Published in final edited form as:

Med Phys. 2021 September ; 48(9): 4784–4798. doi:10.1002/mp.15075.

Deformation driven Seq2Seq longitudinal tumor and organs-at-risk prediction for radiotherapy

Donghoon Lee, Sadegh R. Alam, Jue Jiang, Pengpeng Zhang, Saad Nadeem*, Yu-chi Hu*

Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Abstract

Purpose: Radiotherapy presents unique challenges and clinical requirements for longitudinal tumor and organ-at-risk (OAR) prediction during treatment. The challenges include tumor inflammation/edema and radiation-induced changes in organ geometry, whereas the clinical requirements demand flexibility in input/output sequence timepoints to update the predictions on rolling basis and the grounding of all predictions in relationship to the pre-treatment imaging information for response and toxicity assessment in adaptive radiotherapy.

Methods: To deal with the aforementioned challenges and to comply with the clinical requirements, we present a novel 3D sequence-to-sequence model based on Convolution Long Short-Term Memory (ConvLSTM) that makes use of series of deformation vector fields (DVF) between individual timepoints and reference pre-treatment/planning CTs to predict future anatomical deformations and changes in gross tumor volume as well as critical OARs. High-quality DVF training data are created by employing hyper-parameter optimization on the subset of the training data with DICE coefficient and mutual information metric. We validated our model on two radiotherapy datasets: a publicly available head-and-neck dataset (28 patients with manually contoured pre-, mid-, and post-treatment CTs), and an internal non-small cell lung cancer dataset (63 patients with manually contoured planning CT and 6 weekly CBCTs).

Results: The use of DVF representation and skip connections overcomes the blurring issue of ConvLSTM prediction with the traditional image representation. The mean and standard deviation of DICE for predictions of lung GTV at weeks 4, 5, and 6 were 0.83 ± 0.09 , 0.82 ± 0.08 , and 0.81 ± 0.10 , respectively, and for post-treatment ipsilateral and contralateral parotids, were 0.81 ± 0.06 and 0.85 ± 0.02 .

Conclusion: We presented a novel DVF-based Seq2Seq model for medical images, leveraging the complete 3D imaging information of a relatively large longitudinal clinical dataset, to carry out longitudinal GTV/OAR predictions for anatomical changes in HN and lung radiotherapy patients, which has potential to improve RT outcomes.

Correspondence: Yu-chi Hu and Saad Nadeem, Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 485 Lexington Ave., New York, NY 10017, USA. huj@mskcc.org, (Y. H.); nadeems@mskcc.org, (S. N.).

*Co-senior authors.

CONFLICT OF INTEREST

The authors have no conflict to disclose.

Keywords

ConvLSTM; deformation vector field; longitudinal prediction

1 | INTRODUCTION AND PURPOSE

The recent development of imaging techniques in modern radiation therapy (RT) has enabled an improved understanding of the geometric variation of patient anatomy during the treatment course.¹ The innovative methods to manage this longitudinal variation fall under the umbrella of adaptive radiotherapy (ART).^{1–3} ART is a conceptually attractive approach to correct for daily tumor and normal tissue variation. For example, ART for head and neck cancer is an emerging tool to counter morphologic changes in patient and tumor anatomy during a course of RT by creating new radiation plans mid-treatment.^{4–7} The most common anatomical variation is excessive weight loss and radiation-induced side effects.^{7–9} These longitudinal anatomical changes have been shown to reduce doses to target volumes while increasing the dose to the organs at risk (OARs).^{10,11}

According to retrospective studies, the setup variation can be characterized early on during the treatment process.^{5,12} Therefore, ART may potentially benefit from predictions of patient's anatomical deformation in the later timepoints since these predictions could directly be used in the replanning process and improve the therapeutic outcome.^{10,13,14} In this paper, we present a novel 3D deep learning sequence-to-sequence model (Seq2Seq) using ConvLSTM to predict patient anatomy deformation (with reference to the planning/pre-treatment image) given any number of input/output sequence timepoints.

The Seq2Seq model is developed to address the history-dependent response prediction problem¹⁵ and has been used in various applications such as natural language process, image captioning, weather forecasting, and video frame prediction.^{15–20} In the Seq2Seq model, long short-term memory (LSTM) or gated recurrent unit (GRU), which are representative recurrent neural network (RNN) models, have been used.^{21,22} LSTM and GRU are designed for the next time-step status prediction in a temporal sequence and can be naturally extended to predict the later frames from previous ones in a sequential image dataset.^{23,24} Video or image prediction by LSTM or GRU has been combined with fully convolutional networks (FCNs) in order to learn spatial and temporal filters. However, 3D feature maps of shape $[C, H, W]$ are flattened into 1D vectors of size $C \times W \times H$, where C represents the channels, H is the height, and W is the width of the feature maps. A disadvantage of this approach is that the data that flow through the LSTM is 1D, and as such we may lose spatial information. One approach that mitigates this issue and is suitable for dealing with the sequence of images is using ConvLSTM.^{23,24} It is a recurrent model, just like the LSTM, but internal matrix multiplications are exchanged with convolution operations. As a result, the data that flow through the ConvLSTM cells preserve its input dimensionality instead of being flattened to 1D feature vectors. Because of the spatio-temporal advantage of ConvLSTM, it has been applied to medical image sequential datasets. Zhang et al.²³ studied tumor growth prediction using ConvLSTM with cropped 32×32 patches of tumor CT, tumor contour, and intercellular volume fraction images. Even though the results achieved 0.80 dice overlap,

their model suffered from the known blurring and limited long-term dependency—issues from ConvLSTM.^{25,26} Moreover, the study only focused on predicting a single timepoint from two earlier timepoints. GAN-based approaches such as BeyondMSE²⁷ have also been proposed for video frame prediction but as was shown in,²³ it achieved lower prediction performance than ConvLSTM given the lack of explicit temporal dynamics modeling in BeyondMSE. Zhang et al.²³ also noted that the GAN-based prediction methods have a higher risk of overfitting and the network architectures can be over-complicated for the relatively small-sized medical datasets.

In this paper, we make the following contributions:

1. We present a novel 3D deep learning sequence-to-sequence model using ConvLSTM for longitudinal tumor and OAR prediction during radiotherapy given any number of input/output sequence timepoints (e.g., predicting weeks 4–6 given weeks 1–3 input sequence—or predicting weeks 3–6 given weeks 1–2 input sequence) and using the whole imaging field-of-view (rather than cropped patches).
2. Deformation vector field (DVF) representation is used to overcome the ConvLSTM blurring issue with the image representation and to keep the predictions grounded with respect to the reference planning/pre-treatment images (radiotherapy clinical requirement). Skip connections are used to resolve the limited long-term dependency ConvLSTM issue between sequence elements.
3. A thorough validation is done on two radiotherapy datasets: (1) publicly available head-and-neck (HN) 28 patient dataset with three pre-, mid-, and post-treatment CT timepoints, and (2) internal non-small cell lung cancer (NSCLC) 63 patient dataset with planning CT and 6 weekly cone-beam CT (CBCT) images. All timepoints in both datasets were contoured for tumor and OAR by expert radiation oncologists, providing data for evaluation and for creating high-quality training DVFs via hyper-parameter optimization. For OAR, we focus on longitudinal prediction for esophagus in lung (due to the debilitating radiation-induced esophagitis which develops in 50% of the NSCLC radiotherapy patients) and for parotid glands in HN (due to the prevalent radiation-induced xerostomia). The use of 6 weeks' sequence 3D CT images makes our work the first to study such a long sequence in the longitudinal prediction context with a relatively large dataset.

In the following sections, we will describe the two HN and lung datasets, the DVF creation process via deformable image registration (DIR) hyper-parameter optimization, the Seq2Seq deep learning prediction model and finally conclude with results and discussion.

2. | MATERIALS AND METHODS

2.1 | Longitudinal dataset

We used two longitudinal radiotherapy datasets: a publicly available head-and-neck squamous cell carcinoma (HNSCC) and an internal non-small cell lung cancer datasets. The HNSCC dataset, which is available via The cancer image archive (TCIA), contains three-

dimensional high-resolution 3D fan-beam CT scans collected pre-, mid-, and post-treatment for 28 patients who underwent RT treatment (to a total dose of 58–70 Gy using daily 2–2.20 Gy fraction for 30–35 fractions)²⁸; there were total 31 patients but 3 were missing parotid gland contours. The dataset also contains radiation oncologist drawn contours of anatomical structures and dose maps. The pre-treatment CT scans were acquired a median of 13 days before treatment, the mid-treatment acquired at fraction 17, and the post-treatment at fraction 30. The lung dataset is an internal dataset and contains three-dimensional CT and weekly CBCT scans collected during RT of 63 patients with NSCLC. Patients were treated via intensity-modulated RT in 2–3 Gy daily fractions in a 5 days/week fraction (total dose 50–73 Gy).²⁹ High-resolution planning CT (pCT) was used to define the gross tumor volume (GTV) and planning the treatment. All patients had weekly CBCT to monitor positional uncertainties and tumor changes. In addition to imaging data, we have the contours of GTV and esophagus at planning CT and each weekly CBCT. The contours were generated by a radiation oncologist. We split each dataset into training/validation/testing with 21/2/5 for HN and 50/3/10 for lung.

2.2 | Deformable image registration and hyper-parameter optimization

The purpose of our study is to predict future sequence of DVF from earlier DVF sequence so we need accurate DVF data from deformable image registration (DIR) to train our model. First, we roughly aligned the global structures by matching the center of the tumor and rigidly registered mid- and post-treatment CTs to pre-treatment CT for HNSCC dataset and weekly CBCTs to pCT for the lung dataset. Then we resampled the datasets with an isotropic voxel size, $1 \times 1 \times 1 \text{ mm}^3$ for HNSCC and $2 \times 2 \times 2 \text{ mm}^3$ for lung dataset, to have the same physical deformation intensity in the three-axis direction. The HNSCC volumes were then cropped to include only the HN information, and the lung pCT volumes were cropped to match the CBCT field of view which is inherently different. Using these final images and the corresponding OAR/GTV contours, optimal hyper-parameters for B-spline regularized diffeomorphic registration³⁰ were computed using hyper-parameter optimization^{31,32} with dice output; planning and pre-treatment CTs were used as references/targets and the weekly CBCTs/mid/post-treatment CTs as sources for DIR, and the final transformation in the diffeomorphic registration is obtained using a symmetrized large deformation diffeomorphic metric mapping (LDDMM) algorithm³⁰ available in advanced normalization tools (ANTs). The transformation in LDDMM maps the corresponding points between two images by finding a geodesic solution. The integrated B-spline regularization models the DVF as a B-spline object to capture large deformation. This gives free-form elasticity to the converging/diverging vectors that represent a morphological shrinkage/expansion.³⁰

For deciding the optimal DIR hyper-parameters, we focused on grid size and gradient step which greatly influence the final results. We investigated different combinations of grid size (16, 32, 64, 128) and gradient step size (0.01, 0.1, 0.3, 0.5). The grid size number denotes grid size at the lowest resolution stage in three-step multi-resolution DIR and the grid size is reduced by multiples of two as the resolution increases. The mutual information was used as a metric and the iteration time for DIR was 100, 70, and 40 for each resolution step. The dice between the target and the warped contours (GTV/parotid for HN and GTV/esophagus

for lung) was used as the final output variable. The hyper-parameter optimization (solved using grid search) is formulated as follows:

$$\theta_{DIR}^* = \arg \max_{\theta_{DIR}} f_{Dice}(x | \theta_{DIR})$$

where f_{Dice} is the objective function that is a combined GTV+OAR dice in our context, θ_{DIR} is the set of hyper-parameters in DIR (grid size and gradient step size in this study). The purpose of hyper-parameter tuning is to retrieve optimal values that result in best DIR performance across datasets.

This resulting optimal combination of DIR hyper-parameters was used to create the training DVF. Five randomly selected patients from the training sets for each dataset were used in hyper-parameter optimization. We show that increasing the subset of patients for hyper-parameter optimization did not change the optimal DIR hyper-parameters combination in both HN and lung as shown in Figure 1. After deciding the optimal hyper-parameters for the DIR, we chose two additional hyper-parameter combinations near the optimal one to augment DVF dataset for training.

2.3 | The Seq2Seq deep learning model

LSTM networks are recurrent network with memory cell units capable of learning long-range dependency in a temporal sequence. Let T be the number of timepoints of interest and $(X_1, X_2, \dots, X_T)$ be a sequence of input to the recurrent network, where $X_t \in \mathbf{R}^{M \times N \times 3}$ is a vector field for a timepoint t , and $M \times N$ is the dimension of the vector field in the spatial domain. An LSTM unit contains a memory cell with cell status c_t and the input gate i_t , a forget gate f_t , an output gate o_t and an output hidden state H_t . Compared to conventional LSTM, ConvLSTM is capable of modeling 2D spatio-temporal data sequences by replacing LSTM's multiplication with spatial local convolution. The operation of a ConvLSTM unit is defined by the following equations^{18,20,23}:

The first step in ConvLSTM is to decide what information to be retained from the cell state by examining the hidden state from the previous timepoint and the input of the current timepoint through the forget gate, f_t ,

$$f_t = \sigma(W_f * [H_{t-1}, X_t] + b_f)$$

And then, the input gate, i_t decides which new features will be stored and a new candidate value, $\tilde{C}_t$, is formed.

$$i_t = \sigma(W_i * [H_{t-1}, X_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c * [H_{t-1}, X_t] + b_c)$$

Next the new cell state is updated with the new candidate value, $\tilde{C}_t$ and the previous cell state C_{t-1} , regulated by the forget and input gates.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

Finally, the ConvLSTM decides what features are going to pass through the output gate, o_t .

$$o_t = \sigma(W_o * [H_{t-1}, X_t] + b_o)$$

$$H_t = o_t \odot \tanh(C_t)$$

where σ is the sigmoid activation function, $*$ is the convolution operator, and $\odot$ is the Hadamard product. The input X_b , cell status C_b , hidden states H_b , forget gate f_b , input gate i_b , input-modulation gate $\tilde{C}_t$, and output gate o_t are all 3D tensor with the dimensions of [Width $\times$ Height $\times$ Features]. We used each slice's three-axis DVF $\in \mathbf{R}^{M \times N \times 3}$ for inferences to predict corresponded slice's future three-axis DVF. Through this model, we can predict each slice's three-axis deformation, and stacked each slice's deformation to generate future 3D volume three-axis deformation.

Figure 2a is a schematic image of the training process by our deep learning architecture, figure2b is ConvLSTM block. We built our recurrent network from a recurrent block containing an encoder-decoder sub-network similar to U-Net. Unlike²³ in which ConvLSTM is only used in the bottleneck layer, we replaced each convolution layer in the encoder and decoder with ConvLSTM to learn the encoding and decoding dependency from previous timepoints at different resolutions. The encoder consists of three ConvLSTM and two max-pooling layers; the decoder includes three ConvLSTM layers, two upsampling, and two convolution layers. We maintained the long skip connections crossing the encoder and the decoder at the same level of resolution such that fine-grained details can be recovered in the prediction. At the end of the decoder layers, there are two convolution layers to generate the final prediction of future sequential volumetric DVFs.

The recurrent block predicts the DVF of the next timepoint, $\hat{X}_{t+1}$. We defined the network training loss as follows:

$$Loss = \frac{1}{T-k} \sum_{t=k}^{T-1} LogCosh(X_{t+1}, \hat{X}_{t+1})$$

and

$$LogCosh(x, y) = \frac{1}{M \times N} \sum_{i=1}^{M \times N} \log(\cosh(|x_i - y_i|))$$

Log-Cosh loss works like L2 for small differences and as L1 for large differences. This means that the loss will not be so strongly affected by the occasional highly incorrect predictions.³³

During the training, we used all available timepoints. At inference, the input could be any sequence of previous K timepoints, $1, \dots, K$, $K < T$, and the network will predict the future sequence of timepoints $K + 1, \dots, T$. The proposed network has the following advantages: (1) it is possible to predict the future anatomical shape changes at a relatively earlier timepoint during the RT so that, if necessary, the treatment plan could be adapted to the predicted anatomical change in the early phase of treatment, and (2) it can be applied to different input/output sequences of timepoints to update the predictions on a rolling basis as more longitude images are acquired.

To compare with image-based prediction,²³ in addition to DVF, we trained the proposed network with sequences of $X_t \in \mathbf{R}^{M \times N \times 2}$, where the two channels are the image and its anatomical segmentation. We used adaptive moment estimation (Adam) for optimizer in training both DVF-based and image-based networks and learning rate α , β_1 , β_2 , and ϵ in Adam were 0.001, 0.9, 0.999, and 0.1, respectively.

We used the DVF representation for predicting future anatomical deformations due to its inherent nature in defining relationships with respect to the reference image (planning/pre-treatment images). Representing this relationship is also a clinical requirement since all contouring and dose planning is done on the planning/pre-treatment images and the later timepoints are normally just used for patient setup. We evaluated our model on both image- and DVF-based predictions and showed superior performance with DVF representation. The workstation that we used had Intel® Xeon® Silver 4110 CPU @ 2.10 GHz, 96 GB RAM, and NVIDIA 2080Ti graphic card.

2.4 | Evaluation

2.4.1 | Jacobian evaluation—The purpose of the Jacobian evaluation is to compare the deformation between the DIR and the predicted DVF. For the Jacobian calculation, a 3×3 Jacobian matrix j and its determinant J were calculated from the DVF $\vec{u}$ at every voxel:

$$j(u) = \begin{bmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{bmatrix}$$

$$J(u) = \text{Det}(j(u))$$

j is the first derivative of the DVF and is calculated at every voxel to produce a map of J .^{34–36}

The Jacobian map indicates the volumetric ratio of an object before and after transformation. $J > 1$ means volume expansion and $J < 1$ corresponds to volume shrinkage. As shown in Figures 3 and 4, we also quantitatively compared J from DIR-derived DVF and predicted DVF for GTV and parotid glands in HN dataset and for GTV and esophagus in the lung dataset.

2.4.2 | Geometric matching—The purpose of geometric matching evaluation is to assess the ability of the proposed approach to predict anatomical deformation between initial images and future images. In the case of image-based prediction, we directly measured the DICE coefficient, and average Hausdorff distance (d_{AVD}) between manual contours and predicted contours at a later timepoint. For geometric matching evaluation of DVF-based prediction, we measured the dice coefficient and d_{AVD} of contours warped by predicted DVFs and manual contours. The manual contours were delineated by experienced radiation oncologists. Moreover, we compared the volume of manual and predicted contours for weekly CBCT and for mid/post-treatment CTs by relative volume difference (RVD).^{23,37–40}

$$DICE = \frac{2|X \cap Y|}{|X| + |Y|}$$

$$d_{AVD}(X, Y) = \frac{1}{|X|} \sum_{x \in X} \min_{y \in Y} d(x, y) + \frac{1}{|Y|} \sum_{y \in Y} \min_{x \in X} d(x, y)$$

$$RVD = \frac{|V_x - V_y|}{|V_x|}$$

where X is ground truth and Y is the predicted contours, and V_x and V_y are the volume of ground truth and prediction, respectively. The $|X|$ and $|Y|$ are the number of points in X and Y , respectively. $d(x, y)$ is Euclidean distance metric between two points $x \in X$ and $y \in Y$

3 | RESULTS

Figure 5 shows two examples of Seq2Seq prediction results for lung GTV (blue) and esophagus (red). Both lung sequence examples demonstrate weeks 4–6 sequence prediction results using DVF and image representations in our Seq2Seq model. The lung sequence example 1 is from GTV undergoing inflammation (increase in volume) from weeks 1–3 but at the end exhibits overall 20% shrinkage from the start of the treatment.

Since our training data contained 18 cases (Figure 7) from the inflammation category, even though the lung example 1 input weeks 1–3 sequence volume is monotonically increasing, our DVF model is correctly able to predict the eventual shrinkage with sharper boundaries whereas the image-based prediction gets thrown off in later weeks (and produces significantly blurry results). Lung sequence example 2 belongs to the typical monotonically decreasing volume category and exhibits overall 50% shrinkage. The training/validation data

contained 35 cases from this category (Figure 7) and the Seq2Seq model shows good GTV and esophagus prediction performance on this category (Figure 5 and Table 1).

Note that the DVF-based prediction deforms planning CT images and hence produces sharper noise-/artifact-free predictions corresponding to the weekly CBCTs, showing another advantage of using DVF rather than image representation (where artifacts/noise in CBCT images can potentially throw off predictions).

Detailed quantitative evaluations on inflammation and monotonically decreasing volume categories for lung dataset are given in Figure 7 and Table 1. There were 18 category 1 (inflammation) and 35 category 2 (monotonic volume decrease) cases in training/validation sets, and 3 category 1 and 7 category 2 cases in the test set. The presented DVF Seq2Seq model was shown to accurately predict future GTV volumes even if there were volume fluctuations in the input sequence. Image-based prediction overestimated GTV volumes in the inflammation category cases by over 30% *RVD*. As compared to modest overall GTV shrinkage in category 1, the monotonic decreasing category exhibited larger shrinkage ranges, some as high as 50%. Understandably, the prediction accuracy was relatively lower in category 2 than category 1 but the overall prediction results for both categories with DVF representation are clinically reasonable (over 0.8 *DICE*).

Figure 6 shows the post-treatment prediction results for a HN patient given pre- and mid-treatment sequence input. We investigated GTV (blue) and the PG (red) for HN evaluation. Both GTV and PG shrunk during the RT and both image- and DVF-based predictions performed well though DVF-based prediction achieved 0.75 GTV *DICE*, approximately 13% better than image-based prediction. The PG normally exhibits the largest volume change among OARs during HN RT and the proposed DVF prediction obtained 0.72 *DICE* for iPG and 0.83 for cPG, approximately 23% and 4.5% better than image-based iPG and cPG predictions, respectively. Not only does DVF prediction achieves better accuracy, but the resulting images are also clear/sharp, overcoming the major bottleneck with the image-based prediction.

Since ConvLSTM models suffer from mode-averaging and limited long-term dependencies between sequence elements, we overcame the mode-averaging issue via DVF representation and limited long-term dependency issue via skip connections by directly propagating extracted multi-resolution longitudinal features from encoder to decoder ConvLSTM block. The metric evaluation for predictions via image versus DVF representation and with versus without skip connections is reported in Table 2 which shows *DICE*, average Hausdorff, and relative volume difference evaluation for GTVs and OARs in both lung and HN datasets. The DVF and skip connection combination outperforms image and with/without skip connection combination. Since our model has a flexible input/output sequence structure, we also report weeks 1–2 and weeks 1–3 sequence input predictions. The prediction results from weeks 1–2 were slightly lower than weeks 1–3, but the performance was still clinically valuable especially since DVF prediction method maintained relatively good results even though the input data for future prediction were insufficient. Furthermore, if we input sequences up to week 4 or week 5, we were able to obtain week 6 prediction with *DICE* of 0.83 ± 0.08 and 0.84 ± 0.08 , respectively. From these results, we demonstrated the advantage

of our proposed seq2seq model in handling various lengths of input sequence on a rolling basis as the timepoints are acquired to improve prediction accuracy.

The mean and standard deviation of GTV volume change from planning CT to week 6 CBCT were 111.2 ± 81.7 , 106.9 ± 79.9 , 101.4 ± 76.9 , 95.6 ± 74.6 , 88.7 ± 68.7 , 81.6 ± 62.6 , and 75.3 ± 60.1 , respectively. The testing data's GTV volume from week 4 to 6 was (110.2 ± 62.3 , 102.9 ± 57.3 , 101.1 ± 58.2), and the volume of predicted GTV by DVF-based method was (120.3 ± 70.4 , 118.8 ± 71.2 , 108.4 ± 61.5), and image-based method was (131.1 ± 65 , 136.8 ± 64.9 , 134.1 ± 64.5), respectively. Figure 7 is the volume change tendencies of GTV and we can confirm that the decreased tendencies of ground truth GTV and predicted GTV volume, especially predicted by DVF-based method, were similar to each other.

And we also confirmed that the prediction performance was slightly degraded as predicting timepoint was increased, and this phenomenon was derived from error accumulation of previous predictions.

4 | DISCUSSION

During the RT, the patient's anatomy may significantly change due to weight loss and radiation-induced biological effects. These changes can increase uncertainty of radiation dose delivered during the RT. In effect, it is highly desirable to predict patient's radiation-induced anatomical changes in advance. Recently, longitudinal prediction deep learning models have been proposed for medical imaging which outperform more traditional mathematical models.²³ This attests to the promising potential for time-series medical image prediction for the RT. Previous studies were evaluated on relatively small datasets with only single timepoint prediction and small region-of-interest (ROI) image patches.²³

In this paper, we used three timepoint CT 28 patient HN dataset and six timepoint multimodal CT/CBCT 63 patient lung dataset for developing GTV and OAR prediction. To the best of our knowledge, ours is the largest longitudinal clinical imaging dataset used for developing and evaluating longitudinal sequential prediction. Furthermore, our study has the following strengths: (1) we used the entire 3D field-of-view (FOV) of the images rather than the cropped ROI images. In RT, GTV and the surrounding OARs are confounding factors in planning for delivering a high conformal dose to the target while minimizing radiation dose to the normal tissue. Hence, it is necessary to predict anatomical changes in large FOV during the RT. For our study, we only cropped air and table from the original volume data and used $256 \times 256 \times 256$ volume size in HN dataset and $180 \times 180 \times 128$ volume size in lung dataset. Both included all the anatomical structures originally scanned.

(2) We proposed DVF representation instead of image representation for driving the prediction in the deep learning model since the intensity of a pixel or voxel for the training is one of the factors that have the greatest influence on the extracted features. DVF-based training (with respect to the planning CT) for longitudinal sequential anatomical prediction is more effective since there is no image noise or artifacts involved especially for the lung CBCT cases. The low-dose CBCT modality, generally used for checking patient = position weekly during the RT, has significant image noise and artifacts. Thus, training with CBCT

imaging may suffer from significant bias from these image noise and artifacts. Using the DVF prediction, the high-quality planning CT images (along with the contour and planned dose information) can be deformed to the later timepoints thus allowing for re-planning given the same Hounsfield unit of the deformed image as the planning CT. Moreover, the DVF representation per-voxel contains three-axis directional deformation which aids in 3D spatio-temporal prediction.

Because we used the DVF for training, the accuracy of DVF is critical. Therefore, we used LDDMM-based DIR, which is one of the state-of-the-art DIR algorithms. Given that we had manual contours on all the timepoints, we were able to use a fraction of the training data in the hyper-parameter optimization framework to find optimal values for the most sensitive LDDMM hyper-parameters. This resulted in a high-quality DVF training data.

(3) We used a multi-resolution ConvLSTM-based model. The previous Seq2Seq model used ConvLSTM or LSTM block only for latent features at the bottleneck whereas our proposed model uses ConvLSTM block in a multi-resolution block thus allowing for deeper learning of the spatial-temporal features at different resolutions. For example, the week 6 prediction of lung GTV with the multi-resolution ConvLSTM model was 0.81 ± 0.10 average DICE, but the prediction result with the model that has ConvLSTM block only in the bottleneck part was 0.79 ± 0.09 .

(4) Lastly our study comprises of larger and longer real clinical sequence datasets than reported in previous works. We used six-timepoint sequences from 63 non-small cell lung cancer patients. In comparison, Zhang et al.²³ only predicted a single future timepoint given two previous time points. Furthermore, we studied multiple prediction results from different length input/output sequences. Our model provides the flexibility in the clinical treatment workflow of radiotherapy to make longitudinal predictions on a rolling basis as more weekly images are acquired during the treatment.

To summarize, our proposed study can effectively predict future DVFs hence it can be used for various purposes including image, contour transformation, and dose accumulation. Figure 8 is two examples of DVF and week 6 GTV contour prediction. The predicted DVF has high similarity with the actual DVF and the predicted GTV agreed well to the ground truth GTV. In addition, as shown in Figure 9, the proposed method can predict high-quality weekly images by warping the high-quality planning CT using the predicted DVF. The GTV regions of the predicted images closely resembled the GTV region of the warped image by using the actual DVF from the registration of weekly CBCT with the planning CT.

There are several future directions for improving our model. First, we plan to extend ConvLSTM block to 3D. Our model used three-channel DVFs of multiple slices as input; each slice contains three-axis DVF. The multiple slice 3D DVFs are combined into the mini-batch structure and fed to the trained model. By replacing the 2D convolution operation of the ConvLSTM block with 3D convolution, we plan to develop a model that can train 3D DVF of the entire volume at once. To the best of our knowledge, there are no previous studies with a complete 3D seq2seq model due to the memory constraint. We are currently extending our model to 3D using high-performance computing cluster. Second, we plan to

apply a spatial transformation layer for unsupervised Seq2Seq learning. Recently, several unsupervised learning models have been proposed, such as VoxelMorph⁴¹ for medical imaging DIR. Since unsupervised learning has the advantage of predicting DVFs from image pairs, it is expected to be able to predict DVF at each timepoint from the sequence image. Lastly, it is expected that the performance can be further improved by applying attention to features that have greater influence on the future sequence using an attention block. In the future, we will also evaluate incorporating more recent ConvLSTM^{25,42} and transformer-based modules (handling missing/noisy observations more naturally)^{43,44} in our proposed seq2seq model along with evaluating these on a much larger 1000 patient lung radiotherapy cohort we are assembling.

5 | CONCLUSION

In this work, we presented a novel Seq2Seq model for medical images, leveraging the complete 3D imaging information of a relatively large longitudinal clinical dataset, to carry out longitudinal GTV/OAR time-series predictions for HN and lung radiotherapy patients. We also quantitatively confirmed that we can derive better longitudinal predictions using DVF rather than image representation. To the best of our knowledge, ours is the first DVF-based time-series prediction study applicable to the field of RT. The presented approach has the potential to improve RT patient outcomes by predicting the longitudinal GTV/OAR changes in advance.

ACKNOWLEDGMENTS

This project was supported by MSK Cancer Center Support Grant/Core Grant (P30 CA008748).

Funding information

Varian Medical Systems (Varian); Memorial Sloan-Kettering Cancer Center (MSK); MSK Cancer Center, Grant/Award Number: P30 CA008748

DATA AVAILABILITY STATEMENT

The head and neck squamous cell carcinoma (HNSCC) data that support the findings of this study are openly available at <https://doi.org/10.7937/K9/TCIA.2018.13upr2xf>. Reference number for HNSCC data [28].

REFERENCES

1. Hugo GD, Weiss E, Sleeman WC, et al. A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer. *Med Phys*. 2017;44:762–771. [PubMed: 27991677]
2. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys*. 2017, 44, e43–e76. [PubMed: 28376237]
3. Jaffray DA. Emergent technologies for 3-dimensional image-guided radiation delivery. *Semin Radiat Oncol*. 2005;15:208–216.
4. Schwartz DL, Garden AS, Shah SJ, et al. Adaptive radiotherapy for head and neck cancer—dosimetric results from a prospective clinical trial. *Radiother Oncol*. 2013;106:80–84. [PubMed: 23369744]

5. Jaffray DA. Image-guided radiotherapy: from current concept to future perspectives. *Nature Reviews Clinical Oncology*. 2012;9:688.
6. Yan D, Ziaja E, Jaffray D, et al. The use of adaptive radiation therapy to reduce setup error: a prospective clinical study. *Int J Radiat Oncol Biol Phys*. 1998;41:715–720. [PubMed: 9635724]
7. Mencarelli A, van Kranen SR, Hamming-Vrieze O, et al. Deformable image registration for adaptive radiation therapy of head and neck cancer: accuracy and precision in the presence of tumor changes. *Int J Radiat Oncol Biol Phys*. 2014;90:680–687. [PubMed: 25151537]
8. Munshi A, Pandey MB, Durga T, Pandey KC, Bahadur S, Mohanti BK. Weight loss during radiotherapy for head and neck malignancies: what factors impact it? *Nutr Cancer*. 2003;47:136–140. [PubMed: 15087265]
9. Lee C, Langen KM, Lu W, et al. Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration. *Int J Radiat Oncol Biol Phys*. 2008;71:1563–1571. [PubMed: 18538505]
10. Wang C, Rimner A, Hu Y-C, et al. Toward predicting the evolution of lung tumors during radiotherapy observed on a longitudinal MR imaging study via a deep learning algorithm. *Med Phys*. 2019;46:4699–4707. [PubMed: 31410855]
11. Brouwer CL, Steenbakkers RJ, Langendijk JA, Sijtsma NM. Identifying patients who may benefit from adaptive radiotherapy: does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiother Oncol*. 2015;115:285–294. [PubMed: 26094076]
12. Yan D, Wong J, Vicini F, et al. Adaptive modification of treatment planning to minimize the deleterious effects of treatment setup errors. *Int J Radiat Oncol Biol Phys*. 1997;38:197–206. [PubMed: 9212024]
13. Nadeem S, Zhang P, Rimner A, Sonke J-J, Deasy JO, Tannenbaum A. LDeform: longitudinal deformation analysis for adaptive radiotherapy of lung cancer. *Med Phys*. 2020;47:132–141. [PubMed: 31693764]
14. Lee D, Zhang P, Nadeem S, et al. Predictive dose accumulation for HN adaptive radiotherapy. *Phys Med Biol*. 2020;65(23):235011. [PubMed: 33007769]
15. Wang C, Xu L-Y, Fan J-S. A general deep learning framework for history-dependent response prediction based on UA-Seq2Seq model. *Comput Methods Appl Mech Eng*. 2020;372:113357.
16. Liu T, Wang K, Sha L, Chang B, Sui Z. Table-to-text generation by structure-aware seq2seq learning. *arXiv preprint arXiv:1711.09724*. 2017.
17. Liu B, Yan S, Li J, et al. A sequence-to-sequence air quality predictor based on the n-step recurrent prediction. *IEEE Access*. 2019;7:43331–43345.
18. Kim S, Hong S, Joh M, Song S-K. Deeprain: ConvLstm network for precipitation prediction using multichannel radar data. *arXiv preprint arXiv:1711.02316*. 2017.
19. Kim S, Kim H, Lee J, Yoon S, Kahou S. Deep-hurricane-tracker: tracking and forecasting extreme climate events. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2019;1761–1769. <https://ieeexplore.ieee.org/abstract/document/8658402>
20. Mukherjee S, Ghosh S, Ghosh S, Kumar P, Roy PP. Predicting video-frames using encoder-convlstm combination. *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019; 2027–2031. <https://ieeexplore.ieee.org/document/8682158>
21. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735–1780. [PubMed: 9377276]
22. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. 2014.
23. Zhang L, Lu L, Wang X, et al. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. *IEEE Trans Med Imaging*. 2020;39:1114–1126. [PubMed: 31562074]
24. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-C. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2015;28:802–810.

25. Lange B, Itkina M, Kochenderfer MJ. Attention Augmented ConvLSTM for Environment Prediction. arXiv preprint arXiv:2010.09662.2020.
26. Tang C, Salakhutdinov RR. Multiple futures prediction. In: Advances in Neural Information Processing Systems (NeurIPS). Vol 32. Red Hook, NY: Curran Associates Inc.; 2019:15424–15434.
27. Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error. In: International Conference on Learning Representations. 2016. <https://arxiv.org/abs/1511.05440>
28. Bejarano T, De Ornelas-Couto M, Mihaylov IB. Longitudinal fan-beam computed tomography dataset for head-and-neck squamous cell carcinoma patients. Med Phys. 2019;46:2526–2537. [PubMed: 30806479]
29. Alam S, Thor M, Rimner A, et al. Quantification of accumulated dose and associated anatomical changes of esophagus using weekly Magnetic Resonance Imaging acquired during radiotherapy of locally advanced lung cancer. Phys Imag Radiat Oncol. 2020;13:36–43.
30. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage. 2011;54:2033–2044. [PubMed: 20851191]
31. Le Folgoc L, Delingette H, Criminisi A, Ayache N. Quantifying registration uncertainty with sparse bayesian modelling. IEEE Trans Med Imaging. 2016;36:607–617. [PubMed: 27831863]
32. Le Folgoc L, Delingette H, Criminisi A, Ayache N. Sparse Bayesian registration of medical images for self-tuning of parameters and spatially adaptive parametrization of displacements. Med Image Anal. 2017;36:79–97. [PubMed: 27870999]
33. Jadon SA survey of loss functions for semantic segmentation. arXiv preprint arXiv:2006.14822.2020.
34. Riyahi S, Choi W, Liu C-J, et al. Quantifying local tumor morphological changes with Jacobian map for prediction of pathologic tumor response to chemo-radiotherapy in locally advanced esophageal cancer. Phys Med Biol. 2018;63:145020. [PubMed: 29911659]
35. Chung M, Worsley K, Paus T, et al. A unified statistical approach to deformation-based morphometry. NeuroImage. 2001;14:595–606. [PubMed: 11506533]
36. Fuentes D, Contreras J, Yu J, et al. Morphometry-based measurements of the structural response to whole-brain radiation. Int J Comput Assist Radiol Surg. 2015;10:393–401. [PubMed: 25408306]
37. Wong KCL, Summers RM, Kebebew E, Yao J. Pancreatic tumor growth prediction with elastic-growth decomposition, image-derived motion, and FDM-FEM coupling. IEEE Trans Med Imaging. 2017;36:111–123. [PubMed: 27529869]
38. Roque T, Risser L, Kersemans V, et al. A DCE-MRI driven 3-D reaction-diffusion model of solid tumor growth. IEEE Trans Med Imaging. 2018;37:724–732. [PubMed: 29533893]
39. Zhang L, Lu L, Summers RM, Kebebew E, Yao J. Personalized pancreatic tumor growth prediction via group learning. In: International Conference on Medical Image Computing and Computer-assisted Intervention – MICCAI-2017. Lecture Notes in Computer Science 10434. Basel, Switzerland: Springer; 2017:424–432. 10.1007/978-3-319-66185-8_48
40. Zhang L, Lu L, Summers RM, Kebebew E, Yao J. Convolutional invasion and expansion networks for tumor growth prediction. IEEE Trans Med Imaging. 2018;37:638–648. [PubMed: 29408791]
41. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans Med Imaging. 2019;38:1788–1800.
42. Su J, Byeon W, Huang F, Kautz J, Anandkumar A. Convolutional Tensor-Train LSTM for Spatio-temporal Learning. In: Advances in Neural Information Processing Systems (NeurIPS). Vol 33. Red Hook, NY: Curran Associates, Inc.; 2020:13714–13726.
43. Giuliari F, Hasan I, Cristani M, Galasso F. Transformer networks for trajectory forecasting. arXiv preprint arXiv:2003.08111.2020.
44. Liu Z, Luo S, Li W, