacking. Firstly, algorithm development and testing is frequently carried out on limited datasets, which may not be representative of the clinic and may not

be associated with adequate “gold-standard” diagnoses. Many researchers use publically available data sets, such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI) databases (www.adni.loni.usc.edu). However, these data sets are often acquired under research rather than clinical protocols, using legacy equipment. Thus, there may be substantial uncertainties on the likely performance for clinical data. For example, in the case of ADNI consider the CAD Dementia challenge. Participants were invited to create a classification algorithm based on only very limited training data ($n = 30$). All but 2 of the 29 submitted classification algorithms used ADNI data to supplement the training data. However, in almost every case the estimated algorithm accuracy from training data was lower than that on previously unseen test data³

Furthermore, machine learning algorithms are rarely tested in the environment for which they were notionally designed. Standalone performance results are usually generated in a controlled setting with no human interaction. However, results derived from machine learning tools will at some point need to be interpreted by clinicians. This is particularly pertinent for assistive reporting software, which is designed to directly influence the radiologists’ final decision. Without evidence of impact on the whole

reporting workflow, benefits to the health service and patient care cannot be predicted.

Regulations

When software which is designed to have an impact on patient diagnosis or treatment is released on to the market, it is usually subject to some form of regulation. In Europe, the manufacturer must adhere to the Medical Device Directive. Whatever the classification under the regulations products need to be designed in such a way that patient safety is not compromised and that testing is carried out to ensure that the product performs as intended. Ensuring patient safety may be difficult for certain “black box” algorithms, such as neural networks, where outputs and risks may be challenging to predict.

Most regulatory regimes require clinical trials to confirm the performance of the final design. Ongoing surveillance is usually required in order to identify and fix any bugs associated with the software. Thus, meeting regulations usually requires significant financial resources, and diverse expertise. Costs are generally higher if the risk classification is high, which may be the case for machine learning algorithms designed to directly impact radiologists’ decisions.

Furthermore, the current model of device regulation in Europe (and in other jurisdictions) assumes that medical products are static entities, with any substantial changes to the product requiring reapproval. For machine learning algorithms that are designed to continually relearn and adapt their outputs in clinic, this approach to medical device regulation is impractical.

Health economics

When deciding on whether to invest in particular medical products many healthcare systems utilise economic analysis to inform their decision. In the UK, the National Institute for Health and Care Excellence (NICE, www.nice.org.uk) places strong emphasis on such data when generating guidance on medical technologies. This ensures that developed products have survived a cost–benefit analysis, providing evidence that can facilitate widespread adoption in the clinical community. However, even for the most simplistic economic analysis methods, such as cost–consequence analysis, evidence is required to quantify resource implications of the technology, as well as data on the likely clinical benefits. For many machine learning algorithms, this is difficult. For instance, gathering convincing data on the implications for patient pathways of a computer-aided detection algorithm, as compared to standard reporting methods, is likely to require extensive testing with radiologists under realistic clinical scenarios. Once again, this is likely to be expensive, complex and time-consuming.

Heterogeneity of the clinical environment

Machine learning tools cannot be implemented in isolation. If machine learning is to be used routinely, software needs to be integrated within the hospital infrastructure such that it can be easily accessed and used by reporters, according to local preferences, and data can be transferred to and from the analysis package as required. However, there are substantial differences between hospitals in terms of information technology resources,

associated restrictions and clinical protocols and workflows. The perils of ignoring local circumstances are reflected in the recently reported failure of IBM Watson for Oncology to achieve widespread clinical adoption, with the algorithms’ perceived in-built bias towards the American healthcare system cited as a major reason for lack of sustained uptake outside the United States.⁴ However, designing software that is adaptable to many different settings is difficult and, ultimately, it may not be possible to accommodate all the requirements of different hospital environments.

Data ownership

Machine learning research often relies on the use of retrospective patient data, acquired as part of standard care procedures. The steps necessary to achieve ethical approval in such circumstances are well established in Europe and the United States. However, if patient data are used to train an algorithm that is then sold commercially for profit, issues around data ownership and ethics can arise, particularly when data were originally acquired by a state-funded healthcare system.⁵ Furthermore, if the data are acquired in Europe and is not fully anonymised, the General Data Protection Regulations (GDPR) apply, which requires that the processing of personal data is in line with one of the specified lawful bases.⁶ If consent is chosen as the lawful basis, individual patients must opt-in to allow use of their data for machine learning development. This is another hurdle to development and obtaining and managing such consent adds extra costs to the development process.

Support and promotion

As highlighted by a recent Kings Fund report on adoption of innovation in the NHS, significant investment is usually needed to promote and support implementation of new technology.⁷ Simply generating evidence of impact is not enough to guarantee uptake in a healthcare system. If machine learning is to become a truly game-changing technology in radiology, support is likely to be needed from IT specialists, managers, radiographers as well as radiologists to ensure it is properly integrated in clinic. Not only does this require protected time (and therefore increased financial support) but the end users and patients have to be persuaded of its’ merits. Significant investment is therefore also required to promote the technology, to ensure that clinicians actively push the implementation. However, the perceived threat to radiologists’ role from machine learning, which is often inflated by articles in the popular press, is likely to make it harder to persuade the clinical community of the need for change.

Reflection

Machine learning promises much but given all of the above considerations, it is clear that the resources required to push machine learning technology into the clinic are substantial. Furthermore, it will take more than increased finances to enable machine learning to deliver on its promises.

In recent years, there has been some recognition that the challenges of implementation need to be addressed, rather than continually focusing on development of the algorithms themselves. For instance, the UK government is seeking to implement

recommendations from the life sciences industrial strategy,⁸ which references adoption of artificial intelligence and the need for funding to help move technology beyond the research arena. There is also recognition in the document that issues around data ownership, legislation and economic evaluation processes need attention if widespread implementation is to become a reality. In the United States, the FDA has recently taken a more active role in trying to streamline regulatory approval for software, as laid out in the Digital Health Innovation Action Plan.⁹ Another positive sign is that the FDA recently approved the first medical device that can diagnose disease without input from a clinician (IDx-DR). However, despite these new developments and initiatives the translational burden placed on new machine learning technology remains relatively unchanged.

There are some applications where the benefits from machine learning are likely to be so large that there will be sufficient backing from a multitude of sources to overcome all the challenges described (assuming the technology is sufficiently mature). For example, cancer screening examinations of the breast and lung generate large volumes of imaging data that human reporters must examine. Development of a computer system that can screen such images automatically would save a significant amount of money and reduce the pressure on radiologists' time, giving a strong incentive for adoption. The potential market for developers of such software would be large, encouraging commercial investment.

Furthermore, there are some machine learning applications associated with lower risk activities, such as automated segmentation

of tumours, where the barriers to adoption (particularly in terms of regulation) are likely to be less substantial.

However, we argue that for the majority of radiological applications the balance between potential benefits and likely costs is currently weighted too heavily in favour of costs so that widespread clinical adoption is unlikely to be achieved. A fundamental change is required if this situation is to be improved.

Perhaps the biggest issue facing machine learning developers (particularly those in smaller companies) is a lack of access to realistic clinical data. However, widespread sharing of patient data requires investment in infrastructure and is associated with significant reputational risk for the health provider (as demonstrated by negative publicity around projects such as NHS England's failed Care.data programme). Therefore, in accordance with the UK government's recent statement on artificial intelligence,¹⁰ the authors advocate the establishment of data trusts to create data sharing systems, and to control data flows in a secure, ethical and transparent manner.

Without such actions, there is a danger that the obstacles to routine application of machine learning throughout radiology will be insurmountable.

FUNDING

Jonathan Taylor is funded by a Doctoral Research Fellowship from the National Institute for Health Research (and Health Education England), Funder-id <http://dx.doi.org/10.13039/501100000272>, Award-id: HCS DRF-2014-05-002.

REFERENCES

- Liew C. The future of radiology augmented with artificial intelligence: A strategy for success. *Eur J Radiol* 2018; **102**: 152–6. doi: <https://doi.org/10.1016/j.ejrad.2018.03.019>
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. 2017. Available from: <https://arxiv.org/abs/1711.05225> [cited 2018 Aug 6].
- Bron EE, Smits M, van der Flier WM, Vrenken H, Barkhof F, Scheltens P, et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. *Neuroimage* 2015; **111**: 562–79. doi: <https://doi.org/10.1016/j.neuroimage.2015.01.048>
- Ross C, Swetlitz I. IBM pitched Watson as a revolution in cancer care. It's nowhere close. *Stat* [Internet]. 2017. Available from: <https://www.statnews.com/2017/09/05/watson-ibm-cancer/> [cited 2018 Oct 3].
- Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who owns the data? open data for healthcare. *Front Public Health* 2016; **4**: 7. doi: <https://doi.org/10.3389/fpubh.2016.00007>
- European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* 2016; 1–88.
- Collins B. Adoption and spread of innovation in the NHS [Internet]. 2018. Available from: <https://www.kingsfund.org.uk/publications/innovation-nhs> [cited 2018 Jan 22].
- Bell J. Life Sciences Industrial Strategy - A report to the Government from the life sciences sector [Internet]. 2017. Available from: <https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/115/115.pdf> [cited 2018 Jun 8].
- U.S. Food and Drug Administration. Digital Health Innovation Action Plan [Internet]. U.S. Food and Drug Administration. 2017. Available from: <https://www.fda.gov/downloads/medicaldevices/digitalhealth/ucm568735.pdf> [cited 2018 Oct 3].
- Department of Health, Department of Business, Energy and Industrial Strategy. Making a reality of the Accelerated Access Review: improving patient access to breakthrough treatments [Internet]. 2017. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664685/AAR_Response.pdf.

