

Automated Identification of Optimal Portal Venous Phase Timing with Convolutional Neural Networks

Jingchen Ma, BS, Laurent Dercle, MD, Philip Lichtenstein, MD, Deling Wang, MS, Aiping Chen, MD, Jianguo Zhu, MD, Hubert Piessevaux, MD, PhD, Jun Zhao, PhD, Lawrence H. Schwartz, MD, Lin Lu, PhD[#], Binsheng Zhao, DSc

Abbreviations

AUC

area under the curve

CI

confidence interval

CNN

convolutional neural network

CT

computed tomography

LM-CRC

liver metastases from colorectal cancer

LM-NET

liver metastases from neuroendocrine tumor

PVP

portal venous phase

Objectives: To develop a deep learning-based algorithm to automatically identify optimal portal venous phase timing (PVP-timing) so that image analysis techniques can be accurately performed on post contrast studies.

Methods: 681 CT-scans (training: 479 CT-scans; validation: 202 CT-scans) from a multicenter clinical trial in patients with liver metastases from colorectal cancer were retrospectively analyzed for algorithm development and validation. An additional external validation was performed on a cohort of 228 CT-scans from gastroenteropancreatic neuroendocrine cancer patients. Image acquisition was performed according to each centers' standard CT protocol for single portal venous phase, portal venous acquisition. The reference gold standard for the classification of PVP-timing as either optimal or nonoptimal was based on experienced radiologists' consensus opinion. The algorithm performed automated localization (on axial slices) of the portal vein and aorta upon which a novel dual input Convolutional Neural Network calculated a probability of the optimal PVP-timing.

Results: The algorithm automatically computed a PVP-timing score in 3 seconds and reached area under the curve of 0.837 (95% CI: 0.765, 0.890) in validation set and 0.844 (95% CI: 0.786, 0.889) in external validation set.

Conclusion: A fully automated, deep-learning derived PVP-timing algorithm was developed to classify scans' contrast-enhancement timing and identify scans with optimal PVP-timing. The rapid identification of such scans will aid in the analysis of quantitative (radiomics) features used to characterize tumors and changes in enhancement with treatment in a multitude of settings including quantitative response criteria such as Choi and MASS which rely on reproducible measurement of enhancement.

Key Words: Artificial intelligence; Quality control; Radiography; Abdominal; Tomography; X-ray computed.

© 2019 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

Acad Radiol 2019; ■:1–9

From the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China (J.M., J.Z.); Department of Radiology, Columbia University Medical Center, 630 West 168th Street, New York, NY 10032 (J.M., L.D., P.L., L.H.S., L.L., B.Z.); Gustave Roussy, Université Paris-Saclay, Université Paris-Saclay, Département D'imagerie Médicale, Villejuif, France (L.D.); Department of Medical Imaging, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, Guangdong, China (D.W.); Department of Radiology, First Affiliated Hospital of Nanjing Medical University, Nanjing, China (A.C.); Department of Radiology, the Second Affiliated Hospital of Nanjing Medical University, Nanjing, China (J.Z.); Cliniques Universitaires Saint-Luc, Brussels, Belgium (H.P.). Received January 9, 2019; revised February 25, 2019; accepted February 26, 2019.

Disclosure Statement: There are no disclosures.

IRB Statement: This retrospective study was waived by review board of Columbia University Medical Center, as only de-identified CT-scan images were utilized.

Address correspondence to: L. L. e-mail: ll2860@cumc.columbia.edu

[#] Whom correspondence should be addressed to.

© 2019 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.acra.2019.02.024>

INTRODUCTION

Most abdominal CT-scans are acquired after contrast enhancement at the “portal venous phase” (PVP). The PVP acquisition usually uses a fixed delay time after contrast injection. This parameter is set based upon the CT scanner characteristics and is not tailored for a patient's body habitus or cardiovascular system (1). This leads to a variability in timing and enhancement of the PVP acquisition. The variability is especially apparent in multicenter clinical trials where different vendors' equipment is used and potentially different image acquisition protocols are employed. Additionally, in many of these clinical trials as well as in clinical practice, the acquisition of serial scans in the same patient to assess for change over time requires an understanding of the variability and its impact on measurements of response and progression. There is no reliable way to quantitatively and automatically

recognize the quality of the timing of acquired images. This study aimed to design a deep learning algorithm to automatically recognize optimal PVP acquisitions on CT-scans to ascertain image quality so that the assessment of density-based biomarkers will be more reliable (2).

Optimal PVP timing is crucial in oncology for the automatic detection and characterization of lesions (3), as well as the estimation of tumor enhancement or vascularity, which is increasingly being used to predict treatment response, as well as outcome and recurrence (2, 4–6). In clinical trials and routine practice, anticancer treatment efficacy may be evaluated using change in tumor size (response evaluation criteria in solid tumors: RECIST) (7, 8), or change in tumor size and/or density (Choi and MASS criteria) (9–12). RECIST criteria define treatment response by tumor shrinkage, while Choi criteria specify that a 15% decrease in tumor density on CT-scan acquired at the PVP could be an additional surrogate marker of treatment efficacy (10–12). The incremental value of Choi criteria as compared with RECIST is in the detection of additional patients with clinical benefit, and the response rate can double (from 46% using RECIST to 83% using Choi criteria) (13).

Previously, a visual assessment quality control (QC) tool, based on a quality control scoring system differentiating early, optimal, and late PVP, was demonstrated to greatly improve the estimation of cancer drugs' efficacy using Choi criteria (14). The application of this QC for the evaluation of anticancer treatment efficacy demonstrated that a nonoptimal PVP timing (too early or too late) significantly altered response evaluation when using the Choi criteria. A nonoptimal PVP-timing induced an apparent, false 15% decrease in the measurement of tumor density. This study demonstrated the actual clinical utility of such QC tools. These previously designed QC tools used two different methodologies, both of which had limitations: (1) radiologists' visual assessment or (2) a semiautomatic classification software (14). The major limitation of the visual assessment is that it is highly reader dependent and subjective even after appropriate reader training. The ability to manually perform this type of QC depends on readers' skills and experience. Additionally, both visual assessment by a radiologist and the semiautomatic classification software can be time intensive due to the multitude of tasks that must be performed including optimizing viewing window width/level, and visually estimating density relationships between the relevant organs (or putting region of interests (ROIs) on the organs to calculate density ratios).

In the new era of radiomics and deep learning, convolutional neural networks (CNN)s are highly effective for automated detection and classification tasks such as lung nodule detection (15, 16), differentiation of liver masses (17), assessing skeletal maturity (18), and anatomical recognition (19). Our hypothesis was that a deep learning method can automatically process the information embedded in CT images without the need for visual assessments and the manual selections of ROIs used in the previous machine learning algorithm (14). The purpose of this study was to design a deep learning algorithm to recognize optimal PVP acquisitions on CT-scans, PETAL

(Portal vEin & aorTa based Automatic quaLiTy control), to evaluate if a PVP-timing acquisition is optimal or nonoptimal, so that other sensitive algorithms and techniques can be evaluated with optimal quality scans that require accurate assessment of tumor enhancement over time.

MATERIALS AND METHODS

Patient Image Data

This retrospective study was waived by our institutional review board because only de-identified CT-scan images were utilized. We selected liver metastasis from colorectal cancer (LM-CRC) to train and validate our PETAL algorithm, as colorectal cancer is the second leading cause of cancer death and liver metastasis involve greater than half of the patients with colorectal cancer (20). To study our algorithm's generalizability, we chose liver metastasis from gastroenteropancreatic neuroendocrine tumors (LM-NET) as external validation dataset. The gastroenteropancreatic NET is indeed associated with prolonged survival and relatively high prevalence (21–23). In the training and validation sets, we included all LM-CRC patients. No patients had previous hepatectomy in the LM-CRC datasets: all patients were nonresectable. In the external validation dataset, we excluded 25 out of 99 patients with hepatectomy or any severe mass effect with portal vein anatomy changed, because several LM-NET patients had performed previously severe cytoreduction surgery (including hepatectomy) that are very specific to LM-NET and could affect the detection of portal vein. Additionally, the size and anatomic distribution of lymphadenopathy of LM-NET patients was different from LM-CRC patients. In colorectal cancer, the average size of the enlarged lymph node caused no mass effect or portal vein distortion. In contrast, with neuroendocrine tumors, several patients initially had such large portocaval lymph node masses that there was compression and inflow/outflow vascular phenomena which dramatically change following therapeutic cytoreduction surgery to the liver.

Both data sets were obtained sequentially from previous, already completed, multicenter clinical studies of colorectal and neuroendocrine tumors. Patients' CT images were acquired with standard-of-care abdominal imaging protocols at the PVP after intravenous injection of an iodinated contrast-enhancement product. The PVP acquisition used fixed delay time after contrast injection. The CT imaging settings are shown in Table 1.

We randomly divided LM-CRC dataset into training and validation sets on a ratio of 7:3 for the deep learning algorithm development. CT-scans from the same patients were assigned to the same group: assigned either to the training set or to the validation set.

Reference Standard

For each patient in both LM-CRC and LM-NET datasets, only one pair of radiologists reviewed the images and provided their PVP timing score to create radiologists' initial

TABLE 1. CT Imaging Settings Statistics of LM-CRC and LM-NET Dataset

Dataset	LM-CRC	LM-NET	p Value
Scanner	GE, Philips, SIEMENS, etc.	GE, Philips, SIEMENS, etc.	
Tube voltage median (Q1, Q2)	120 (120, 120) kVp	120 (120, 120) kVp	0.11
Tube current time median (Q1, Q2)	156 (121, 200) mAs	169 (114, 212) mAs	0.24
Recon kernel	smooth	smooth	
Slice thickness median (Q1, Q2)	5 (5, 5) mm	3 (2, 5) mm	6.1×10^{-31}
Pixel spacing median (Q1, Q2)	0.74 (0.68, 0.79) mm	0.71 (0.66, 0.77) mm	0.0035

Q1 is lower (first) quartile and Q2 is upper (third) quartile; LM-CRC, liver metastases from colorectal cancer; LM-NET, liver metastases from neuroendocrine tumor.

visual consensus. Four radiologists interpreted LM-CRC dataset due to a large amount of CTs. Their experience were 3, 5, 8, and 15 years respectively. In the external validation stage, two radiologists with 8 and 12 years of experience interpreted the LM-NET dataset. The reference standard was the three-phase consensus process shown as follow.

Visual Scoring of PVP Timing

We categorized the PVP timing as optimal or nonoptimal. Nonoptimal PVP timing included either early or late PVP timing. The definition was based on the relative contrast enhancement within vessels and tissues (24–28). Optimal PVP demonstrated peak enhancement of the liver parenchyma and portal vein, as well as some enhancement of the hepatic veins. Early PVP was defined by contrast still predominantly in the arterial supply as compared with the portal vein. Late PVP was associated with a washout of the hepatic contrast enhancement and approached the nephrogenic phase with more enhancement of the renal medulla. Six examples from three patients are shown in Figure 1.

(1) *Initial Visual Consensus.* Each CT scan was initially scored by two radiologists. The radiologists independently read their CT scans and scored the PVP timing. If the two radiologists' scores agreed, then the score was finalized. If the two radiologists disagreed, then a joint consensus reading was performed between both radiologists after a time interval to avoid memory bias. If the two radiologists could not reach an agreement during the consensus reading, then a third radiologist adjudicated to reach a majority consensus for the score.

(2) *Semiautomatic classification.* Each CT scan was also scored by a semiautomatic classifier. We used a previously described semiautomatic classification of PVP timing in (14) to ensure the quality and reproducibility of our PVP timing classification. To this end, we used a previously described probability map that the PVP timing was optimal based on the average density (in Hounsfield Unit) of the image pixels inside a manually drawn ROI in the portal vein and in the aorta as well (14). The details of semiautomatic classification method were shown in the supplemental materials S1.

(3) *Final (including 1 and 2) Reference standard Consensus.* For any semiautomatic classification that differed from the initial visual consensus, an additional consensus reading was performed by the pair of radiologists to determine whether the

optimal PVP timing classification, after the semiautomatic classification scores were shown to the pair of radiologists. The goal was to obtain the most accurate and optimal reference-standard PVP-timing consensus for subsequent analyses. In summary, this was derived from (1) a visual consensus between a pair or trio of radiologists, (2) a semiautomatic classification of all scans to find discrepant scans of which computer-based classifications were different from radiologists' visual consensus, and (3) a final visual consensus by the radiologists for any discrepant scan.

Algorithm Building

The contrast enhancement in portal vein and aorta was demonstrated to be the dominant information that contributed to classify the optimal PVP-timing as the ROIs in liver, spleen, inferior vena cava, psoas muscle, and kidney did not provide incremental value (14). With this knowledge, an automated algorithm for deriving a probability of optimal PVP-timing by analyzing the portal vein and aorta was developed with the following 6 steps (Fig. 2).

Step 1. Input data for the algorithm was the volumetric data (voxel size is $N \times 512 \times 512$, where N is the number of slices) from a contrast-enhanced abdominal CT scan.

Step 2. Identification of CT slices containing portal vein was determined using the Alex-net (29) which pushed classification accuracy of the ImageNet Challenge (30) by a significant stride in comparison to traditional methodologies. By training the first CNN, Alex-net, our algorithm predicted the probability of appearance of portal vein in every CT slice. Consecutive CT slices with probabilities greater than 0.5 were determined as the axial slices containing portal vein and then fed in step 3 and step 4. These axial slices also contained aorta according to the anatomy. An example is shown in Figure 3b. The rest slices were excluded for the subsequent steps to reduce false positives detections of portal vein and aorta.

Step 3. Detection of portal vein's bounding-boxes was performed next with another CNN, the Faster R-CNN (31) which is a deep-learning based object detection system with high performance to run at near real-time frame rates. Figure 3c illustrates the detection of the portal vein. The bounding-box of portal vein contained portal vein and its adjacent tissues, such as a part of the liver. At most

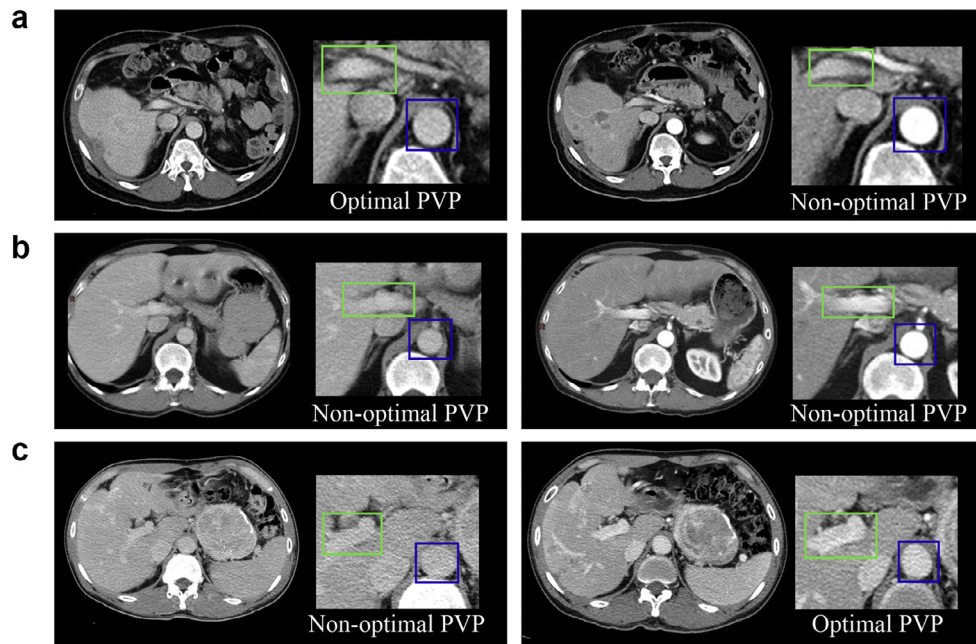


Figure 1. Clinical implication of inpatient variability in portal venous phases timing using CT images acquired at baseline and after treatment initiation. Optimal PVP timing (a left and c right) was defined by peak enhancement of the liver parenchyma and portal vein, as well as some enhancement of the hepatic veins. Patient A was treated for a NET. At baseline (a, left), the biodistribution of the contrast enhancement product was optimal and significantly differed from the subsequent follow-up (a, right) which was nonoptimal as demonstrated by the contrast still being predominantly in the arteries. This might mimic treatment effect if tumor density is used as a response criterion (e.g., Choi, MASS, mRECIST). Patient B was treated for a CRC, at baseline (b, left), the PVP timing was nonoptimal because the contrast enhancement in the liver parenchyma and portal vein was low due to the washout of the contrast enhancement. The acquisition after treatment initiation (b, right) was nonoptimal as explained by the fact that the contrast agent was still predominantly in the arteries. Patient C was treated for a NET. The PVP timing was nonoptimal at baseline (c, left) and optimal during the follow-up (c, right). This will lead to an increase in liver and tumor density at follow-up due to acquisition protocol and might mask treatment effect if tumor density is used as a response criterion.

one bounding-box was created for each slice identified by the step 2. For each cases, if the number of the bounding-boxes was less or equal to five, all the bounding-boxes were retained. If the number is greater than five, top five bounding-boxes with highest probability values were retained and the rest bounding-boxes were excluded. A case is shown in Figure 4. These retained bounding-boxes were used to feed in step 5.

Step 4. Detection of aorta's bounding-boxes was identical to the portal vein localization in the previous step. Figure 3d showed a case to illustrate the bounding-box detections of the aorta.

