Development of deep neural network for individualized hepatobiliary toxicity prediction after liver SBRT

Bulat Ibragimov, Diego Toesca, and Daniel Chang

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA, USA

Yixuan Yuan

Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China

Albert Koong

Department of Radiation Oncology, MD Anderson Cancer Center, Houston, TX, USA

Lei Xing^{a)}

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA, USA

(Received 10 April 2018; revised 31 July 2018; accepted for publication 31 July 2018; published 10 September 2018)

Background: Accurate prediction of radiation toxicity of healthy organs-at-risks (OARs) critically determines the radiation therapy (RT) success. The existing dose–volume histogram-based metric may grossly under/overestimate the therapeutic toxicity after 27% in liver RT and 50% in head-and-neck RT. We propose the novel paradigm for toxicity prediction by leveraging the enormous potential of deep learning and go beyond the existing dose/volume histograms.

Experimental Design: We employed a database of 125 liver stereotactic body RT (SBRT) cases with follow-up data to train deep learning-based toxicity predictor. Convolutional neural networks (CNNs) were applied to discover the consistent patterns in 3D dose plans associated with toxicities. To enhance the predicting power, we first pretrain the CNNs with transfer learning from 3D CT images of 2644 human organs. CNNs were then trained on liver SBRT cases. Furthermore, nondosimetric pretreatment features, such as patients' demographics, underlying liver diseases, liver-directed therapies, were inputted into the fully connected neural network for more comprehensive prediction. The saliency maps of CNNs were used to estimate the toxicity risks associated with irradiation of anatomical regions of specific OARs. In addition, we applied machine learning solutions to map numerical pretreatment features with hepatobiliary toxicity manifestation.

Results: Among 125 liver SBRT patients, 58 were treated for liver metastases, 36 for hepatocellular carcinoma, 27 for cholangiocarcinoma, and 4 for other histologies. We observed that CNN we able to achieve accurate hepatobiliary toxicity prediction with the AUC of 0.79, whereas combining CNN for 3D dose plan analysis and fully connected neural networks for numerical feature analysis resulted in AUC of 0.85. Deep learning produces almost two times fewer false-positive toxicity predictions in comparison to DVH-based predictions, when the number of false negatives, i.e., missed toxicities, was minimized. The CNN saliency maps automatically estimated the toxicity risks for portal vein (PV) regions. We discovered that irradiation of the proximal portal vein is associated with two times higher toxicity risks (risk score: 0.66) than irradiation of the left portal vein (risk score: 0.31).

Conclusions: The framework offers clinically accurate tools for hepatobiliary toxicity prediction and automatic identification of anatomical regions that are critical to spare during SBRT. © 2018 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.13122>]

Key words: convolutional neural networks, liver cancer, SBRT, toxicity prediction

1. INTRODUCTION

Liver cancer is one of the deadliest cancers accounting for approximately 28,920 deaths in the US in 2017.¹ Moreover, the liver cancer incidence rate has been increasing with a 3.8% annual change, whereas the overall cancer incidence rate has been declining by −1.5% annual change. Stereotactic body radiation therapy (SBRT) has been rapidly growing into one of the main treatment strategies for liver cancer. Studies demonstrate 46–82% 2-yr survival rate for primary liver cancer,^{2–4} and 36–50% 2-yr survival rate for metastatic liver

cancer^{5–7} after SBRT treatments. Treatment fractionation is one of the factors for SBRT success allowing high doses to be delivered to the tumor during several iterations.⁸ Dose escalation to the tumor is, however, accompanied by extensive irradiation of the surrounding healthy organs-at-risk (OARs), which may result in toxicities of OARs. Acute and late toxicities can manifest after up to 30% of liver SBRTs significantly compromising patients' life after the treatment and potentially increasing mortality rates.^{9,10}

Radiation-induced toxicities of healthy organs-at-risk (OARs) surrounding tumor remain the main limiting factor

toward dose escalation during radiation therapy (RT).^{11–13} Regardless the RT modality used for cancer treatment, accurate assessment of potential toxicity risks represents one of the essential steps toward a successful treatment planning and subsequent dose delivery.^{13–16} In the current protocol, normal tissue complication probability (NTCP) is estimated by a dose/volume-based nomogram or models based on the assumption that an organ has a parallel or serial structure,^{12,17} where the NTCP of the OARs is numerically predicted by a power-law relation with the proportion of OAR irradiated.¹⁸ Such approaches for toxicity prediction rely on the use of oversimplified population-averaging models. The inherent functional or physiological heterogeneities of some OARs are tacitly ignored.¹⁹ Practically, clinical understanding of human anatomy suggests that irradiation of some regions of an OAR may result in higher toxicity risks than irradiation of other regions, as was recently demonstrated on head-and-neck cancers, where sparing of the consistently located regions of parotid glands known to harbor stem cells was shown to reduce the incidence of radiation-induced xerostomia.²⁰ Identification of such crucial regions and effective use of this information for dose optimization are urgently needed for the development of the next generation RT planning.

The recently developed concept of deep learning has been evolving into a powerful tool with close-to-human classification performance.²¹ Convolutional neural network — a concept from deep learning — revolutionized the field of computer vision²² and rapidly expanding into medical image analysis.²³ The analysis of dose plans for radiotherapy planning, however, remains an uninvestigated area. Dose plans are more challenging to be visually comprehended than natural images due to the relationships between dose plans and patient's anatomy, and the lack of clearly distinguishable dose features. Recently, Zhen et al.²⁴ proposed the first attempt to apply CNNs to predict toxicities from two-dimensional (2D) doses computed for the rectum surface. Working with 2D dose plans opens a room to use well-established CNN architectures with pretrained parameters and consequently compensate for potentially insufficient size of RT databases. At the same time, such an approach cannot be extrapolated into any organ. The rectum can be roughly approximated with a surface of revolution, so rectum unfolding follows the same algorithm for all patients and is unlikely to result in unexpected distortions in 2D dose plans. Unfolding the surface of an arbitrary object with closed shape will require stretching some surface parts resulting in considerable dose distortions. Moreover, surfaces of some objects, like multibranch vessels are not homeomorphic to cylinder or torus, and therefore cannot be unfolded into 2D image without ruptures. Our work aims to pioneer the concept of 3D dose plan analysis for volumetric toxicity prediction. To preserve the advantages of transfer learning, we collected a database of anatomy images for pretraining of 3D CNNs prior to toxicity prediction. The behavior of the resulting 3D CNN predictor is analyzed to identify the critical regions of the irradiated OAR.

The clinical question of the paper is predicting grade 3+ acute and late hepatobiliary (HB) toxicities after liver stereotactic body RTs (SBRTs). For this purpose, we rigidly registered segmented portal veins (PVs) of the SBRT patients to the reference anatomy, transformed 3D dose plans delivered to the PVs, used CNNs to identify consistent patterns in the aligned dose plans and associated the patterns with HB toxicities. The toxicity risks originated from irradiation of PV branches are then calculated. We additionally studied the correlation between numerical treatment features and HB toxicities by utilizing support vector machines (SVMs), random forests (RFs), and fully connected neural networks (FcNNs) and combined them with CNNs. After validation on a clinical database, we found out that CNNs are effective for HB toxicity prediction. We also observed that CNNs can be enhanced by feature-based prediction and reduces by half the number of false-positive toxicity predictions for highly sensitive prediction settings. Furthermore, CNNs discovered that irradiation of the right PV branch imposes two times higher toxicity risks than irradiation of the left PV branch.

