Received: 11 August 2022 | Revised: 7 March 2023 | Accepted: 15 March 2023

DOI: 10.1002/mp.16386

RESEARCH ARTICLE

MEDICAL PHYSICS

Deep learning prediction of post-SBRT liver function changes and NTCP modeling in hepatocellular carcinoma based on DGAE-MRI

Lise Wei[1] | **Madhava P. Aryal**[1] | **Kyle Cuneo**[1] | **Martha Matuszak**[1] | **Theodore S. Lawrence**[1] | **Randall K. Ten Haken**[1] | **Yue Cao**[1] | **Issam El Naqa**[2]

[1]Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan, USA

[2]Moffitt Cancer Center, Tampa, Florida, USA

Correspondence
Lise Wei
Email: liswei@umich.edu

Funding information
National Institutes of Health, Grant/Award Numbers: P01-CA059827, R01-CA233487

Abstract

Background: Stereotactic body radiation therapy (SBRT) produces excellent local control for patients with hepatocellular carcinoma (HCC). However, the risk of toxicity for normal liver tissue is still a limiting factor. Normal tissue complication probability (NTCP) models have been proposed to estimate the toxicity with the assumption of uniform liver function distribution, which is not optimal. With more accurate regional liver functional imaging available for individual patient, we can improve the estimation and be more patient-specific.

Purpose: To develop normal tissue complication probability (NTCP) models using pre-/during-treatment (RT) dynamic Gadoxetic Acid-enhanced (DGAE) MRI for adaptation of RT in a patient-specific manner in hepatocellular cancer (HCC) patients who receive SBRT.

Methods: 24 of 146 HCC patients who received SBRT underwent DGAE MRI. Physical doses were converted into EQD2 for analysis. Voxel-by-voxel quantification of the contrast uptake rate (k1) from DGAE-MRI was used to quantify liver function. A logistic dose-response model was used to estimate the fraction of liver functional loss, and NTCP was estimated using the cumulative functional reserve model for changes in Child-Pugh (C-P) scores. Model parameters were calculated using maximum-likelihood estimations. During-RT liver functional maps were predicted from dose distributions and pre-RT k1 maps with a conditional Wasserstein generative adversarial network (cWGAN). Imaging prediction quality was assessed using root-mean-square error (RMSE) and structural similarity (SSIM) metrics. The dose-response and NTCP were fit on both original and cWGAN predicted images and compared using a Wilcoxon signed-rank test.

Results: Logistic dose response models for changes in k1 yielded D50 of 35.2 (95% CI: 26.7–47.5) Gy and k of 0.62 (0.49–0.75) for the whole population. The high baseline ALBI (poor liver function) subgroup showed a significantly smaller D50 of 11.7 (CI: 9.06–15.4) Gy and larger k of 0.96 (CI: 0.74–1.22) compared to a low baseline ALBI (good liver function) subgroup of 54.8 (CI: 38.3–79.1) Gy and 0.59 (CI: 0.48–0.74), with p-values of < 0.001 and $= 0.008$, respectively, which indicates higher radiosensitivity for the worse baseline liver function cohort. Subset analyses were also performed for high/low baseline CP subgroups. The corresponding NTCP models showed good agreement for the fit parameters between cWGAN predicted and the ground-truth during-RT images with no statistical differences for low ALBI subgroup.

Conclusions: NTCP models which incorporate voxel-wise functional information from DGAE-MRI k1 maps were successfully developed and feasibility was

 © 2023 American Association of Physicists in Medicine.

demonstrated in a small patient cohort. cWGAN predicted functional maps show promise for estimating localized patient-specific response to RT and warrant further validation in a larger patient cohort.

KEYWORDS
deep learning, DGAE-MRI, HCC, NTCP

1 | INTRODUCTION

Stereotactic body radiation therapy (SBRT) is now a widely applied treatment,[1–4] and, in particular produces excellent local control for patients with hepatocellular carcinoma (HCC).[5,6] However, the risk of toxicity for uninvolved liver currently limits the dose that can be safely delivered, which potentially will lead to a reduction in local control.[7,8] An improved understanding/ability to predict the local dose-response relationship of uninvolved liver could help to guide more effective RT where delivery is modulated to spare functionally sensitive healthy liver.

Many attempts have been made to develop normal tissue complication probability (NTCP) models to minimize the risk of liver toxicity.[9] Earlier, the QUANTEC (Quantitative Analysis of Normal Tissue Effects in the Clinic) has provided recommendations for radiation induced liver toxicity estimation based on consensus data.[10] However, the data at that time were mostly from conventional fractionation regimes. More recently, HyTEC reviewed the normal tissue dose-volume effects for SBRT-induced liver toxicities.[11] In these models, dose distribution, volume, and outcome data can be used to fit complication probabilities to the observed frequencies of occurrence for liver toxicity in the patient population to determine tolerable normal liver doses. Also, a parallel architecture model was developed for organs whose function is carried out in parallel subunits.[12] In this model, a treatment is assumed to cause a complication if the fraction of the organ damaged exceeds the functional reserve of the patient. The NTCP for the treatment is then the proportion of patients with functional reserves less than the fraction of the organ damaged.[12] However, due to the lack of local functional data at that time, assumptions of uniform function across the liver and uniform radiosensitivity were applied. Later, when functional imaging data became available, the functional loss parameters (usually logit model parameters) could be fit to the functional imaging data themselves, which is more accurate than the uniform distribution of function assumption. There are different types of functional imaging studies, including SPECT/CT, portal venous perfusion maps from dynamic contrast enhanced (DCE) MRI, or dynamic Gadoxetic Acid-enhanced (DGAE) MRI, etc. Miften *et al*. developed a method that incorporates tumor heterogeneous functionality using FDG-PET images and SPECT perfusion data for critical structures by expanding the EUD concept by FEUD.[13] A previous work used the portal venous perfusion maps for the parallel architecture model.[14] Bowen *et al*. characterized patient-specific radiation dose-response relationships as surrogates for liver radiosensitivity based on SPECT/CT images.[15] Though these studies contribute to the understanding of populational dose response, they did not assess regional function directly and none of them predicted the post-RT regional function.

Therefore, we carried out a study using the direct measurement of regional hepatic function by Gadoxetic Acid-enhanced (DGAE) MRI. Although portal venous perfusion significantly correlates with overall function evaluated via indocyanine green (ICG) clearance or retention,[16] this is an indirect surrogate for liver function. Gadoxetic Acid is a hepatobiliary contrast-based imaging agent, where the contrast is taken up into the liver cells, allowing more direct interrogation of liver function. Compared with SPECT images, MRI images in general have better resolution. Recent studies showed that DGAE MRI is significantly correlated with global liver function (ICGR, ALBI).[17–20] There are also studies showing the significant correlation of patient outcome (change of global liver function) using DGAE MRI data.[21–23] In addition, we developed a deep learning method to predict the voxel-by-voxel post-RT liver function (i.e., a conditional generative adversarial network [cGAN] with Wasserstein loss [cWGAN]), which could be useful in treatment planning for adaptive radiotherapy. Potentially, this prediction could be used pre-treatment to optimize personalized treatment plans so that one could not only avoid irradiating the high functional regions but also avoid high sensitivity regions to achieve better treatment outcomes.