Step 5. The optimal PVP-timing was automatically classified based on the pair of portal vein's bounding-box and aorta's bounding-box attained in steps 3 and 4. Portal vein bounding-box paired with aorta bounding-box according to the probability order. If the number of portal

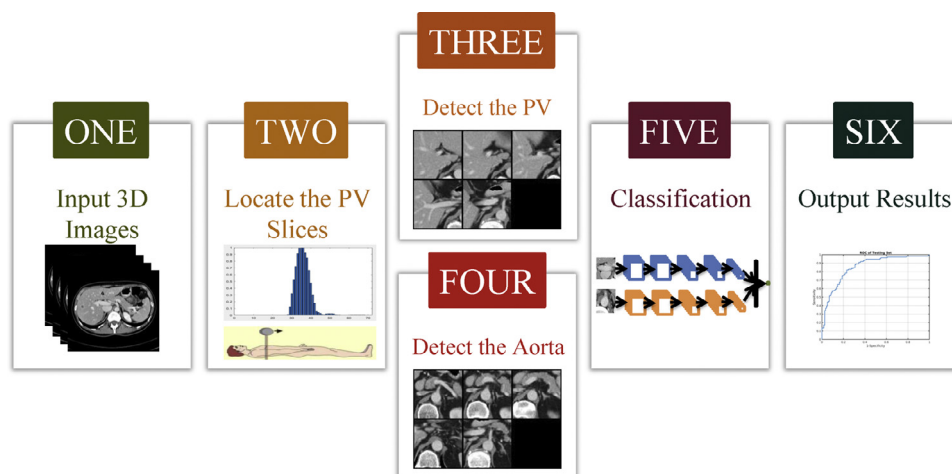


Figure 2. This graph illustrates the pipeline of our proposed PETAL algorithm.

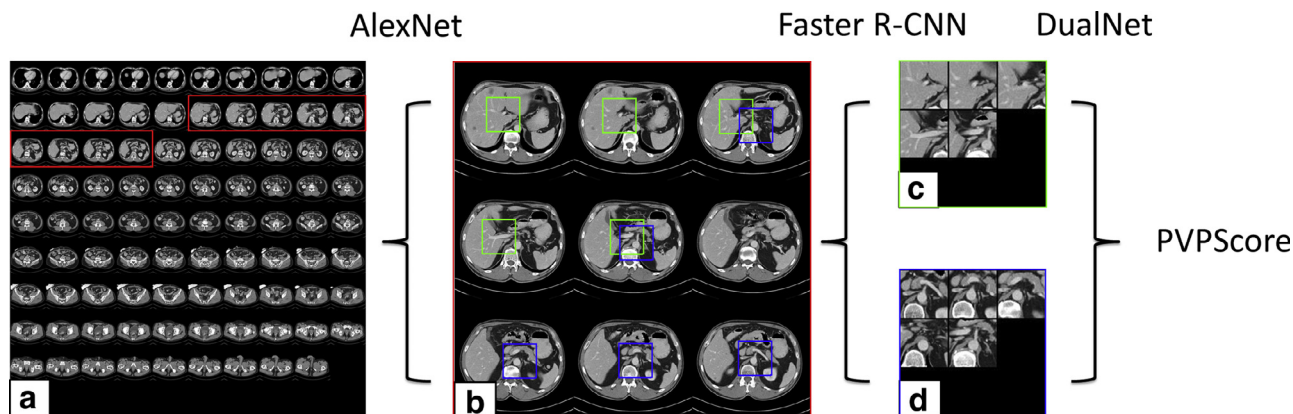


Figure 3. An example case predicted by PETAL. (a) All the image series of the input CT. The AlexNet identified the slices containing portal vein shown in the red boxes. (b) The axial slices containing portal vein with AlexNet. Faster R-CNN detects the bounding-boxes of portal vein (green boxes) and aorta (blue boxes), which are shown in (c) and (d) respectively. Finally, PVP score is predicted based on the detected portal vein and aorta bounding-boxes with our proposed dual input CNN. (Color version of figure is available online.).

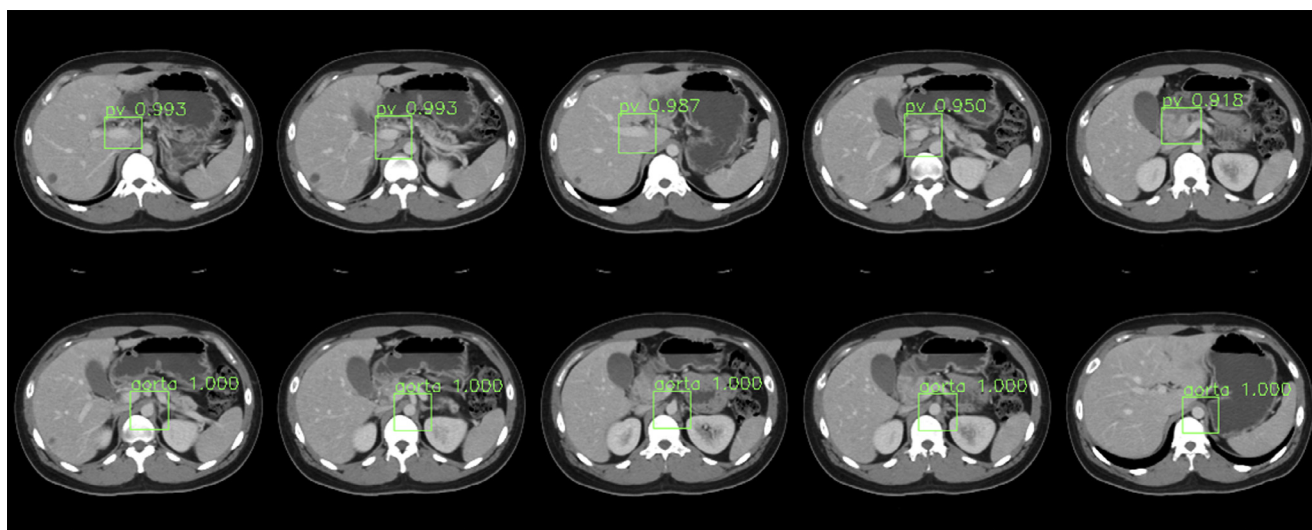


Figure 4. An example of the automatically portal vein and aorta detections. The images show a case where PETAL detected top five highest probabilities bounding-boxes from five slices for both portal vein and aorta. The bounding-boxes in the same column were in the same pair to feed in dual input CNN.

vein bounding-box was less than the number of aorta bounding-boxes, the portal vein bounding-box with the highest probability would be replicated multiple times to equal the number, and vice versa. Here, we developed the third CNN, a new dual input CNN as shown in Figure 5, to process the image patches of portal vein and aorta. Each score can be obtained by analyzing each pair of portal vein bounding-box and aorta bounding-box.

Step 6. Output of final PVP probability was the average of five scores resulted from five pairs of portal vein's and aorta's bounding-boxes.

Implementation details of PETAL were shown in the supplement materials S2.

Performance Evaluation

We used Free-Response ROC (FROC) to evaluate the detection performance of the portal vein and aorta. All the bounding-boxes of PETAL were inspected by one radiologist and were classified into true positives or false positives. A true positive detection (hit) was defined by the bounding-box being centered on the target (portal vein or aorta). Otherwise, the bounding-box was considered as a false positive detection. The sensitivity and false positive per scan of the FROC were defined as following:

$$\text{Sensitivity} = \frac{\# \text{ cases that at least one bounding box hit the target}}{\# \text{ total cases}}$$

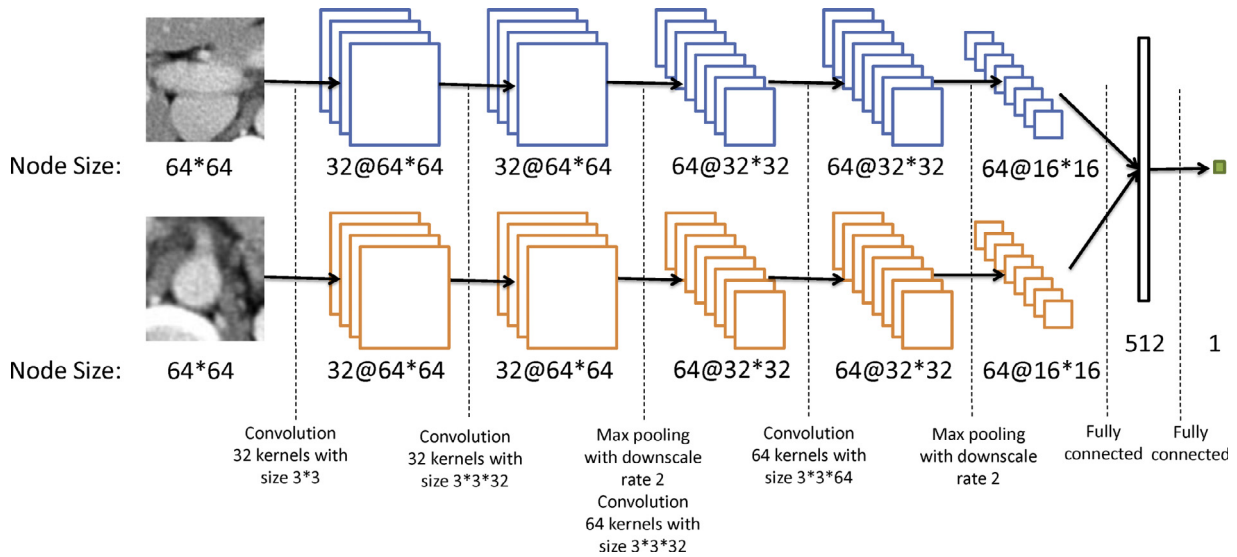


Figure 5. Proposed dual input CNN architecture. Input patches are processed by convolutional and max-pooling layers. The output PVP score is processed with the concatenation of portal vein and aorta information and fully connected layers. In this network, two blocks of two consecutive convolutional layers with 32 and 64 filters were used. Each filter had 3-by-3 pixels convolutional kernels with ReLU activation, zero padding on the edge. Each block was followed by a 2-by-2 max-pooling layer and a dropout layer of dropout rate 50%. Finally, the fully connected layer of 512 nodes with ReLU activation combined the information about the portal vein and the aorta. The final 1 node layer was connected with the previous layers with sigmoid activation.

False positive per scan

$$= \frac{\# \text{ bounding boxes that did not hit the target}}{\# \text{ total cases}}$$

The performance of PETAL was evaluated in terms of area under the curve of ROC curve (AUC). The ROC curve was calculated by comparing the output probabilities with the reference standards. The real positive case was a case of which the reference standard was optimal PVP and the real negative case was a case of which the reference standard was nonoptimal PVP. The sensitivity and specificity in the ROC were defined as following:

$$\begin{aligned} \text{Sensitivity} &= \frac{\# \text{ true positive cases}}{\# \text{ real positive cases}} \\ &= \frac{\# \text{ cases that both PETAL outputs and reference standards were non-optimal PVP}}{\# \text{ total cases that reference standards were non-optimal PVP}} \end{aligned}$$

$$\begin{aligned} \text{Specificity} &= \frac{\# \text{ true negative cases}}{\# \text{ real negative cases}} \\ &= \frac{\# \text{ cases that both PETAL outputs and reference standards were non-optimal PVP}}{\# \text{ total cases that reference standards were non-optimal PVP}} \end{aligned}$$

The 95% confidence interval of AUC was estimated by using bootstrapping with 1000 bootstraps of prediction scores. We collected two cohorts of data independently. The LM-CRC cohort was used for training and validating PETAL and the LM-NET was used for testing PETAL blindly.

RESULTS

In total, 479 PVP CT-scans (300 optimal and 179 nonoptimal) from LM-CRC dataset were used for training and 202 PVP CT-scans (137 optimal and 65 nonoptimal) from LM-CRC dataset were used for validation. The external validation set utilized LM-NET dataset with 228 PVP CT-scans (111 optimal and 117 nonoptimal).

Of note, PETAL algorithm consisted with six steps. The errors in the previous steps would affect the subsequent steps, so all results contained the propagating errors from the prior steps.

Overall Optimal PVP Classification

The optimal PVP probability of PETAL outputs ranged from 0 to 1. In the external validation set, the optimal threshold of the PVP probability was 0.84. At this operating point, the sensitivity, specificity, positive predictive, and negative predictive values

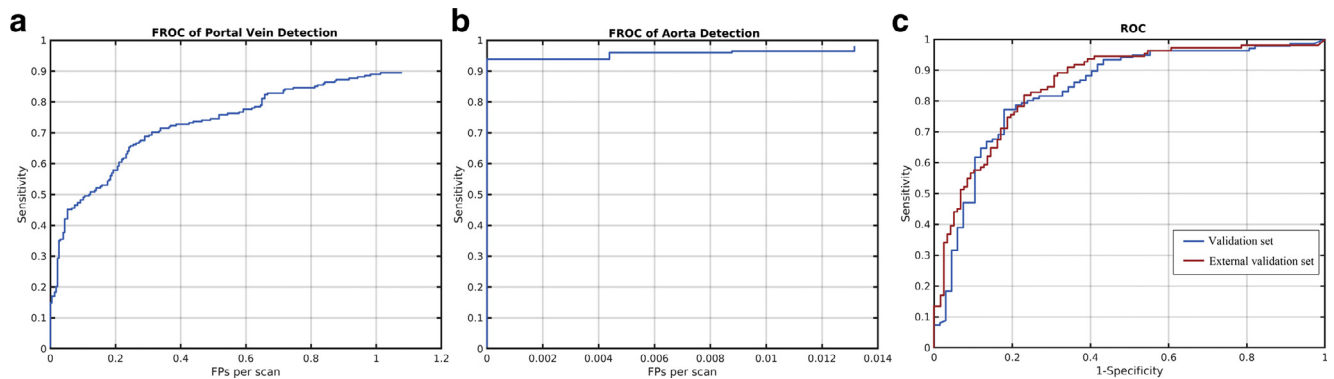


Figure 6. Performance evaluations with FROC and ROC curves. (a) FROC of portal vein detection that shows PETAL achieved 89.5% sensitivity with 1.08 false positive result per scan in the external validation set. (b) The aorta detection was almost perfect since it achieved 98.2% sensitivity with 0.0132 false positive results per scan. (c) The ROC curve of optimal PVP classification of validation set, AUC of 0.837 (95% CI: 0.765, 0.890) and the ROC in the external validation set, AUC of 0.844 (95% CI: 0.786, 0.889).

were 82% (91/111), 74%(87/117), 75%(91/121), and 81%(87/107), respectively. At the operating point of 0.91, the optimal threshold on the validation set, the sensitivity, specificity, positive predictive, and negative predictive values of external validation set were 78%(87/111), 77%(90/117), 76%(87/114), 79% (90/114), respectively. Good performances were also observed in the validation set, the LM-CRC dataset, with an AUC of 0.837 (95% CI: 0.765, 0.890) and in the external validation set, the LM-NET dataset that was a different tumor type than the training set, with an AUC of 0.844 (95% CI: 0.786, 0.889) (Fig. 6). There was no significant difference between colorectal cancer and neuroendocrine cancer in terms of AUC. It is likely that these two different cancer types do not vary the pharmacokinetics of contrast media in the vessels we used in this study. Additionally, PETAL algorithm was robust to the CT slice thickness and pixel spacing which were statistically significantly different between the LM-CRC and LM-NET datasets as shown in Table 1.

Intermediate Results

(1) Portal vein slices classification. In the external validation set, four cases failed to locate the axial slices containing portal vein (1.75%) due to poor contrast in the PV ($n = 1$), displaced portal vein caused by mass effect ($n = 2$) and systematic failure of the algorithm ($n = 1$, contrast enhancement was unexpectedly high in the PV with 360 HU, which is above our abdominal CT window of -160~240 HU).

(2) Portal vein detection. PETAL algorithm extracted totally 1046 bounding-boxes of portal veins from 228 cases in the external validation set and received 89.5% sensitivity with 1.08 false positive per scan. The FROC curve is shown in Figure 6a. One radiologist analyzed the detections of 1046 bounding-boxes visually. There were some errors due to selection of multiple vessels (especially in the coeliac trunk 93 cases, 8.9%), detections of bone vertebra or rib (71 cases, 6.8%), detection of structures with oral contrast enhancement (opacification digestive tract, 18 cases, 1.7%), and selection of normal structures such as liver, kidney, stomach (65 case, 6.2%).

(3) Aorta detection. The aorta detection was almost perfect; it achieved 98.2% sensitivity with 0.0132 false positives per scan. We analyzed the detections of all 1046 bounding-boxes in the external validation set. There were three mistakes out of 1046 cases, two of which were caused by oral contrast enhancement and one was caused by selection of multiple vessels in the coeliac trunk.

Duration of Data Processing

PETAL algorithm can evaluate a CT scan in 3 seconds on average by a computer with a GTX 1060 GPU and an Intel i7-7700HQ 2.8GHz CPU. The algorithm was fully automatic. We applied PETAL algorithm to the LM-NET dataset twice and got exactly the same probability of being optimal PVP for every case, because the algorithm did not contain human interactions or random factors.

DISCUSSION

In this study, we developed a deep learning method to identify CT-scan acquisitions with an optimal PVP-timing, based on a fully automated analysis of CT scan images. This method first screens the entire abdominal CT scan, proposes five bounding-boxes of the aorta and five bounding-boxes of the portal vein, and finally outputs a probability of the CT scan being optimal PVP; all automatically in 3 seconds. This algorithm was developed and validated with CT scans from two multicenter clinical studies and involving two different primary tumor types, colorectal metastases which are generally hypointense on the PVP phase and neuroendocrine metastases which are generally hyperintense. The performance of the algorithm was comparable with the radiologists for the visual assessment of the PVP timing, which was reported to be 81.7% accuracy in (14). It achieved an AUC of 0.837 (95% CI: 0.765, 0.890) in the validation set and an AUC of 0.844 (95% CI: 0.786, 0.889) in the external validation set. Thus, the deep learning method was able to differentiate an optimal PVP-timing versus a nonoptimal PVP-timing (early or late

PVP) acquisition. This algorithm can assist in determining whether an intended PVP acquisition is optimal and can also help researchers by identifying nonoptimal PVP acquisitions on retrospective studies, which they may consider evaluating separately. In this study, 43% (87/202) LM-CRC cases and 50% (114/228) LM-NET cases were recognized as nonoptimal PVP by this algorithm. However, the patients with nonoptimal PVP scans may have heart failure or other conditions that negatively affect studies. It still need to be studied in the future by using the quality of the PVP as a new factor in statistical analysis or as a new variable to integrate into radiomics model (normalizing features or model building).

This algorithm potentially provides additional and incremental value to the current bolus tracking technique used in CT imaging. Bolus tracking is used during the imaging acquisition to control phase timing so that the likelihood of the PVP-timing being optimal is increased. However, it does not consider the individual patient's biological variation and thus cannot ensure if the optimal PVP-timing was successfully reached in a given patient (24, 32–35). In bolus tracking, a small bolus of radio-opaque contrast media is injected into a patient via a peripheral intravenous cannula. The flow of contrast to a specific vessel can be tracked using an ROI. When the contrast in the vessel reaches a certain threshold, the acquisition timing is considered optimal and the CT-scan acquisition is triggered. Additionally, bolus tracking is not used in all center and most centers still rely on image acquisition on the historical and empirical 60-second delay between contrast-agent product injection and image acquisitions. The problem is that the individual optimal hepatic enhancement timing cannot be assumed a priori, because it is associated with multiple variables as well as the anthropomorphic characteristics and hemodynamic status of the patient (36–38). Consequently, an optimal PVP timing is only reached in two out of three patients in multicenter clinical trials (14). Hence, the need for such quality control techniques is critical.

This algorithm may help to ensure an optimal extraction of imaging biomarkers that could be useful to classify new patterns of progression (39–41) and response (42–44) in novel anticancer agents by automatically identifying optimal PVP timing and potentially excluding imaging data sets with improper acquisition timing. It could also be applied to improve radiomics signatures by selecting CT images of optimal quality (45–50). The algorithm is capable of ensuring that conditions are optimal to appraise tumor vascularity and density-based response criteria such as CHOI (9–11). A previous study (14) demonstrated that a 14.8% variability in tumor density measurement was directly caused by variation in PVP acquisition timing which is deleterious to the use of density as a biomarker, because tumor density reduction would not be robust as its variability could be due to acquisition timing (rather than tumor lesion biology) and would therefore be poor at predicting overall survival and progression-free survival (9–12). Increasing the reproducibility of tumor density measurements is essential to acceptance of density-based biomarkers as the basis for imaging endpoints in clinical trials and response in new anticancer agents. Identifying optimal PVP

enhancement may also benefit some organ segmentations due to increasing the contrast between the target organ and the background. For the patients studied over time, the consistency of PVP enhancement can make the assessments more reliable.

There are some limitations in this work. First, the gold standard of optimal PVP-timing for each patient's CT scan required the creation of a consensus reference standard based on the subjective visual assessment of trained radiologists and the objective computer-aided classification. Second, PETAL cannot be applied to the patients with major hepatic surgery or any severe mass effect with portal vein anatomy changed. Finally, the overall performance was limited by the accuracy of portal vein localization. Since the algorithm consists of multiple parts, an error in any one part will propagate through the entire algorithm. The greatest challenge in the algorithm was in the portal vein localization, for which we will continue to optimize.

CONCLUSION

Our work demonstrates that a fully automatic, deep-learning derived PVP-timing recognition system can reliably and rapidly identify the optimal PVP-timing based on CT images. This technique could be invaluable for any type of image analysis, such as radiomics, which rely on features that might be affected by the phase of contrast administration and scanning.

ACKNOWLEDGMENT

National Cancer Institute (NCI) supported this work in part by Grants U01 CA140207 and U01 CA225431. The content is solely the responsibility of the authors and does not necessarily represent the funding sources.

REFERENCE

1. Bae KT. Intravenous contrast medium administration and scan timing at CT: considerations and approaches. *Radiology* 2010; 256:32–61.
2. Smith AD, Shah SN, Rini BI, et al. Morphology, attenuation, size, and structure (MASS) criteria: assessing response and predicting clinical outcome in metastatic renal cell carcinoma on antiangiogenic targeted therapy. *AJR Am J Roentgenol* 2010; 194:1470–1478.
3. Perrin T, Midya A, Yamashita R, et al. Short-term reproducibility of radiomic features in liver parenchyma and liver malignancies on contrast-enhanced CT imaging. *Abdom Radio* 2018; 43:3271–3278.
4. Smith AD, Lieber ML, Shah SN. Assessing tumor response and detecting recurrence in metastatic renal cell carcinoma on targeted therapy: importance of size and attenuation on contrast-enhanced CT. *AJR Am J Roentgenol* 2010; 194:157–165.
5. Krajewski KM, Guo M, Van den Abbeele AD, et al. Comparison of four early posttherapy imaging changes (EPTIC; RECIST 1.0, tumor shrinkage, computed tomography tumor density, Choi criteria) in assessing outcome to vascular endothelial growth factor-targeted therapy in patients with advanced renal cell carcinoma. *Eur Urol* 2011; 59:856–862.
6. Ammari S, Thiam R, Cuenod CA, et al. Radiological evaluation of response to treatment: application to metastatic renal cancers receiving anti-angiogenic treatment. *Diagn Interv Imaging* 2014; 95:527–539.
7. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45:228–247.
8. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000; 92:205–216.

9. Brufau BP, Cerqueda CS, Villalba LB, et al. Metastatic renal cell carcinoma: radiologic findings and assessment of response to targeted antiangiogenic therapy by using multidetector CT. *Radiographics* 2013; 33:1691–1716.
10. Benjamin RS, Choi H, Macapinlac HA, et al. We should desist using RECIST, at least in GIST. *J Clin Oncol* 2007; 25:1760–1764.
11. Choi H, Chamsangavej C, Faria SC, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007; 25:1753–1759.
12. Choi H, Chamsangavej C, de Castro Faria S, et al. CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. *AJR Am J Roentgenol* 2004; 183:1619–1628.
13. Benjamin RS, Choi H, Macapinlac HA, et al. We should desist using RECIST, at least in GIST. *J Clin Oncol* 2007; 25:1760–1764.
14. Dercle L, Lu L, Lichtenstein P, et al. Impact of variability in portal venous phase acquisition timing in tumor density measurement and treatment response assessment: metastatic colorectal cancer as a paradigm. *JCO Clin Cancer Inf* 2017; 1–8.
15. Dou Q, Chen H, Yu L, et al. Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection. *IEEE Trans Biomed Eng* 2017; 64:1558–1567.
16. Setio AAA, Traverso A, De Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal* 2017; 42:1–13.
17. Yasaka K, Akai H, Abe O, et al. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology* 2017. 170706.
18. Larson DB, Chen MC, Lungren MP, et al. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology* 2017. 170236.
19. Yan Z, Zhan Y, Peng Z, et al. Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition. *IEEE Trans Med Imaging* 2016; 35:1332–1343.
20. Van Cutsem E, Cervantes A, Nordlinger B, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2014; 25(Suppl_3):iii1–iii9.
21. Yao JC, Hassan M, Phan A, et al. One hundred years after “carcinoid”: epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol* 2008; 26:3063–3072.
22. Öberg K, Knigge U, Kwakkeboom D, et al. Neuroendocrine gastroenteropancreatic tumors: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2012; 23(Suppl_7):vii124–vii130.
23. Ramage JK, Ahmed A, Ardill J, et al. Guidelines for the management of gastroenteropancreatic neuroendocrine (including carcinoid) tumours (NETs). *Gut* 2012; 61:6–32.
24. Silverman PM, Brown B, Wray H, et al. Optimal contrast enhancement of the liver using helical (spiral) CT: value of SmartPrep. *AJR Am J Roentgenol* 1995; 164:1169–1171.
25. Soyer P, Pocard M, Boudiaf M, et al. Detection of hypovascular hepatic metastases at triple-phase helical CT: sensitivity of phases and comparison with surgical and histopathologic findings. *Radiology* 2004; 231:413–420.
26. Ganeshan B, Miles KA, Young RC, et al. Hepatic enhancement in colorectal cancer: texture analysis correlates with hepatic hemodynamics and patient survival. *Acad Radiol* 2007; 14:1520–1530.
27. Huang Y-q, Liang C-h, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol* 2016; 34:2157–2164.
28. Liang C, Huang Y, He L, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget* 2016; 7:31401.
29. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process Syst* 2012; 1097–1105.
30. Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database. In: *Computer Vision and Pattern Recognition, 2009 CVPR 2009 IEEE Conference on: IEEE; 2009. p. 248–255.*
31. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017; 39:1137–1149.
32. Birnbaum BA, Jacobs JE, Langlotz CP, et al. Assessment of a bolus-tracking technique in helical renal CT to optimize nephrographic phase imaging. *Radiology* 1999; 211:87–94.
33. Cademartiri F, Nieman K, van der Lugt A, et al. Intravenous contrast material administration at 16-detector row helical CT coronary angiography: test bolus versus bolus-tracking technique. *Radiology* 2004; 233:817–823.
34. Kim T, Murakami T, Hori M, et al. Small hypervascular hepatocellular carcinoma revealed by double arterial phase CT performed with single breath-hold scanning and automatic bolus tracking. *Am J Roentgenol* 2002; 178:899–904.
35. Silverman PM, Roberts S, Tefft M, et al. Helical CT of the liver: clinical application of an automated computer technique, SmartPrep, for obtaining images with optimal contrast enhancement. *AJR Am J Roentgenol* 1995; 165:73–78.
36. Soyer P, Pocard M, Boudiaf M, et al. Detection of hypovascular hepatic metastases at triple-phase helical CT: sensitivity of phases and comparison with surgical and histopathologic findings. *Radiology* 2004; 231:413–420.
37. Leggett DA, Kelley BB, Bunce IH, et al. Colorectal cancer: diagnostic potential of CT measurements of hepatic perfusion and implications for contrast enhancement protocols. *Radiology* 1997; 205:716–720.
38. Tirumani SH, Kim KW, Nishino M, et al. Update on the role of imaging in management of metastatic colorectal cancer. *Radiographics* 2014; 34:1908–1928.
39. Dercle L, Seban RD, Lazarovici J, et al. 18F-FDG PET and CT-scan detect new imaging patterns of response and progression in patients with Hodgkin lymphoma treated by anti-PD1 immune checkpoint inhibitor. *J Nucl Med* 2018; 59:15–24.
40. Champiat S, Dercle L, Ammari S, et al. Hyperprogressive disease is a new pattern of progression in cancer patients treated by Anti-PD-1/PD-L1. *Clin Cancer Res* 2017; 23:1920–1928.
41. Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. *J Clin Oncol* 2015; 33:3541–3543.
42. Michot JM, Mazon R, Dercle L, et al. Abscopal effect in a Hodgkin lymphoma patient treated by an anti-programmed death 1 antibody. *Eur J Cancer* 2016; 66:91–94.
43. Dercle L, Chisin R, Ammari S, et al. Nonsurgical giant cell tumour of the tendon sheath or of the diffuse type: are MRI or 18F-FDG PET/CT able to provide an accurate prediction of long-term outcome? *Eur J Nucl Med Mol Imaging* 2015; 42:397–408.
44. Seymour L, Bogaerts J, Perrone A, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017; 18:e143–ee52.
45. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016; 6:23428.
46. Dercle L, Ammari S, Bateson M, et al. Limits of radiomic-based entropy as a surrogate of tumor heterogeneity: ROI-area, acquisition protocol and tissue site exert substantial influence. *Sci Rep* 2017; 7:7952.
47. Limkin EJ, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* 2017.
48. Ganeshan B, Miles KA, Young RC, et al. Hepatic enhancement in colorectal cancer: texture analysis correlates with hepatic hemodynamics and patient survival. *Acad Radiol* 2007; 14:1520–1530.
49. Liang C, Huang Y, He L, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget* 2016; 7:31401–31412.
50. Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol* 2016; 34:2157–2164.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.acra.2019.02.024.