2. MATERIALS AND METHODS

After the institutional review board approval, we assembled a database of patients treated with liver SBRT between July 2004 and November 2015 at our institution. The generated database included patients treated for primary liver cancer, mainly hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA), as well as patients treated for liver metastases. From each selected patient case, we retrieved treatment characteristics, RT simulation CT image, target and OARs segmentations, and SBRT treatment plans. We also retrieved pretreatment disease and demographic features, and complete follow-up records including a detailed assessment of acute and late post-treatment toxicities. To ensure that toxicity manifestation was solely related to the current liver SBRTs, all patients who had received liver or abdominal radiation therapy within 1 yr of SBRT were excluded.

2.A. Alignment of dose plans delivered to the portal vein

Let the set of SBRT cases be defined as $\mathbf{W} = \{\mathbf{I}, \mathbf{S}, \mathbf{D}\}$, where $\mathbf{I}$ is the set of liver CT images, $\mathbf{S}$ is the set of portal vein (PV) masks manually segmented from CT images $\mathbf{I}$ during the standard RT planning protocol, and $\mathbf{D}$ is the set of the executed SBRT dose plans. The dose plans $\mathbf{D}$ alone are not sufficient for prediction of toxicities because dose plans do not store anatomy information. A dose hot spot at the same location x on different dose plans $D_1, D_2 \in \mathbf{D}$ can correspond to completely different anatomical regions in $I_1, I_2 \in \mathbf{I}$. To compensate for such anatomical disagreement, we propose to align all dose plans to an arbitrary selected reference anatomy S_{REF} . As PVs are highly concave objects and the volume of PVs vary a lot from patient to patient, nonrigid registration of PV segmentation will result in highly distorted

transformation fields. Distorted transformations field will distort dose plans during alignment and the information on the total dose and dose-volume coverage will be lost. We avoided distortions by developing a rigid registration approach for PVs.

All PV masks $\mathbf{S}$ were first subdivided into four anatomical regions, namely the right branch, left branch, proximal branch, and central PV bifurcation. We then converted the four anatomical regions of each PV mask S_t into four sets of surface points $M_t = \{M_t^R, M_t^L, M_t^P, M_t^C\}$, where R, L, P, and C correspond 135 to right, left, proximal, and central PV regions. To compensate for translation differences among masks, the surface points for mask S_t are normalized against the s_t centerpoint of the central PV bifurcation M_t^C :

$$\begin{aligned} \mathcal{S}_t &= \frac{M_t^C}{|M_t^C|}; \bar{M}_t^R = M_t^R - \mathcal{S}_t, \bar{M}_t^L = M_t^L - \mathcal{S}_t, \\ \bar{M}_t^P &= M_t^P - \mathcal{S}_t, \bar{M}_t^C = M_t^C - \mathcal{S}_t. \end{aligned} \quad (1)$$

In the same way, the surface points for other PV masks are normalized, i.e., moved to the center of coordinate space, including the surface points $\bar{M}_{\text{REF}} = \{\bar{M}_{\text{REF}}^R, \bar{M}_{\text{REF}}^L, \bar{M}_{\text{REF}}^P, \bar{M}_{\text{REF}}^C\}$ for a reference mask S_{REF} . To compensate for rotation differences, the normalized surface points $\bar{M}_t$ are optimally rotated toward $\bar{M}_{\text{REF}}$ by minimizing:

$$\begin{aligned} Q_t = \operatorname{argmin}_Q & \left(\frac{\Delta(Q(\bar{M}_t^R), \bar{M}_{\text{REF}}^R)}{|\bar{M}_t^R| + |\bar{M}_{\text{REF}}^R|} + \frac{\Delta(Q(\bar{M}_t^L), \bar{M}_{\text{REF}}^L)}{|\bar{M}_t^L| + |\bar{M}_{\text{REF}}^L|} \right. \\ & \left. + \frac{\Delta(Q(\bar{M}_t^P), \bar{M}_{\text{REF}}^P)}{|\bar{M}_t^P| + |\bar{M}_{\text{REF}}^P|} + \frac{\Delta(Q(\bar{M}_t^C), \bar{M}_{\text{REF}}^C)}{|\bar{M}_t^C| + |\bar{M}_{\text{REF}}^C|} \right), \end{aligned} \quad (2)$$

where Q is a rotation matrix applied to its input, and Δ is the base of the squared symmetric surface distance defined as:

$$\begin{aligned} \Delta(K, M) &= \sum_{k \in K} \left(\min_{m \in M} \|k - m\| \right)^2 \\ &+ \sum_{m \in M} \left(\min_{k \in K} \|k - m\| \right)^2, \end{aligned} \quad (3)$$

where K and M are two sets of surface points. Rigid registration of S_t to S_{REF} was therefore formulated as rotating S_t against the centerpoint s_t until the distance between anatomical parts of S_t and S_{REF} is minimized. Due to the anatomical variability of PV, rigid registration will not result in a perfect alignment. At the same time, some level of misalignment would not necessarily be harmful for toxicity prediction. First, misalignments can be considered as a special type of noise to the data and learning on noisy data often leads to more robust models even on the cost of reduced prediction accuracy. Second, natural movements in the abdominal area do not allow doses to be delivered in exactly the same way they were defined in the treatment plan.

We rigidly aligned all dose plans $\mathbf{D} \in \mathbf{D} \setminus \mathbf{D}_{\text{REF}}$ to the reference anatomy S_{REF} according to the computed rotation matrices Q . The resulting doses are rotated accordingly so that

anatomical disagreement between dose plans is reduced (Figs. 1 and 2).

2.B. Dose plan analysis with convolutional neural networks

We designed a 3D CNN to search for the consistent pattern in the aligned dose plans $\mathbf{D}$ and associate such patterns with HB toxicities. The CNN methodological principles allow automated features extraction and remove a need to hand-craft dose features, which is an extremely nontrivial task due to lack of visually recognizable edges and corners in dose plans. The key component of CNNs is the convolution layer for generating simple intensity filters over the layer input that will be further combined into a more complex hierarchy. The first convolution layer with m filters forms m response layers, where the l -th filter $\mathbf{W}_1^l$ slides across the input dose plan and form response map $\mathbf{X}_1^l$:

$$\begin{aligned} X_1^l(i, j) &= \\ \operatorname{ReLU} & \left(\sum_{t=1}^{C_0} \sum_{p=1}^{P_1^h} \sum_{r=1}^{P_1^w} \mathbf{W}_1^{l,t}(p, r) \mathbf{D}(i - p + 1, j - r + 1) + b_1^l \right), \end{aligned} \quad (4)$$

where $\mathbf{W}_1^{l,t}$ is a convolution filter with a linear kernel of size P_1^h and P_1^w , b_1^l is additive bias and ReLU is a rectified nonlinear unit function defined as $\operatorname{ReLU}(x) = \max(x, 0)$. Parameter C_k correspond to the number of channels in the previous layer, in the case of the first convolution layer, $C_0 = 1$. Considering a potentially limited number of liver SBRTs, CNN overfitting may become an obstacle toward prediction of HB toxicities after liver SBRTs. As a result, the network may exhibit poor performance on new SBRT cases. We reduced overfitting by randomly dropout,²⁵ i.e., turning off some filters in convolution layers $\mathbf{X}$ with probability ρ . Two neighboring convolution layers are separated by a max-pooling layer for input downsampling. Max-pooling also introduces small shift invariance to potentially account for natural dose variability. A max-pooling layer $\mathbf{H}_{t+1}^l$ applied after convolution layers $\mathbf{X}_t^l$ is computed as:

$$\mathbf{H}_{t+1}^l(i, j) = \max_{i=[0, \dots, P_{t+1}^h], j=[0, \dots, P_{t+1}^w]} (\mathbf{X}_t^l(i - p + 1, j - r + 1)), \quad (5)$$

where P_{t+1}^h and P_{t+1}^w are max-pooling operator dimensions.

The CNN for dose plan analysis consisted of three sets of convolution layers with dropouts and two max-pooling layers separating the convolution layers. The second convolution layer was a bottleneck layer with fewer features than the first and third convolution layers, as bottleneck layers are shown to increase network robustness.^{26–28} The output of the CNN was defined as a binary value indicating if a patient is at high risk to develop an acute or late HB toxicity. The complete network architecture and layer parameters are summarized in Fig. 3, where all convolution and max-pooling layer have valid padding and the stride of 1, while the network training

FIG. 1. Examples of portal vein (PV) with superimposed SBRT doses. (a) The portal vein with the corresponding CT image and isodose curves. (b) A 3D dose plan delivered to the central hepatobiliary tract (15 mm expansion of the portal vein), which was analyzed using deep learning for hepatobiliary toxicity prediction. [Color figure can be viewed at wileyonlinelibrary.com]

Alignment of portal veins

Alignment of dose plans

FIG. 2. Alignment of 3D dose plans delivered to portal vein according to anatomical segments of the portal vein. Rigid registration is used to avoid deformations of dose plans. [Color figure can be viewed at wileyonlinelibrary.com]

was performed in 100 epochs. To enhance CNN performance, the training follows two stages: pretraining on general purpose 3D anatomy images, and tuning on augmented dose plans, where original dose plans were artificially modified with Gaussian noise and small rotation and translations. Such a strategy named transfer learning is increasingly used to empower CNNs to deal with problems with moderate-sized databases. For pretraining, we have assembled a database of 3D CT images with annotations depicting 2644 human organs of 28 types from Stanford University Hospital and three public databases.^{29–31} The organs were from head-and-neck region including larynx (number of samples = 45), eye globes (93), optic nerves (94), parotid glands (91), submandibular glands (62), mandible (50), optic chiasm (46), and pharynx (39); abdominal regions including adrenal glands (100), bladder (50), duodenum (53), gall bladder (50), inferior vena cava (50), large bowel (51), kidney (394), liver (173), pancreas (132), portal vein (125), rectum (50), small bowel (72), spleen (78), splenic vein (50), stomach (126), and uterus (50); spinal region including spinal cord (205) and

vertebra (135); aorta (50), chest wall (25), and esophagus (105). Using these 2644 organs, we generated 264400 training samples by extracting random regions of each organ. Each sample represented a $19 \times 19 \times 19$ size 3D image patch with isotropic 5 mm^3 resolution depicting a region of an organ from the anatomy database. A CNN was trained to automatically recognize the appearance of different organs by associating 264400 3D samples and 264400-element vector with organ codes depicted in the samples.

2.C. Treatment feature analysis

The risks of toxicities are known to depend not only on dose patterns but also on nondosimetric features. For example, patients with liver cirrhosis or/and hepatitis B are shown to develop the radiation-induced liver disease more often than patients without such underlying liver comorbidities.^{32,33} Demographic and genetic features are applicable for prediction of post-treatment fibrosis³⁴ and internal bleeding.³⁵ Potential of using machine learning on treatment features for

FIG. 3. The convolution neural network architecture developed for hepatobiliary toxicity prediction.

prediction of radiotherapy outcome has been shown on prostate and lung cancer examples.^{36–38}

We assembled a set of 10 explicit dosimetric features and 17 nondosimetric features for HB toxicity prediction. The dosimetric features corresponded to DVH bins at dose values of 10, 20... 100 Gy biologically effective doses (BED) multiplied by the volume of central HB tract (cHBT). We included the total SBRT dose and the number of SBRT fractions. Other features consisted of patient's gender and age, maximal tumor radius, the number of tumors, the volumes of liver, cHBT, tumor — gross tumor volume (GTV), tumor volume with clinical margins — planning target volume (PTV), underlying liver cirrhosis, HB stenting, hepatitis B (HBV) and C viruses (HCV), and concurrent to SBRT treatments including liver resection, chemoembolization, and chemotherapy. Three machine learning algorithms, namely SVMs, RFs, and FcNNs, were used to associate 27 treatment features with HB toxicities. The SVMs were trained using the sequential minimization algorithm with linear kernel and parameters $c = 0.05$, tolerance = 0.001, and $\varepsilon = 0.001$. The RFs consisting of 100 randomized trees of maximal depth equals five, and five randomly selected features per level. The FcNNs had the same architecture and the number of filters per layer as the CNNs with all convolution layers replaced with fully connected layers and all max-pooling layers removed.

2.D. Predicted toxicity risks for PV anatomical regions

The CNNs offer tools for identification of images parts that receive the highest CNN attention during classification.³⁹ In the case of HB toxicity prediction, such parts correspond to anatomical regions of the irradiated OAR that are most critical to be spared. We artificially modified dose plans to evaluate the CNN prediction changes. Having a selected dose plan $\vec{D}_t$ and spatial location $\mathbf{x}$, we compute the artificial dose plan $\vec{D}_t$ with the dose reduced around $\mathbf{x}$ and the artificial dose plan $\vec{D}_t$ with dose increased around $\mathbf{x}$:

$$\begin{aligned} \vec{D}_t(\mathbf{y}) &= 0 : \forall \|\mathbf{x} - \mathbf{y}\| < 2\sigma \\ \vec{D}_t(\mathbf{y}) &= \min\left(d_{\max}, D_t(\mathbf{y}) + d_{\max}e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}\right), \end{aligned} \quad (6)$$

where σ is a predefined parameter corresponding to the size of the modified area. Normalization against the maximum dose $d_{\max}$ delivered during SBRT treatment prevents situations when adding artificial doses to already highly irradiated PV regions results in PV receiving much higher dose than the tumor. We calculated the difference in CNN response for two artificial dose plans to estimate how critical is irradiation of the region around $\mathbf{x}$.

The presented above methodology was utilized to examine how irradiation of different subregions of PV contributes to the risk of HB toxicity. First, at a specific location of a subregion in PV, an artificial hot spot was added to each SBRT dose plans, and the proposed CNN-based toxicity prediction model was used to predict the corresponding toxicity risks. Second, we reduced the dose at the same location as the PV subregion for all SBRT plans region and again calculated the CNN-based toxicity risks from the modified dose plans. If the difference between the two prediction sets is small, it suggests that irradiation of the analyzed location of the PV subregion is associated with low toxicity risks, and *vice versa*. This procedure was repeated for all locations over PV subregions for all SBRT cases. To quantify irradiation risks, a scoring system was devised ζ_t^T :

$$\zeta_t^T = \frac{1}{|\mathcal{S}_t^T|} \sum_{\mathbf{y} \in \mathcal{S}_t^T} \left(\text{CNN}(\vec{D}_t(\mathbf{y})) - \text{CNN}(\vec{D}_t(\mathbf{y})) \right) \quad (7)$$

where t is the case number and $T \in \{R, L, P, C\}$ is the portal vein anatomical region for which the score is computed. The score 0 indicating that irradiation of a PV subregion does not change the risks of HB toxicity, and the score 1 indicating that irradiation of a subregion makes HB toxicity to be predicted with the complete certainty. The scores were computed over four anatomical regions of PV, namely right, left, and proximal PV branches, and central PV bifurcation.

3. RESULTS

3.A. Patient and treatment characteristics

We identified 150 patients treated with liver SBRT between July 2004 and November 2015 at our institution. After excluding 20 patients with incomplete follow-ups and

five patients with inconsistencies in the dose delivery plans in the MIM 6.5 software (MIM Software Inc., Cleveland, OH), we devised a database of 125 patient cases. Among 125 patients (63 males and 62 females), 58 (46.4%) were treated for liver metastases, 36 (28.8%) for HCC, 27 (21.6%) for CCA, and 4 (3.2%) for other primary liver tumor histologies. The acute and late grade 3+ HB toxicities were observed in 33 cases, respectively. Grade 3+ HB toxicities occurred on 9 patients treated for liver metastasis, 7 for HCC, 15 for CCA, and 2 for patients with other tumor histologies. The complete

list of demographic and treatment characteristics is given in Table I.

The CT images from the database were axially reconstructed with an in-plane voxel size of 0.830–1.523 mm and cross-sectional thickness of 0.8–3.27 mm. Each dose plan was of the same resolution as the corresponding CT image. Experienced radiation oncologists manually delineated each OAR structure and individual target structures on a dual-phase simulation CT image using the Eclipse Treatment Planning Software (Varian Medical Systems, Palo Alto, CA) or

TABLE I. Extended description of the SBRT database.

Characteristic	Number of cases (%) or median (range)			
	Metastasis	Hepatocellular carcinoma	Cholangiocarcinoma	Other
Number of cases	58	36	27	4
Gender				
Male	27 (47%)	24 (67%)	10 (37%)	3 (75%)
Female	31 (53%)	12 (33%)	17 (63%)	1 (25%)
Age, yr	59 [38–84]	66 [49–84]	69 [45–82]	61.5 [51–71]
SBRT dose, Gy	45 [25–54]	40 [30–50]	40 [26–50]	36.5 [30–42.5]
Number of fractions	3 [1–5]	5 [2–7]	5 [1–5]	5 [5]
Number of tumors				
Single	42 (72%)	32 (89%)	27 (100%)	4 (100%)
Multiple	16 (28%)	4 (11%)	0 (0%)	0 (0%)
Anatomy				
Max tumor radius, mm (range)	35 [10–111]	54 [19–120]	42 [10–75]	20 [15–50]
GTV volume, cc	18.8 [0.5–482.6]	60.5 [2–607.8]	35 [9.1–222.1]	42.4 [1.8–243.3]
PTV volume, cc	40.2 [6.6–608.5]	101.9 [10.7–875.5]	80.4 [23.2–280.4]	71.2 [7.9–326.9]
cHB tract volume, 10 cc	17.4 [9.2–26.3]	18.8 [12.6–27.7]	16.8 [8.2–23.4]	18.7 [17.3–28.4]
Liver volume, 100 cc	14.3 [7.6–33.8]	14.2 [6.9–46.1]	13.9 [7.2–24.3]	13.2 [10.9–21.8]
Biliary stent	2 (3.4%)	1 (2.8%)	13 (48%)	1 (25%)
Underlying liver diseases				
Cirrhosis	1 (1.7%)	24 (67%)	4 (15%)	0 (0%)
Chronic HCV	0 (0%)	20 (56%)	3 (11%)	0 (0%)
Chronic HBV	1 (1.7%)	5 (14%)	2 (7.4%)	0 (0%)
Other	0 (0%)	1 (2.8%)	1 (3.7%)	0 (0%)
Prior-to-SBRT liver treatments				
Liver resection	16 (28%)	8 (22%)	8 (30%)	4 (100%)
Chemoembolization	3 (5.2%)	33 (92%)	4 (15%)	0 (0%)
Radiofrequency ablation	14 (24%)	3 (8.3%)	1 (3.7%)	1 (25%)
Rad. therapy (>1 yr prior)	7 (12%)	1 (2.8%)	2 (7.4%)	0 (0%)
Non-liver rad. therapy (>1 yr prior)	20 (34%)	0 (0%)	0 (0%)	0 (0%)
Chemotherapy				
Prior to SBRT (>1 month)	43 (74%)	0 (0%)	7 (26%)	2 (50%)
Prior to SBRT (<1 month)	10 (17%)	1 (2.8%)	0 (0%)	0 (0%)
Child–Pugh class				
A (score 5–6)	49 (84%)	25 (69%)	15 (56%)	4 (100%)
B (score 7–9)	5 (8.6%)	11 (31%)	10 (37%)	0 (0%)
C (score 10–15)	1 (1.7%)	0 (0%)	1 (3.7%)	0 (0%)
Not accessible	3 (5.2%)	0 (0%)	1 (3.7%)	0 (0%)
Albumin–Bilirubin grade				
A1	—	3 (8.3%)	6 (22%)	3 (75%)
A2	—	28 (77.8%)	11 (41%)	1 (25%)
A3	—	5 (14%)	10 (37%)	0 (0%)

Multi-Plan Software (Accuray Inc., Sunnyvale, CA). All structures and radiation plans were exported to MIM 6.5 for dosimetric analysis with doses converted into biologically effective doses (BED_{10}) by applying the linear quadratic model with an α/β ratio of 10. An SBRT treatment was delivered in 1-5 consecutive daily fractions with the total prescribed doses ranged from 25 to 54 Gy, determined by the attending physicians. Liver dose constraints were the following: spare ≥ 700 cc of normal liver from receiving ≥ 15 Gy; spare ≥ 500 cc of normal liver from receiving ≥ 7 Gy; and limit the mean liver dose to < 15 Gy. The patients were followed up for 1–98 months with the median time of 13 months. The follow-up time has been mainly based on survival time, and patients with short follow-ups died shortly after the treatment. A patient, who did not develop acute 3+ grade HB toxicity and died less than 6 months after the treatment was considered toxicity-free in this study. During the follow-up period, acute and late HB, gastrointestinal and general toxicities were graded and documented following the National Cancer Institute Common Terminology Criteria for Adverse Event version 4.03.⁴⁰