2 | MATERIALS AND METHODS

2.1 | Patient characteristics

A total of 146 patients who had received SBRT were retrospectively enrolled in this study on institutional review board approved protocols. Twenty four of these patients have pre/post-RT DGAE MRI images. An adaptive treatment was designed to deliver 3 or 5 fractions, based on Indocyanine Green (ICG) clearance at 15-min (ICG-R15).[15] Global liver function is quantified by Child-Pugh (CP) and Albumin-Bilirubin (ALBI) scores. A summary

TABLE 1 Patient characteristics.

Age (y)	
Median (range)	67.0 (50.5–88.7)
Sex	
Female/male	5/19
Cirrhotic status	23/24
Mean liver-GTV EQD2 (Gy)	
Median (range)	6.8 (1.1–18.0)
Prescription doses (Gy)	
Median (range)	33.0 (19.5–50.0)
Pre-RT CP score	
Median (range)	6 (5–10)
Pre-RT ALBI Score	
Median (range)	−2.14 (−3.56–1.09)
Post-RT CP score	
Median (range)	7 (5–10)
Post-RT ALBI score	
Median (range)	−1.73 (−3.1 to −0.14)
GTV size (cm^3)	
Median (range)	47.9 (4.1–1157.4)
Fraction number	
Median (range)	3 (3–5)

of the patient characteristics is given in Table 1. The radiation dose distributions for the total uninvolved liver exclusive of the gross tumor volume were computed using the Varian Eclipse treatment planning system (v15) and the AAA algorithm. All total dose values were converted to their 2-Gy equivalents (EQD2) using the linear-quadratic linear model with an $\alpha/\beta = 2.5$ Gy (α/β of 2–3 Gy is commonly used for normal liver tissue to convert physical dose into 2-Gy equivalent dose).[24–27] We used median ALBI score to stratify patients into good/poor liver function subgroup.

2.2 | Imaging acquisition and processing

The MRI scans were acquired on a 3T scanner (Skyra, Siemens Healthineers) within 2 weeks before RT and 1 month after 3 or 5 fractions of SBRT. The acquisition parameters were ~420 mm field, 64 axial slices with ~3.5 mm thickness, in-plane resolution of ~2.2 × 2.2 mm, TE/TR~1.16/2.68 ms, flip angle of 13.5°, and a median temporal sampling rate of 8.8 s. An overdetermined method of rigid-body transformation was used to register the image volumes to correct motion.[28] The voxel-wise uptake rate (k1 values) were calculated by fitting the images to a single-input two-compartment model that was shown to be more robust and computationally faster than the dual-input two compartment model.[29] The obtained k1 maps were used as a potential imaging biomarker for local liver function with axial slice thickness ~3.5 mm, in-plane resolution ~2.2 × 2.2 mm. Co-registration (rigid registration) of MRI volumes pre- and post-SBRT were performed using an in-house software (imFIAT) that has been evaluated in collaborative projects of the Quantitative Image Network of NCI and granted level 5 benchmark by NCI. Gross tumor volumes (GTVs) after uniform expansion of 1 cm were excluded in the training and NTCP modeling to eliminate the GTVs and any potentially involved proximal tissue from analysis. Voxels with values greater than the top 1 % quantile threshold of the intensities were set to the threshold values, and an erosion process (1 cm) was applied to remove the artifacts for liver mask due to hyperintense artifacts for DGAE images. Motion effects are especially apparent at the edges of the liver and liver vasculature. In these locations, slight motion can cause apparent jumps in uptake as a motion artifact.[29] Since these hyperintense artifacts are outliers, it is necessary to remove these voxels to get reasonable results. An automated segmentation approach based on Otsu's multithreshold histogram decomposition was used for the subvolumes extraction.[30] The morphologic operators were used to remove small and isolated clusters in the map. For the neural network input, images were standardized and resized to the size of 116 × 101 × 51, which is the largest dimension of liver for all the patients.

2.3 | Conditional GAN (cGAN) and conditional Wasserstein GAN (cWGAN) for post-RT k1 map prediction

Studies have shown that pure taxicab norm (L1) and Euclidean norm (L2) (L1/L2 norm) loss will result in blurry images since these distances are minimized by averaging.[31] GAN loss can bypass this issue because it preserves both high and low frequency components. It can capture more detailed and high frequency information, while L1/L2 loss mainly learns the low frequency part. The combination of GAN loss and L1 loss can generate a sharp and realistic prediction.[32] In our experiment, we adopted this GAN and L1 loss combination. We applied a cGAN, which learns a conditional generative model, conditioning on pre-RT k1 map and incorporating dose distribution with pre-RT k1 map in the generator. The objective of the cGAN is

$$L_{cGAN}(G, D) = \mathrm{E}_{preRT\ k1, postRT\ k1}[\log D(preRT\ k1, postRT\ k1)] + E_{preRT\ k1, dose}[\log(1 - D(preRT\ k1, G(preRT\ k1, dose)))], \quad (1)$$

where G tries to minimize this objective against an adversarial D that tries to maximize it. To mix with a

reconstruction L1 loss, the generator is optimized using the GAN loss as well as the L1 loss. So, it can force the output to be close to the ground truth.

$$L_{L1}(G) = E_{preRT\ k1,postRT\ k1,dose} \times [\|postRT\ k1 - G(preRT\ k1, dose)\|_1], \quad (2)$$

The original GAN loss minimizes the Jensen-Shannon (JS) divergence between the true data distribution and the one assumed by the generator. The training sometimes can be unstable and difficult in practice because the training is hard to converge when the model trained distribution differs from the data distribution too much. In detail, it will result in a vanishing gradient and the model cannot be optimized further. The original GAN proposed an alternative cost function to address the gradient vanishing problem, but the new function has large variance of gradients that makes the model unstable. Wasserstein-1 distance was found to be a more appropriate distance to measure two distributions, which has the minimum cost of transporting mass in converting one data distribution to another data distribution.[33,34] It provides a continuous mapping from the parameter space to data distribution, and thus has a usable gradient that eliminates the gradient vanishing problem for original GAN.[35] To ensure the discriminator function to be K-Lipshitz, penalization of the gradient was used, which has been called WGAN with gradient penalty.[36] Both the original cGAN and cWGAN were explored. The cWGAN objective is

$$L_{cWGAN}(G, D) = E_{preRT\ k1,postRT\ k1}[f(preRT\ k1, postRT\ k1)] - E_{preRT\ k1,dose}[f(G(preRT\ k1, dose))], \quad (3)$$

The general shape of a "U-Net" with skip connections was used for the generator (4 layers). The convolutional kernel size was 4, stride was 2, with padding of 1. Both generator and discriminator use modules of the form convolution-BatchNorm-ReLu. The training of G and D was alternated for vanilla GAN, and the objective of D was divided by 2 to slow down the rate relative to G. Adam optimizer was used with initial learning rates of 0.002 for generator and 0.0002 for discriminator, and β_1 is 0.5, β_2 is 0.999. Batch size is 1. The learning rate decreased linearly, after 100 epochs of training. The discriminator was updated $n = 4$ times for every generator update for WGAN, other training strategies are the same. The data augmentation method of random crop was used during training, with input image dimension being $96 \times 96 \times 32$.

Root mean square error (RMSE) and structural similarity (SSIM) metrics were used to evaluate the difference between ground truth and predicted images. Gamma analysis was used to summarize the distance to agreement between voxels in the post-RT map and predicted post-RT map. Passing criterion is voxels in the predicted map are within 5% threshold intensity difference and 2 voxels distance (slice thickness) in the post-RT map. The two-sided Wilcoxon test was performed for all high and low ALBI groups to compare the cGAN and cWGAN prediction results. Five-fold cross validation was used for the model evaluation (19 patients in the training set and 5 patients in the test set for each fold).

2.4 | NTCP modeling and calculation

NTCP modeling was based on a generalization of the parallel architecture (PA).[12] The complication that NTCP is computing is the ALBI score change. We chose a 1-grade change in ALBI as an event. ALBI grade offers a simple, evidence-based, objective method of assessing liver function. It was found that a one-grade increase in ALBI score was associated with a significantly increased risk for poor prognosis in liver disease patients.[37,38] NTCP is represented by the cumulative functional damage for the population, which is modeled by the formula below:

$$NTCP = H(f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(f-f_{50})/\gamma} exp\left[-\frac{x^2}{2}\right] dx, \quad (4)$$

where f is the fraction of the organ damaged by the treatment, f_{50} is the fraction of the total organ damaged which would produce a 50% complication rate, $\gamma$the steepness of the organ response, and $H(f)$ is the cumulative distribution of functional reserves in the patient population. Assuming organs are composed of numerous subunits that respond to radiation independently, and that these subunits are small enough that they receive homogeneous dose. The subunit damage is calculated by the change of k1 map (pre-/post-RT), which quantifies the regional liver function. There have been studies showing quantification of hepatobiliary agent uptake rate can be used as a direct measure of regional liver function.[17,29,39–42] because hepatobiliary agents, such as the one used in this study, are taken up by the hepatocytes compared to other vascular contrast agents which are not taken up by the hepatocytes and thus not a direct analog to liver function. Binning of the input was performed. The subunits of the k1 maps were extracted using an automated segmentation approach based on the optimal Otsu's multi-threshold histogram, which is used to separate the pixels of an input image into several different classes (4 levels: high, medium, low, and background clusters was used for this method), each one obtained according to the intensity of the gray

levels within the image. The resulting subvolumes were postprocessed using morphological operators based on opening to remove small and isolated clusters in the map using a spherical structure element. The subunit dose response was described phenomenologically by a sigmoid dose response function.

$$p(d) = \frac{1}{\left[1 + \left(\frac{D_{50}}{d}\right)^{k}\right]}, \quad (5)$$

where d is dose one subunit received, D_{50} is the dose at which 50% of the subunits are damaged, and a slope parameter, k, determines rate at which the probability of damage subunits increases with dose.

The overall functional damage then can be estimated:

$$f = \sum_i v_i p(d_i), \quad (6)$$

where v_i is the fractional volume receiving dose d_i. The subunit dose-response model was fit using 24 patients' imaging data, and the NTCP model was fit using all 146 patients' data. Model parameters were fit by maximum likelihood estimation. The 95% error bars of binned data were calculated using Agresti's approximation.[43] Model confidence intervals were calculated using the approximate profile likelihood. p-values were estimated using χ^2statistics. The NTCP models were applied to both real patient data and the synthetic post-RT k1 data generated by cGAN and cWGAN. Assuming patients with different pre-RT liver conditions (based on ALBI scores) dose-responses vary, high and low pre-RT ALBI subgroups were formed based on the median ALBI scores, and NCTP models were built respectively on each subgroup. We used median ALBI value to divide the patients into good ($N = 12$) and poor ($N = 12$) liver function subgroups. We provided a flowchart to describe the workflow of for model training and validation, as well as the inputs and outputs for the GAN and NTCP (Figure 1).

3 | RESULTS

3.1 | Predicted post-RT functional map analysis

The prediction performance of post-RT k1 mapping using a deep learning method has been analyzed and compared. Cross validated testing showed that cWGAN predicted post-RT liver function better than the traditional cGAN, and the prediction is significantly more accurate for better liver baseline global condition patients. cWGAN yielded a mean Structural Similarity Index (SSIM) of 0.53 (CI: 0.46–0.59) and mean root-mean square error (RMSE) of 13.32 (CI: 8.51–18.12) comparing with cGAN SSIM 0.37 (CI: 0.31–0.42), RMSE 23.76 (CI: 13.30–34.23). A Wilcoxon test showed that the cWGAN performed significantly better in all three groups: all ($p < 0.001$), high ($p = 0.045$), and low baseline ALBI ($p = 0.001$) groups, as shown in Figure 2. cWGAN prediction yielded a gamma passing rate of 0.867 (CI: 0.824–0.910). Comparing poor/good liver baseline function subgroups showed that cWGAN images could achieve 0.61 (CI: 0.53–0.68) mean SSIM for the high liver function group, while the SSIM is 0.45 (CI: 0.35–0.55) for the low liver function group (Table S1). cWGAN achieved mean gamma passing rate of 0.884 (CI: 0.824–0.944) for high liver function group compared to 0.850 (CI: 0.781–0.919) for low liver function group (Figure S2). Examples of predicted post-RT k1 maps (cGAN and cWGAN) along with the corresponding ground truth, pre-RT k1 maps, error maps, and doses (EQD2) are shown in Figure 3. Error map is calculated by subtracting ground truth post-RT map from the predicted post-RT map.

3.2 | NTCP analysis

NTCP parameters were fitted and compared for different baseline liver function subgroups as well as using original and model-predicted post-RT k1 maps. The logistic models yielded a D50 of 35.2 (CI: 26.7–47.5) Gy, 11.7 (CI: 9.1–15.4) Gy, and 54.8 (CI: 38.3–79.1) Gy and k of 0.62 (CI: 0.49–0.75), 0.96 (CI: 0.74–1.22), and 0.59 (CI: 0.48–0.74) based on original images for all, high and low ALBI subgroups, respectively (Figure 4a-c), where D50 is the dose at which 50% of the volume are damaged and k determines rate at which the probability of volume damage increases with dose. The poor liver function group showed a significantly smaller D50 ($p < 0.001$) and larger k values ($p = 0.008$) compared to the good liver function subgroup, indicating higher sensitivity to lower dose and steeper dose-response curves as shown in Figures 4b,c and 5a,b. The NTCP models yielded a $f50$ of 0.53 (CI: 0.50–0.56), 0.73 (CI: 0.70–0.77), and 0.43 (CI: 0.41–0.46) and γ of 0.84 (CI: 0.76–0.90), 0.90 (CI: 0.82–0.96), and 0.85 (CI: 0.77–0.91) based on original images for all, high and low ALBI subgroups, respectively (Figure 4g-i), where $f50$ is the fraction of the total organ damaged which would produce a 50% complication rate, γ the steepness of the organ response. A fractional loss of 0.73 would cause a 50% complication rate for poor liver function patients, which is substantially larger than that of 0.43 for good liver function patients as shown in Figure 4h,i. The dose response and NTCP models were also fitted using cWGAN predicted images for all and the sub-populations of good/poor liver function patients (Figure 4d-f,j-l). The fit parameters are summarized in Tables S2 and S3.

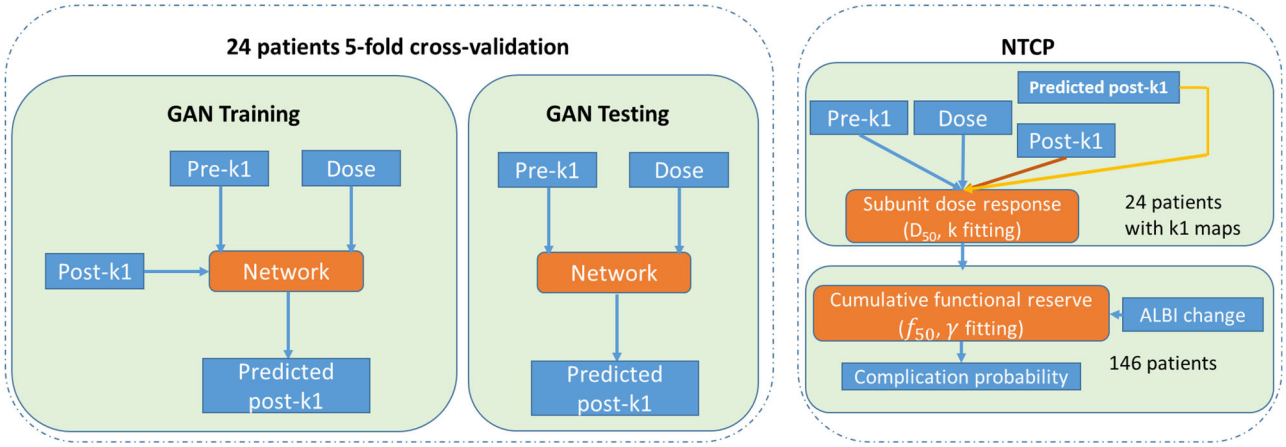


FIGURE 1 Workflow for the GAN and NTCP models. The NTCP was fit twice to compare the four parameters D50, k, f_50, γ. Orange line in the right figure (NTCP) means using the post-RT k1 image, pre-RT k1 and dose to fit the NTCP, which is the ground truth fitting. The yellow line means using predicted post-RT k1 map for the fitting instead of the original post-RT k1.

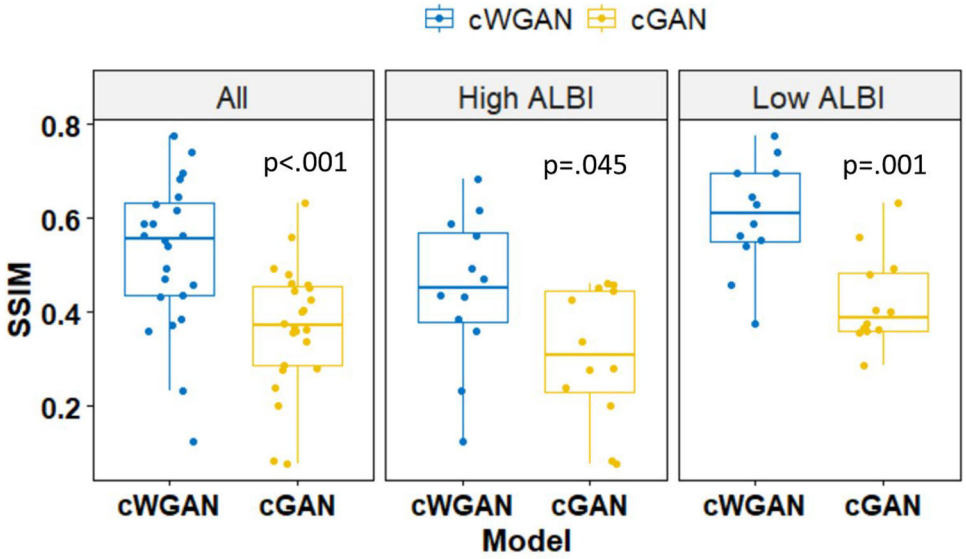


FIGURE 2 Box plot for all, high and low ALBI patients with Wilcoxon signed-rank test between predicted images from cGAN and cWGAN based on SSIM metric.

4 | DISCUSSION

In this work, we built NTCP models using the Gadoxetic Acid uptake rate (k1 map) derived from DGAE-MRI to represent local hepatic function for HCC patients treated with SBRT. D50 was found to be significantly smaller and *k* significantly larger for the poor liver function subgroup compared with the good liver function subgroup, which indicates higher radiosensitivity for the poor baseline liver function cohort. We used a novel conditional Wasserstein adversarial neural network based on pre-RT liver function maps and dose distributions to try to predict post-RT liver function maps. The proposed model performed better for good liver function patients than for poor. The NTCP models obtained from original and predicted images for all and the good baseline liver function patient groups yielded similar results though this was not observed to be the case for the low baseline function group. This shows the potential usage of this developed cWGAN model to optimize the treatment for individual patients in advance and provide dose limit guidance to good liver function patients,

Effective prediction of post-RT functional maps could provide localized response information which could be used to maximize dose to the tumor while minimizing toxicity on a patient-specific basis. This allows one to analyze the radiosensitivity for each patient and optimize the treatment plan based on the predicted response. A Wilcoxon test showed that the fitted parameters are all statistically not different between

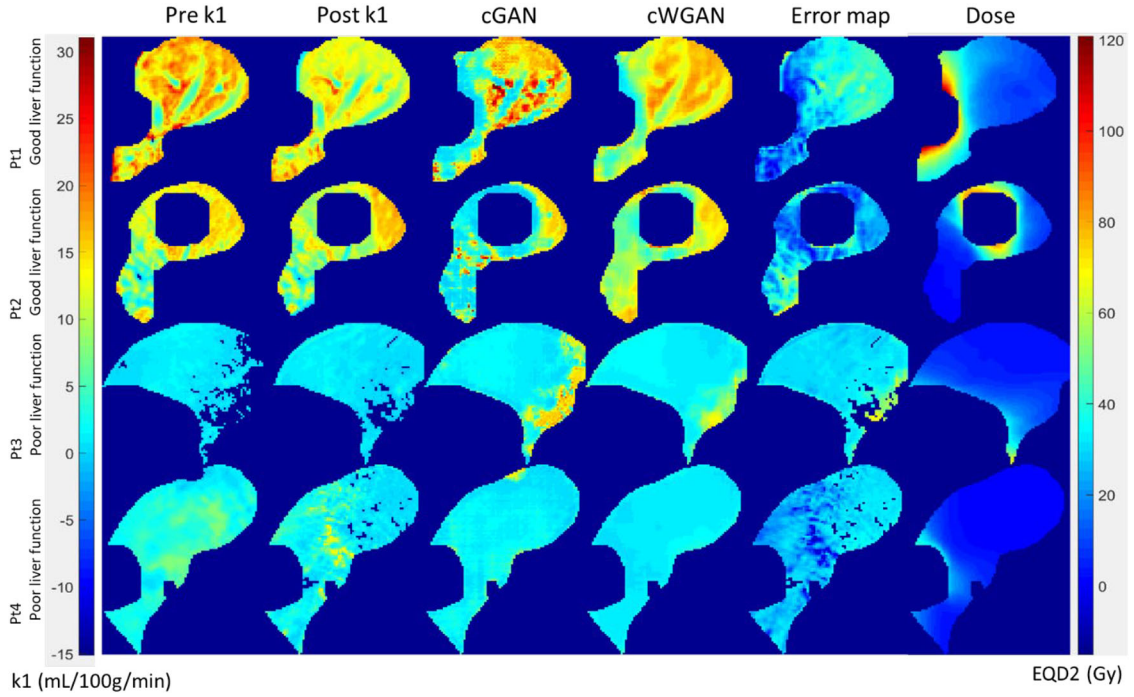


FIGURE 3 Examples of good (top two rows) and poor (bottom two rows) baseline liver function patients. From left to right: pre-RT k1 maps, ground truth post-RT k1 maps, cGAN, cWGAN predicted post-RT k1 maps, error maps for cWGAN and doses are shown for each patient. Unit for k1 maps is mL/100 g/min and for dose (EQD2) is Gy. Top row patient CP: 5, ALBI: −2.6; second row CP: 5, ALBI: −2.53; third row CP: 8, ALBI: −0.96; fourth row CP: 8, ALBI: −1.29. The left colorbar is for the k1 maps (1–5th column) and the right colorbar is just for the dose (6th column).

original image and cWGAN predicted image based NTCP models for the good liver function subgroup (Figure 6). In addition, we found this dose-response model yielded significantly smaller D50 (11.7 Gy) and larger *k* (0.959) for the poor liver function compared with D50 (54.8 Gy) and *k* (0.59) for good liver function subgroup. This indicates that poor liver function patients are more sensitive to low doses and show steeper functional loss with dose. Similar low dose sensitivity has been reported by another study by Price *et al*., where patients with low baseline liver function exhibited combined extrema of low dose threshold and steep dose-response slope.[44] Associating global liver function (pre-RT ALBI/CP) with regional/functional dose response and radiation-induced normal tissue complication can be informative for assigning patients to different response groups. These findings are applicable in treatment planning practice. We could simply stratify patients into high/low pre-RT ALBI groups and optimize treatment plans with the dose threshold for each group as well as sparing high functional regions. Specifically, we can use different dose constraints for different subgroups of patients for the optimization, for example, if a patient has poor liver function, we can use the fitted parameters for poor function subgroup and calculate the NTCP and set constraint based on the results. Another, more personalized way would be to take full advantage of an individual's response map and obtain the dose threshold for each patient to optimize the plan. This will require future studies to optimize and validate dose distribution using pre-RT and predicted post-RT functional maps to minimize the functional loss and complication probability.

A limitation of this study is the lack of satisfactory predictions for the poor function group. The pre-RT k1 map has lower intensities for the poor function group, and the prediction was not accurate and does not provide much useful information. In comparison, the cWGAN model prediction is significantly better for good function subgroup, as shown in Figures 3 and 6. The possible contributing factors for the worse prediction results include: 1. Lack of data makes it difficult to learn the patterns in the complex cases for patients with poor liver function; 2. There are probably confounding factors other than dose causing the regional function loss for these patients. The response patterns are different for different global liver function (quantified by ALBI score) patients. A recent study has shown that inflammation level is also associated with toxicity and overall survival of the patients after liver irradiation.[18] We expect large improvements in the performance of our proposed model once we have more data for different subgroups, including comprehensive genetics and proteomics information for each individual patient. Another limitation of our approach is the use of rigid registration for pre- and post-RT k1 maps, which assumes that the liver is

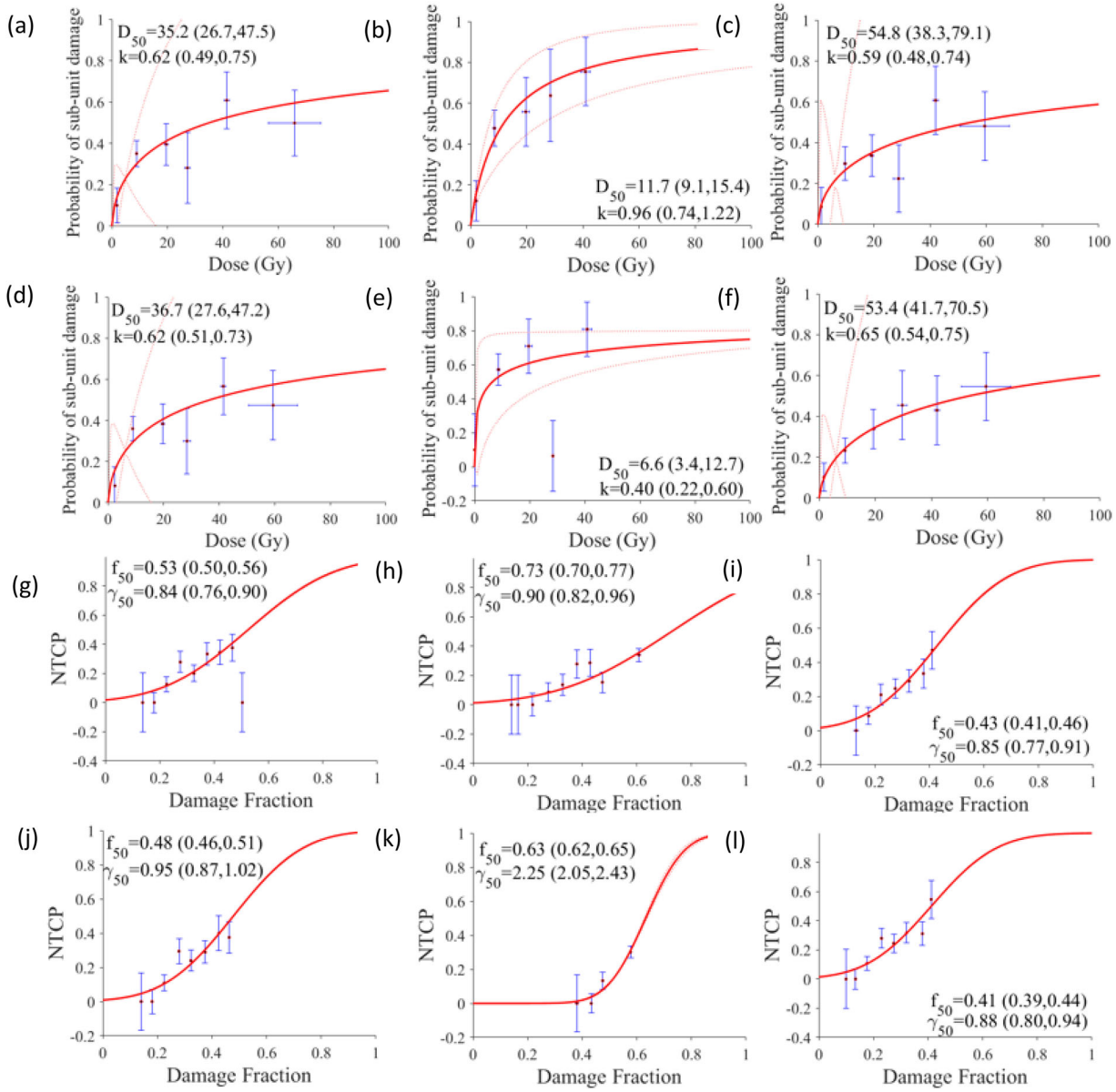


FIGURE 4 Dose response and NTCP fraction damage model results. Dose response results using original images for all (a), high (b), and low (c) ALBI patients and using cWGAN predicted images for all (d), high (e), and low (f) ALBI patients; NTCP fit results using original images for all (g), high (h), and low (i) ALBI patients and using cWGAN predicted images for all (j), high (k), and low (l) ALBI patients. The blue error bars represent 95% confidence intervals of binned data. The solid red lines represent the best fit, with the dotted lines representing 95% confidence intervals. The 95% confidence intervals are much smaller for plots (g)–(l), so the dotted lines overlap with the solid curves in (g)–(l).

a rigid object. However, the liver is a deformable organ and residual misalignment between pre- and post-RT scans after rigid registration could introduce variation in the maps that is not solely due to the RT response. To address this issue, we employed a binning method for the NTCP modeling, which aggregates data for many voxels per data point instead of voxel-wise, thereby reducing the impact of registration uncertainty. Nonetheless, we acknowledge that the uncertainty introduced by the registration method may still affect the deep learning model we trained. We plan to investigate this further in future studies.

RMSE and SSIM are the metrics we used to evaluate the image prediction performance. Although we showed that the proposed cWGAN method outperformed the original cGAN method, it is hard to determine a target performance in an absolute sense. The SSIM index measures the similarities of three elements of two images: the similarity of luminances (brightness values), contrasts, and structures. These similarities are

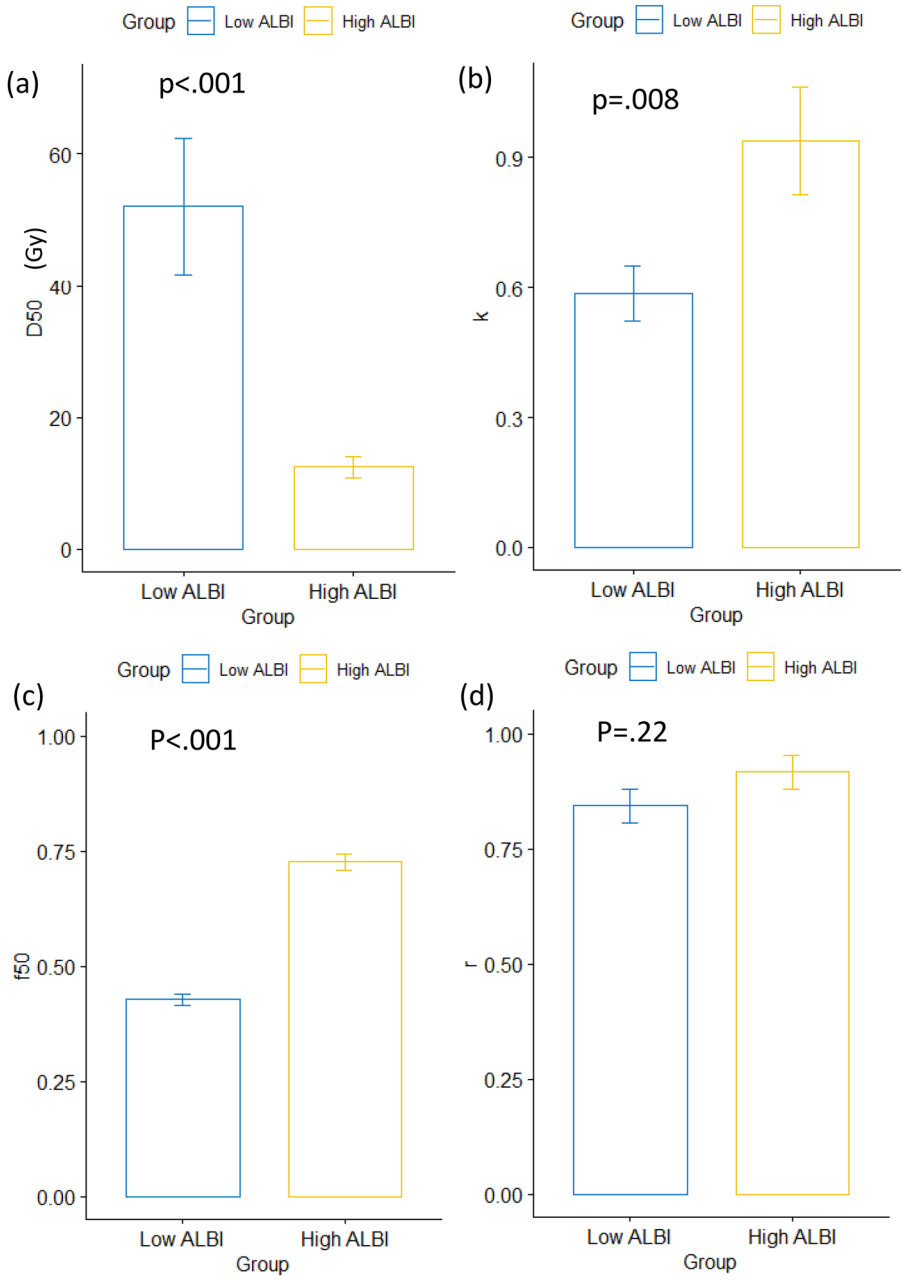


FIGURE 5 Wilcoxon test between low and high ALBI groups model fitted parameters. Dose response fit parameters D50 (a) and k (b); NTCP fraction damage fit parameters f50 (c) and r (d).

expressed using simple, easily computed statistics, and combined to form SSIM.[45] The maximum achievable value is 1., if and only if two images are identical. SSIM has been used for evaluating image processing results in different medical applications, such as image synthesis,[46] artifact suppression,[47] predicting post-contrast MR images from pre-contrast images,[48] predicting PET images from MRIs,[49] image generation using GANs.[50,51] The range of SSIM for these applications are wide (~0.2 to ~0.99). It mainly depends on how challenging the specific task is. For tasks like image reconstruction,[52,53] denoising,[54] artifact reduction,[55] in general the achievable SSIM is high (~0.9). However, the achievable SSIM is low for certain tasks due to the inherent difficulty of problem. In a work that was developed to simulate T1-weighted postcontrast images from eight precontrast sequences in 400 patient, the SSIMs for whole brain, whole tumor, and tumor core are about 0.83, 0.58, and 0.45, respectively,[48] We can see the SSIM is not satisfactory for challenging tumor prediction. One study did similar work as us to predict the voxelwise relative cerebral blood flow (CBF) change map with before acetazolamide administration PET (baseline PET), baseline arterial spin labeling (ASL) perfusion

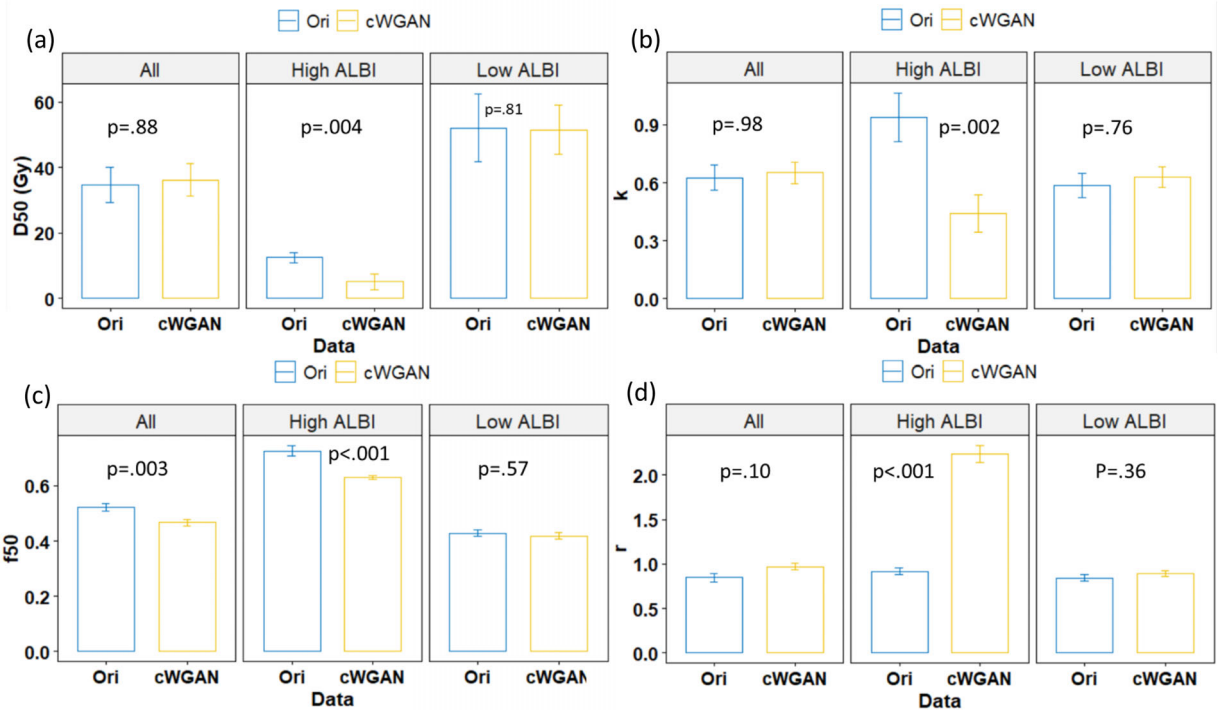


FIGURE 6 Wilcoxon test between original image and cWGAN predicted image-based dose response model parameters D50 (a), k (b) and fraction damage model parameters f50 (c), r (d) for all, high and low ALBI patients.

images (5 images) and 6 structural MRI images, in total 12 input images for 36 patients[49] with best model SSIM ~ 0.58. They did not interpret the absolute values but discuss how various models' results are different from each other based on SSIM. The SSIM for good baseline liver function subgroup in our study is 0.61, given only 1 pre-RT MRI image for 24 patients is available. This is probably the best we can get with the current dataset. To be noted, the SSIM is not a direct indicator of the clinical relevance of the developed models. To show the clinical relevance, we used the predicted images to fit the NTCP models and compared the fitted parameters with the ground truth values (Figure 6). The rationale is that currently we use the NTCP models for making treatment plans for individual patients. As long as the NTCP models are accurate, the prediction models we developed should be able to inform clinical decision-making process. However, with more data accumulated and potentially better techniques developed, the prediction performance should be improved in the future.

Although we came up with novel methods to minimize the limitation caused by lack of data, these interesting findings warrant large and external validation. We did not make complex or individualized models for NTCP modeling but continued using the parallel architecture model. With a large number of voxels available, for the single output outcome modeling, more complex models are prone to overfit with 24 data points. Hence, a future work will be to build personalized normal tissue complication models instead of the current population-wise method.

5 | CONCLUSIONS

Voxel-wise prediction of the post-RT liver functional (k1) map was achieved using cWGAN based on the pre-RT k1 map derived from DGAE MRI and dose distribution. NTCP models which incorporate voxel-wise functional information from DGAE-MRI k-maps were successfully developed and feasibility was demonstrated in a small patient cohort. cWGAN predicted functional maps show promise for estimating localized patient-specific response to RT and warrant further validation in a larger patient cohort.

ACKNOWLEDGMENTS

This research is supported by the National Institutes of Health under Award Numbers P01-CA059827 and R01-CA233487.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

REFERENCES

1. Grimm J, Vargo JA, Mavroidis P, et al. Initial Data pooling for radiation dose-volume tolerance for carotid artery blowout and other

bleeding events in hypofractionated head and neck retreatments. *Internat J Radiat Oncol Biol Phys*. 2021;110:147-159.
2. Mahadevan A, Moningi S, Grimm J, Li XA, Forster KM, Palta M, et al. Maximizing tumor control and limiting complications with stereotactic body radiation therapy for pancreatic cancer. *Internat J Radiat Oncol Biol Phys*. 2021;110:206-216.
3. Miften M, Vinogradskiy Y, Moiseenko V, et al. Radiation dose-volume effects for liver SBRT. *Internat J Radiat Oncol Biol Phys*. 2021;110:196-205.
4. Ohri N, Tomé WA, Romero AM, Miften M, Ten Haken RK, Dawson LA, et al. Local control after stereotactic body radiation therapy for liver tumors. *Internat J Radiat Oncol Biol Phys*. 2021;110:188-195.
5. Schefter TE, Kavanagh BD, Timmerman RD, Cardenes HR, Baron A, LE Gaspar. A phase I trial of stereotactic body radiation therapy (SBRT) for liver metastases. *Internat J Radiat Oncol Biol Phys*. 2005;62:1371-1378.
6. Ibarra RA, Rojas D, Snyder L, Yao M, Fabien J, Milano M, et al. Multicenter results of stereotactic body radiotherapy (SBRT) for non-resectable primary liver tumors. *Acta Oncol (Madr)*. 2012;51:575-583.
7. Jackson WC, Tang M, Maurino C, Mendiratta-Lala M, Parikh ND, Matuszak MM, et al. Individualized adaptive radiation therapy allows for safe treatment of hepatocellular carcinoma in patients with child-turcotte-pugh B liver disease. *Internat J Radiat Oncol Biol Phys*. 2021;109:212-219.
8. Velec M, Haddad CR, Craig T, Wang L, Lindsay P, Brierley J, et al. Predictors of liver toxicity following stereotactic body radiation therapy for hepatocellular carcinoma. *Internat J Radiat Oncol Biol Phys*. 2017;97:939-946.
9. Lyman JT. Complication probability as assessed from dose-volume histograms. *Radiat Res Suppl*. 1985;104:S13-S9.
10. Pan CC, Kavanagh BD, Dawson LA, et al. Radiation-associated liver injury. *Internat J Radiat Oncol Biol Phys*. 2010;76:S94-S100.
11. Miften M, Vinogradskiy Y, Moiseenko V, et al. Radiation dose-volume effects for liver SBRT. *Internat J Radiat Oncol Biol Phys*. 2021;110:196-205.
12. Jackson A, Ten Haken RK, Robertson JM, Kessler ML, Kutcher GJ, Lawrence TS. Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Internat J Radiat Oncol Biol Phys*. 1995;31:883-891.
13. Miften MM, Das SK, Su M, Marks LB. Incorporation of functional imaging data in the evaluation of dose distributions using the generalized concept of equivalent uniform dose. *Phys Med Biol*. 2004;49:1711-1721.
14. El Naqa I, Johansson A, Owen D, et al. Modeling of normal tissue complications using imaging and biomarkers after radiation therapy for hepatocellular carcinoma. *Internat J Radiat Oncol Biol Phys*. 2018;100:335-343.
15. Price RG, Apisarnthanarax S, Schaub SK, et al. Regional radiation dose-response modeling of functional liver in hepatocellular carcinoma patients with longitudinal sulfur colloid SPECT/CT: a proof of concept. *Internat J Radiat Oncol Biol Phys*. 2018;102:1349-1356.
16. Stenmark MH, Cao Y, Wang H, et al. Estimating functional liver reserve following hepatic irradiation: adaptive normal tissue response models. *Radiother Oncol*. 2014;111:418-423.
17. Simeth J, Aryal M, Owen D, Cuneo K, Lawrence TS, Cao Y. Gadoxetic acid uptake rate as a measure of global and regional liver function as compared with indocyanine green retention, albumin-bilirubin score, and portal venous perfusion. *Adv Radiat Oncol*. 2022;7:100942.
18. Haimerl M, Verloh N, Zeman F, et al. Gd-EOB-DTPA-enhanced MRI for evaluation of liver function: comparison between signal-intensity-based indices and T1 relaxometry. *Sci Rep*. 2017;7:1-12.
19. Bártulos CR, Senk K, Schumacher M, et al. Assessment of liver function with MRI: where do we stand? *Front Med*. 2022;9:729.
20. Haimerl M, Probst U, Poelsterl S, et al. Hepatobiliary MRI: signal intensity based assessment of liver function correlated to 13C-Methacetin breath test. *Sci Rep*. 2018;8:1-8.
21. Wei L, Simeth J, Aryal MP, et al. The effect of stereotactic body radiation therapy for hepatocellular cancer on regional hepatic liver function. *International Internat J Radiat Oncol Biol Phys*. 2023;115(3):794-802.
22. Lewis H, Ghasabeh M, Khoshpouri P, Kamel I, Pawlik T. Functional hepatic imaging as a biomarker of primary and secondary tumor response to loco-regional therapies. *Surg Oncol*. 2017;26:411-422.
23. El Naqa I, Johansson A, Owen D, et al. Modeling of normal tissue complications using imaging and biomarkers after radiation therapy for hepatocellular carcinoma. *Internat J Radiat Oncol Biol Phys*. 2018;100:335-343.
24. Barendsen G. Dose fractionation, dose rate and iso-effect relationships for normal tissue responses. *Internat J Radiat Oncol Biol Phys*. 1982;8:1981-1997.
25. Cheng JC-H, Wu J-K, Lee PC-T, et al. Biologic susceptibility of hepatocellular carcinoma patients treated with radiotherapy to radiation-induced liver disease. *Internat J Radiat Oncol Biol Phys*. 2004;60:1502-1509.
26. Dawson LA, Biersack M, Lockwood G, Eisbruch A, Lawrence TS, Ten Haken RK. Use of principal component analysis to evaluate the partial organ tolerance of normal tissues to radiation. *Internat J Radiat Oncol Biol Phys*. 2005;62:829-837.
27. Kim TH, Kim DY, Park J-W, et al. Dose–volumetric parameters predicting radiation-induced hepatic toxicity in unresectable hepatocellular carcinoma patients treated with three-dimensional conformal radiotherapy. *Internat J Radiat Oncol Biol Phys*. 2007;67:225-231.
28. Johansson A, Balter J, Feng M, Cao YJT. An overdetermined system of transform equations in support of robust DCE-MRI registration with outlier rejection. *Tomography*. 2016;2:188-196.
29. Simeth J, Johansson A, Owen D, et al. Quantification of liver function by linearization of a two-compartment model of gadoxetic acid uptake using dynamic contrast-enhanced magnetic resonance imaging. *NMR Biomed*. 2018;31:e3913.
30. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9:62-66.
31. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA, Context encoders: feature learning by inpainting. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. 2536-2544.
32. Isola P, Zhu J-Y, Zhou T, Efros AA, Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. 1125-1134.
33. Weng L, From gan to wgan. arXiv preprint arXiv:190408994. 2019.
34. Kodali N, Abernethy J, Hays J, Kira Z, On convergence and stability of gans. arXiv preprint. arXiv:170507215. 2017.
35. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. International conference on machine learning: PMLR. 2017:214-223.
36. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A, Improved training of wasserstein gans. arXiv preprint. arXiv:170400028. *Adv Neural Inf Process Syst*. 2017:30.
37. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach—the ALBI grade. *J Clin Oncol*. 2015;33:550.
38. Pinato DJ, Sharma R, Allara E, Yen C, Arizumi T, Kubota K, et al. The ALBI grade provides objective hepatic reserve estimation across each BCLC stage of hepatocellular carcinoma. *J Hepatol*. 2017;66:338-346.
39. Yamada A. *Quantitative evaluation of liver function within MR imaging. Abdomen and thoracic imaging*. Springer; 2014:233-251.

40. Sourbron S, Sommer WH, Reiser MF, Zech CJ. Combined quantification of liver perfusion and function with dynamic gadoxetic acid–enhanced MR imaging. *Radiology*. 2012;263:874-883.
41. Verloh N, Haimerl M, Zeman F, et al. Assessing liver function by liver enhancement during the hepatobiliary phase with Gd-EOB-DTPA-enhanced MRI at 3 Tesla. *Eur Radiol*. 2014;24:1013-1119.
42. Nilsson H, Blomqvist L, Douglas L, Nordell A, Jonas E. Assessment of liver function in primary biliary cirrhosis using Gd-EOB-DTPA-enhanced liver MRI. *HPB*. 2010;12:567-576.
43. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat*. 1998;52:119-126.
44. Price RG, Apisarnthanarax S, Schaub SK, et al. Regional radiation dose-response modeling of functional liver in hepatocellular carcinoma patients with longitudinal sulfur colloid SPECT/CT: a proof of concept. *Internat J Radiat Oncol Biol Phys*. 2018;102:1349-1356.
45. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.
46. Gourdeau D, Duchesne S, Archambault L. On the proper use of structural similarity for the robust evaluation of medical image synthesis models. *Med Phys*. 2022;49:2462-2474.
47. Jaubert O, Steeden J, Montalt-Tordera J, Arridge S, Kowalik GT, Muthurangu V. Deep artifact suppression for spiral real-time phase contrast cardiac magnetic resonance imaging in congenital heart disease. *Magn Reson Imaging*. 2021;83:125-132.
48. Calabrese E, Rudie JD, Rauschecker AM, Villanueva-Meyer JE, Cha S. Feasibility of simulated postcontrast MRI of glioblastomas and lower-grade gliomas by using three-dimensional fully convolutional neural networks. *Radiol Artif Intell*. 2021;3:e200276.
49. Chen DY, Ishii Y, Fan AP, et al. Predicting PET cerebrovascular reserve with deep learning by using baseline MRI: a pilot investigation of a drug-free brain stress test. *Radiology*. 2020;296:627.
50. Ali H, Biswas R, Ali F, Shah U, Alamgir A, Mousa O, et al. The role of generative adversarial networks in brain MRI: a scoping review. *Insights Imaging*. 2022;13:1-15.
51. Rakhimberdina Z, Jodelet Q, Liu X, Murata T. Natural image reconstruction from fMRI using deep learning: a survey. *Front Neurosci*. 2021;15:795488.
52. Zhou Z, Han F, Ghodrati V, et al. Parallel imaging and convolutional neural network combined fast MR image reconstruction: applications in low-latency accelerated real-time imaging. *Med Phys*. 2019;46:3399-3413.
53. Zhou B, Zhou SK, Duncan JS, Liu C, Limited view tomographic reconstruction using a deep recurrent framework with residual dense spatial-channel attention network and sinogram consistency. arXiv preprint arXiv:200901782. 2020.
54. Li Y, Zhang K, Shi W, Miao Y, Jiang Z. A novel medical image denoising method based on conditional generative adversarial network. *Computat Math Methods Med*. 2021;1-11.
55. Tamada D, Kromrey M-L, Onishi H, Motosugi U, Method for motion artifact reduction using a convolutional neural network for dynamic contrast enhanced MRI of the liver. arXiv preprint arXiv:180706956. 2018.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wei L, Aryal MP, Cuneo K, et al. Deep learning prediction of post-SBRT liver function changes and NTCP modeling in hepatocellular carcinoma based on DGAE-MRI. *Med Phys*. 2023;50:5597–5608. https://doi.org/10.1002/mp.16386