3.B. Hepatobiliary toxicity prediction parameters

The manifestation of acute or late grade 3+ HB toxicity after liver SBRT was considered a positive toxicity event, whereas no, grade 1 or 2 HB toxicity after liver SBRT was considered a negative toxicity event. The manifestation of HB toxicities after liver SBRT is shown to correlate with irradiation of the cHBT ($P < 0.0001$).^{41,42} The cHBT was defined as an isometric 15-mm expansion of the portal vein (PV), from the porto-splenic and superior mesenteric vein confluence until the first bifurcations of the left and right PV branches ($cHBT = PV_{15}$). The standard DVH-based predictors for HB toxicity are computed from the doses delivered to cHBT, with cutoffs $V_{BED_{10}30} = 45.42$ cc and $V_{BED_{10}40} = 36.97$ cc, above which there is a high probability of grade 3+ HB toxicity. We investigated if using different machine learning approaches and studying individual dose patterns will result in more accurate or comparable toxicity prediction compared to the manual $V_{BED_{10}30}$ and $V_{BED_{10}40}$ criteria used in DVH-based toxicity estimation.

All experiments followed 20-fold cross-validation schema, meaning that all cases were divided into 20 subsets with similar size, and to predict the outcome of an SBRT from subset t , the remaining subsets were used for training. Such formulation means that all 125 cases were analyzed, ensuring that every testing SBRT was completely unfamiliar to the corresponding prediction model. The rigid registration performed prior to 3D dose plan analysis resulted in the average Dice coefficient of 0.53 and 0.73 for PV and cHBT, respectively. The CNNs were first pretrained on general anatomy images and then fine-tuned on 3D dose plans. For fine-tuning, all registered dose plans were unified into the same space by rescaling and cropping them into 3D images of $19 \times 19 \times 19$ size and isotropic 5 mm^3 resolution. Rescaling of the dose plans to a unified resolution ensured that

similar size dose hot spots had a similar appearance on any dose plan. The first reason for cropping and rescaling is memory restriction for CNN training. As the 3D dose plans contain empty space that accounts for more than 90% of a plan, it is unnecessarily and memory-demanding to store complete dose plans for training in the graphic processing unit. The second reason for cropping is to keep only doses delivered to the HB tract, as we know from previous clinical studies that doses outside the HB tract are not correlated with the HB toxicity manifestation. Although it is theoretically possible to use original size dose plans, operating with large size 3D volumes would dramatically complicate and slow down CNN training; moreover, this would not have much practical meaning as only a small portion of a patient's body gets irradiated. Generation of SVMs, RFs, and FcNNs also followed 20-fold cross-validation schema. The number of filters for CNN and FcNN layers, number of RF trees and maximal tree depth were considered nontrainable parameters of the predictors and were therefore defined prior to their training. Rationale for nontrainable parameter selection was found from previous studies and methodological principles of each predictor.^{39,43} For example, a high number of trees with the small tree depth is expected to reduce risks of RF overfitting. When tested on a randomly selected SBRT subset, moderate fluctuations of these parameters did not considerably change the performance of the predictors. To facilitate repetition of the analysis, two publicly available Microsoft toolboxes, namely CNTK⁴⁴ for implementation of CNNs and FcNNs and Sherwood⁴³ for implementation of RFs were utilized.

3.C. Hepatobiliary toxicity prediction results

The performance of all predictors was evaluated in terms of the area under the receiving operator characteristic curve (ROC of AUC). In addition, we estimated the behavior of the models in highly restrictive settings where a model is expected to miss no more than 0, 1, 2, 3, 4, or 5 of SBRT cases that resulted in HB toxicities, i.e., make 0, 1, 2, 3, 4, or 5 false-negative predictions. Information from 3D doses delivered to cHBT (AUC = 0.44 for CNN) was not sufficient to reliably predict toxicities after SBRT for metastatic liver cancer. This observation is in agreement with previous studies on HB toxicities showing no correlation between doses delivered to cHBT and HB toxicity manifestation for metastatic liver cancer cases.⁴² The accuracy of HB toxicities for primary liver cancer was of AUC of 0.789, 0.840, 0.786, and 0.759 for CNNs, FcNNs, RFs, and SVMs, respectively [Fig. 4(a)]. The obtained numbers were comparable or superior to the manual predictions through $V_{BED_{10}30}$, $V_{BED_{10}40}$, and combined $V_{BED_{10}30} \cup V_{BED_{10}40}$ that exhibited AUC of 0.805, 0.776, and 0.801, respectively [Fig. 4(b)]. We applied the weighted sum model to estimate the performance of hybrid predictors where CNNs, FcNNs, RFs, and SVMs were differently combined. The weights of the weighted sum model were of 0.5, 0.33, and 0.25 if two, three, and four predictors were combined, respectively. As a result, the best performance was observed when CNN and FcNN are combined

with AUC of 0.850, while adding RF or SVM did not improve the prediction accuracy. We estimated the number of false-positive predictions in highly restrictive conditions (Fig. 5). Augmenting FcNN with CNN resulted in statistically significant improvements ($P < 0.001$) for highly restrictive cases where 2, 1, or 0 FNs were allowed. The Pearson correlation coefficient ρ for CNNs, SVMs, RFs, FcNNs, combined CNN and FcNN, $V_{BED10}30$, and $V_{BED10}40$ predictors computed against the reference vector of actual toxicity manifestations were of 0.45, 0.48, 0.63, 0.51, 0.62, 0.50, and 0.49, respectively. The correlation coefficient of 1 would mean that a target predictor was in the perfect agreement with true toxicity manifestations.

3.D. Spatial maps of radiation-induced toxicity risk

The HB toxicity risks associated with irradiation of different PV regions were estimated automatically by using the CNNs, having $\sigma = 5$ mm for artificial dose hot spots generation. The central PV bifurcation (score: 0.680) and proximal PV (score: 0.656) showed to be the regions, irradiation of which most strongly correlated with manifestation of HB toxicity, while the risk score for the right PV branch was lower (score: 0.555), and irradiation of the left PV branch was shown the weakest correlation with HB toxicity (score: 0.311) (Fig. 6). All differences between scores for PV subregions were statistically significant ($P < 0.005$), except for the difference between the central bifurcation and proximal PV branch ($P = 0.413$). Note that the lowest score for the left PV branch is obtained despite the fact that the left branch receives in average higher dose than the proximal and right

PV branches. The mean BED_{10} received by the whole PV, and its proximal, left and right branches and central bifurcation was of 29.53 BED_{10} , 24.55 BED_{10} , 30.04 BED_{10} , 31.89 BED_{10} , and 37.23 BED_{10} , respectively. We therefore conclude that CNNs are not only able to predict HB toxicity manifestation but also identify the relative importance of different regions within the target OAR.

3.E. Features important for HB toxicity prediction

We studied how different features are associated with HB toxicity manifestation by analyzing the properties of SVM, RF and FcNN. The RF training is based on the idea of identifying features with the highest discriminative powers that would allow the best separation of the training samples according to their classes, i.e., HB toxicity manifestation values. At each training step, the optimal feature is selected to split training samples into two subsets, and the procedure is repeated for each subset until the RF stopping criteria is reached. The most discriminative features are therefore selected closer to tree roots and separate training samples into larger subsets. We can numerically estimate the contribution of a feature by counting the total number of samples it separated over the complete collection of trees in the forest. To estimate the feature importance predicted by SVM and FcNN, we generated sets of artificial cases using the cases from the SBRT database. A binary feature importance, e.g., history of cirrhosis, is computed as a difference between FcNN (SVM) predictions generated for the database cases where the feature values were artificially modified to true, and FcNN (SVM) prediction generated for the database cases

FIG. 4. ROC curves for the proposed deep dose analysis, alternative machine learning-based toxicity predictors and existing DVH-based predictors. (a) All SVM, RFs, FcNNs, CNNs and combined CNN and FcNN predictors achieved the Area under receiver operating characteristic curve (AUC) > 0.7 . (b) The combined predictor with AUC of 0.859 outperforms the existing DVH-based predictors $V_{BED10}30$, $V_{BED10}40$, and $V_{BED10}30 \cup V_{BED10}40$. The 95% confidence interval (95% CI) was reported for all AUC values. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 5. Performance of the proposed deep dose analysis using the number of false-positive and false-negative predictions. For example, when all SBRTs resulted in toxicities were identified (false negatives = 0), the random forest, fully connected, and convolutional neural networks make 18, 43, and 20 false-positive predictions. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

FIG. 6. Maps of crucial portal vein (PV) subregions irradiation of which correlates with hepatobiliary (HB) toxicity manifestation computed for five randomly selected patients. Among four PV anatomical regions, irradiation of the central bifurcation and proximal PV could most likely cause HB toxicity with the average risk scores of 0.680 and 0.656, while irradiation of the left PV branch could least likely cause HB toxicity with the average risk score of 0.311. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

where the feature values were artificially modified to false. A numeric feature importance is computed by iterating through its acceptable values and computing FcNN (SVM) predictions. We observed that DVHs cutoff thresholds for 30 and 40 BED₁₀ were among the most contributing features, which is in agreement with previous work on HB toxicity. Among nondosimetric features, the presence biliary stent, prior-to-SBRT chemoembolization and cHBT volume were the most contributing. The complete list of features, their contributions to HB toxicity measured through SVM, RF, and FcNN predictors are summarized in Table II. The obtained results are compared against individual contributions of features measured with *P*-values for logistic regression

4. DISCUSSION

Traditional toxicity prediction mainly relies on DVH nomograms that are deficient in studying the relationships between irradiation of different anatomical regions of OARs and toxicities. In the case of RT, the ultimate success of RT depends heavily on the actual dose distribution used for the patient treatment,⁴⁵ and therefore, without analyzing the delivered dose patterns, any predictive model would have limited accuracy and applicability. In this study, we have

pioneered the concept of 3D deep dose analysis for toxicity prediction and validated it on HB toxicities after liver SBRT. Moreover, we utilized SVM, RF and FcNNs to analyze numerical pre-SBRT features for more comprehensive toxicity prediction. Feature analysis results go alongside with previous studies showing the potential of radiomic^{46,47} and radiogenomic approaches^{48,49} for toxicity prediction.

Database collecting is among the main challenges for deep dose analysis because all analyzed patients should follow the same treatment protocol and have complete follow-ups to ensure the correct toxicity recording. Despite the moderate size of SBRTs database in this study, the accuracy of CNNs predictions was comparable to the accuracy of the existing DVH-based models. This observation confirms the hypothesis that information in dose distribution patterns is of importance for toxicity prediction. The risks for CNN overfitting were addressed by transfer learning from human anatomy database, the usage of bottleneck convolution layers, dropout layers and L2 regularization with coefficient 0.05. To compensate for potential imbalance between the number of cases with and without toxicity during training, dose maps for randomly selected toxicity cases were duplicated and augmented with Gaussian noise with zero mean and standard deviation equals to 1% of the maximal delivered dose. It is important

TABLE II. The correlation between individual features and hepatobiliary toxicity manifestation. The contribution of patient and treatment features into support vector machine (SVM), random forest (RF), and deep neural network (FcNN) predictors for hepatobiliary toxicity. All contributions are normalized to 100% and compared to correlation coefficients between individual features and HB toxicities.

Characteristic	Feature contribution in predictors (%)			Individual correlation (<i>P</i> value)	Characteristic	Feature contribution in predictors (%)			Individual correlation (<i>P</i> value)
	SVM	RF	FcNN			SVM	RF	FcNN	
cHBT V _{BED10} 10	0.93	4.91	1.07	<0.0001	Chronic HCV	2.45	1.08	4.92	0.2873
cHBT V _{BED10} 20	3.11	5.63	7.30	<0.0001	Chronic HBV	3.19	1.82	4.24	0.1729
cHBT V _{BED10} 30	2.81	5.79	11.06	<0.0001	Prior liver resection	2.81	3.06	3.06	0.0667
cHBT V _{BED10} 40	2.03	5.51	11.05	<0.0001	Prior chemoembolization	6.12	3.68	4.01	0.0058
cHBT V _{BED10} 50	1.28	5.48	6.96	<0.0001	Prior chemotherapy	1.51	1.53	2.21	0.2630
cHBT V _{BED10} 60	0.80	5.19	1.54	<0.0001	# tumors	2.19	0.92	0.25	0.3529
cHBT V _{BED10} 70	0.39	4.79	0.02	0.0001	Biliary stent	32.63	4.81	15.62	< 0.0001
cHBT V _{BED10} 80	0.15	4.45	0.20	0.0058	SBRT dose, Gy	0.40	4.57	3.88	0.0067
cHBT V _{BED10} 90	0.02	3.40	3.90	0.0458	# of fractions	1.04	2.94	0.02	0.0471
Gender	0.91	1.83	0.22	0.1886	GTV volume	7.45	4.32	0.86	0.0371
Age	1.70	3.80	4.02	0.0949	PTV volume	6.37	4.43	3.37	0.0810
Max tumor radius	1.21	4.51	0.03	0.0670	Liver volume	11.38	4.42	0.99	0.3527
Cirrhosis	1.75	2.75	1.07	0.0345	cHBT volume	3.91	4.39	8.11	0.0787

to note that opportunities for dose plan augmentation are restricted because dose plans are dependent on the patients' anatomy. Significant translations and rotations of dose plans may, therefore, invalidate the original SBRT treatments. As nontrainable parameters of each machine learning predictor affect the predictor performance, we can assume that a more optimal set of such parameters may exist potentially improving toxicity prediction results. It cannot be stated that the proposed CNN architecture is globally optimal among all possible architectures that can be developed for 3D dose plan analysis. At the same time, the main goals of this pioneering work were to investigate the principal applicability of CNNs for 3D dose plan analysis and to estimate if CNNs can augment feature-based toxicity prediction. Despite moderate accuracy, rigid registration of irradiated organs was shown to be an acceptable approach for 3D dose plan preprocessing for CNN classification. We did not expect cHBT to be perfectly aligned for all patients without performing nonrigid deformation that may potentially distort dose plans. Nonrigid deformations are especially likely to result in distortion for concave multibranch structures as portal veins. Moreover, 10–15% of patients have unusual portal vein branching configurations,⁵⁰ and nonrigid registration of standard and unusual portal veins will unavoidably result in distortions. At the same time, nonrigid registration may be a suitable option for analyzing dose patterns delivered to more convex structures with easy-to-align shapes, e.g., lung fields, liver, brain, etc. The selection of an appropriate registration algorithm must be based on the nature of the analyzed organs considering the tradeoff between dose alignment and dose distortions. Rigid registration ensures that dose/volume coverage of the target organs is preserved while different patients will not be perfectly aligned with each other, whereas nonrigid registration may change dose/volume coverage of organs while different patients will be better aligned. It is important to note that

certain level of dose/volume coverage changes or/and anatomy misalignment is not only harmless but also desirable and was artificially introduced to increase the robustness of the CNN predictor (See Material and Methods Section). We second observed that other machine learning approaches trained on pretreatment features also exhibit high toxicity prediction powers, with FcNNs outperforming all other approached including the existing manual models. The combination of CNNs with SVMs demonstrated AUC of 0.812, CNNs + RFs demonstrated AUC of 0.794, CNNs + FcNNs demonstrated AUC of 0.850 and CNNs + SVMs + RFs + FcNNs demonstrated AUC of 0.850. The superiority of the combination of CNNs and FcNNs against other pure and hybrid predictors suggests both the existence of some auxiliary information in dose distribution patterns that cannot be captured by numerical treatment features and lack of additional information added by SVMs or RFs when FcNNs are used. Similar to the previous studies, no correlation was found between doses over cHBT and HB toxicities for metastatic cases. The absence of underlying liver metabolic dysfunction, as seen in primary liver cancer cases, may explain this lack of a cHBT dose-effect for the occurrence of HB toxicities.

The radiation sensitivity within an OAR is often heterogeneous, but practical identification of the most crucial regions of an individual OAR is challenging.²⁰ Our next intriguing conclusion was that CNNs can identify and visualize the OAR regions irradiation of which is more (less) likely to result in HB toxicity. We observed that irradiation of the left PV branch had the least influence on HB toxicity. This influence does not depend on the total average dose delivered to the PV or the volume of the left branch for a particular case. Moreover, the average dose received by the left branch is higher than the average dose received by the proximal and right PV branches. In contrast, the proximal PV branch had

nearly the highest influence on HB toxicity but received the average dose of 83% of the total average dose delivered to PV. One of the possible explanations for these observations is the anatomical properties of hepatic ducts that along PV branches, which may be damaged by high doses. The left hepatic duct is responsible for bile drainage of liver segments II, III, and IV, accounting for only about one-third of the total liver volume. On the other hand, along the proximal PV segment runs the common hepatic duct and common bile duct after the cystic duct emergence. Radiation-induced damage to the common bile duct will compromise bile drainage from the whole liver which may be much more devastating than damaging the left duct. The anatomical rationale can be applied to understand the critical importance to spare the central PV bifurcation. Being geometrically located in the center of PV, the bifurcation is likely to receive higher radiation doses. We can conclude that the proximal PV should receive higher priority in sparing than the area around left PV during the planning process for liver SBRT. It must be noted that the above-presented explanation is purely based on the liver anatomy and the conclusion should be additionally supported by clinical analysis of PV and cHBT in the future. To the best of our knowledge, there have been no clinical 424 studies determining the relative importance of PV branch sparing during RTs.

Finally, we studied the contribution of nondosimetric features in SVM, RFs, and FcNNs predictors. We observed that the presence of HB stent contributed the most to SVM and FcNNs, and exhibited strong individual predictive power ($P < 0.0001$). We can assume that the patients with HB stents already had issues with the biliary tract, which makes them more likely to develop HB complications after liver SBRT. Similarly, prior chemoembolization features were extensively used by all predictor and have a statistically significant correlation with HB toxicity manifestation ($P = 0.0058$). Delivering higher doses to the tumor increases the risk of normal tissues receiving higher doses as well, and therefore is shown to be predictive for toxicity risks by SVM and FcNN. We also observed that patient's gender, the number of tumors, prior chemotherapy did not correlate with HB toxicities and marginally contributed to SVM, RFs, and FcNNs. Although SVMs, RFs, and FcNNs agree when estimating the contribution of most of the 435 analyzed features, they disagree for some features, e.g., SVM considers liver volume to be a significant 436 predictor for toxicity, whereas FcNN does not consider it to be very informative. Certain disagreement 437 between predictors can be expected due to different machine learning mechanisms used for each predictor 438. Moreover, as SVMs have the worst performance among all predictors, we can expect them to deviate from 439 RFs and FcNNs in estimating different feature contribution. Considerable disagreement between 440 predictors observed on several features can be also explained by the limited database size and a limited 441 pool of feature values.

In summary, we showed the potential of 3D deep dose analysis for toxicity prediction after liver SBRTs. Being augmented by treatment features, deep dose analysis can open a room toward personalized SBRT planning. Finally, we

emphasize that the presented approach is quite general and applicable to other toxicity and cancer types. Future research directions include enriching the liver SBRT database and checking the performance of the toxicity prediction framework on recently treated liver cancer patients as well as extending the framework applicability to other cancer types and post-SBRT outcomes.

ACKNOWLEDGMENTS

This work was partially supported by NIH (1R01 CA176553 and EB016777), and Varian research grants.

^{a)}Author to whom correspondence should be addressed. Electronic mail: lei@stanford.edu

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin.* 2017;67:7–30.
2. Bibault J-E, Dewas S, Vautravers-Dewas C, et al. Stereotactic body radiation therapy for hepatocellular carcinoma: prognostic factors of local control, overall survival, and toxicity. *PLoS ONE.* 2013;8:e77472.
3. Jang WI, Kim MS, Bae SH, et al. High-dose stereotactic body radiotherapy correlates increased local control and overall survival in patients with inoperable hepatocellular carcinoma. *Radiat Oncol.* 2013;8:250.
4. Sanuki N, Takeda A, Oku Y, et al. Stereotactic body radiotherapy for small hepatocellular carcinoma: a retrospective outcome analysis in 185 patients. *Acta Oncol.* 2014;53:399–404.
5. Høyer M, Swaminath A, Bydder S, et al. Radiotherapy for liver metastases: a review of evidence. *Int J Radiat Oncol Biol Phys.* 2012;82:1047–1057.
6. Scorsetti M, Arcangeli S, Tozzi A, et al. Is stereotactic body radiation therapy an attractive option for unresectable liver metastases? A preliminary report from a phase 2 trial. *Int J Radiat Oncol Biol Phys.* 2013;86:336–342.
7. Scorsetti M, Clerici E, Comito T. Stereotactic body radiation therapy for liver metastases. *J Gastrointest Oncol.* 2014;5:190–197.
8. Murray LJ, Dawson LA. Advances in stereotactic body radiation therapy for hepatocellular carcinoma. *Semin Radiat Oncol.* 2017;27:247–255.
9. Bujold A, Massey CA, Kim JJ, et al. Sequential phase I and II trials of stereotactic body radiotherapy for locally advanced hepatocellular carcinoma. *J Clin Oncol.* 2013;31:1631–1639.
10. Takeda A, Sanuki N, Eriguchi T, et al. Stereotactic ablative body radiotherapy for previously untreated solitary hepatocellular carcinoma. *J Gastroenterol Hepatol.* 2014;29:372–379.
11. Rusthoven KE, Kavanagh BD, Cardenes H, et al. Multi-institutional phase I/II trial of stereotactic body radiation therapy for liver metastases. *J Clin Oncol.* 2009;27:1572–1578.
12. Wang X, Hu C, Eisbruch A. Organ-sparing radiation therapy for head and neck cancer. *Nat Rev Clin Oncol.* 2011;8:639–648.
13. Lacas B, Bourhis J, Overgaard J, et al. Role of radiotherapy fractionation in head and neck cancers (MARCh): an updated meta-analysis. *Lancet Oncol.* 2017;18:1221–1237.
14. Hong TS, Wo JY, Yeap BY, et al. Multi-institutional phase II study of high-dose hypofractionated proton beam therapy in patients with localized, unresectable hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J Clin Oncol.* 2016;34:460–468.
15. Sciubba JJ, Goldenberg D. Oral complications of radiotherapy. *Lancet Oncol.* 2006;7:175–183.
16. Baumann M, Krause M, Overgaard J, et al. Radiation oncology in the era of precision medicine. *Nat Rev Cancer.* 2016;16:234–249.
17. Lambin P, van Stiphout RG, Starmans MH, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat Rev Clin Oncol.* 2013;10:27–40.

18. Bentzen SM, Constine LS, Deasy JO, et al. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat Oncol Biol Phys.* 2010;76:S3–S9.
19. Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. *Lancet Oncol.* 2010;11:85–91.
20. van Luijk P, Pringle S, Deasy JO, et al. Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer. *Sci Transl Med.* 2015;7:147.
21. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–444.
22. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis.* 2015;115:211–252.
23. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* 2017;19:221–248.
24. Zhen X, Chen J, Zhong Z, et al. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. *Phys Med Biol.* 2017;62:8246.
25. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929–1958.
26. Yoshioka T, Takiguchi T, Arikawa Y, Duffner S, Garcia C. Convolutional bottleneck network with dropout for dysarthric speech recognition. *Trans Mach Learn Artif Intell.* 2014;2:1–15.
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition; 2015. ArXiv1512.03385 Cs.
28. Wang M, Liu B, Foroosh H. Design of Efficient Convolutional Layers using Single Intra-channel Convolution, Topological Subdivisioning and Spatial ‘Bottleneck’ Structure; 2016. ArXiv1608.04337 Cs.
29. Ibragimov B, Likar B, Pernuš F, Vrtovec T. Shape representation for efficient landmark-based segmentation in 3-d. *IEEE Trans Med Imaging.* 2014;33:861–874.
30. Multi-atlas labeling beyond the cranial vault - workshop and challenge – syn3193805. Available at: <https://www.synapse.org/#!/Synapse:syn3193805/wiki/>. (Accessed: 14th September 2017).
31. Roth HR, Lu L, Farag A. DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. *Lect Notes Comput Sci.* 2015;000:556–564. ArXiv1506.06448 Cs
32. Wu D-H, Liu L, Chen L-H. Therapeutic effects and prognostic factors in three-dimensional conformal radiotherapy combined with transcatheter arterial chemoembolization for hepatocellular carcinoma. *World J Gastroenterol.* 2004;10:2184–2189.
33. Pan CC, Kavanagh BD, Dawson LA, et al. Radiation-associated liver injury. *Int J Radiat Oncol.* 2010;76:S94–S100.
34. Edvardsen H, Landmark-Høyvik H, Reinertsen KV, et al. SNP in TXNRD2 associated with radiation-induced fibrosis: a study of genetic variation in reactive oxygen species metabolism and signaling. *Int J Radiat Oncol Biol Phys.* 2013;86:791–799.
35. Kerns SL, Stock RG, Stone NN, et al. Genome-wide association study identifies a region on chromosome 11q14.3 associated with late rectal bleeding following radiation therapy for prostate cancer. *Radiother Oncol.* 2013;107:372–376.
36. Gulliford SL, Webb S, Rowbottom CG, Come DW, Dearnaley DP. Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate. *Radiother Oncol.* 2004;71:3–12.
37. Chen S, Zhou S, Yin F-F, Marks LB, Das SK. Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis. *Med Phys.* 2007;34:3808–3814.
38. Pella A, Cambria R, Riboldi M, et al. Use of machine learning methods for prediction of acute toxicity in organs at risk following prostate radiotherapy. *Med Phys.* 2011;38:2859–2867.
39. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks; 2013. ArXiv1311.2901 Cs.
40. National cancer institute. Common terminology criteria for adverse events (CTCAE) version 4.0 : U.S. department of health and human services; national institutes of health; national cancer institute. Available at: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf. (Accessed: 26th May 2017).
41. Osmundson EC, Wu Y, Luxton G, Bazan JG, Koong AC, Chang DT. Predictors of toxicity associated with stereotactic body radiation therapy to the central hepatobiliary tract. *Int J Radiat Oncol Biol Phys.* 2015;91:986–994.
42. Toesca DA, Osmundson EC, Eyben RV, et al. Central liver toxicity after SBRT: an expanded analysis and predictive nomogram. *Radiother Oncol.* 2016;122:130–136.
43. Criminisi A, Shotton J, Konukoglu E. *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning.* Berlin: Springer; 2012.
44. Agarwal A, Akchurin E, Basoglu C, et al. An introduction to computational networks and the computational network toolkit. *Microsoft Tech. Rep;* 2014. MSR-TR-2014-112.
45. Toesca DAS, Osmundson EC, von Eyben R, Shaffer JL, Koong AC, Chang DT. Assessment of hepatic function decline after stereotactic body radiation therapy for primary liver cancer. *Pract Radiat Oncol.* 2017;7:173–182.
46. Itakura H, Achrol AS, Mitchell LA, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. *Sci Transl Med.* 2015;7:138.
47. Fehr D, Veeraraghavan H, Wibmer A, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proc Natl Acad Sci USA.* 2015;112:E6265–E6273.
48. Yu K-H, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun.* 2016;7:12474.
49. Aerts HJWL. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol.* 2016;2:1636–1642.
50. Madoff DC, Hicks ME, Vauthey JN, et al. Transhepatic portal vein embolization: anatomy, indications, and technical considerations. *RadioGraphics.* 2002;22:1063–1076.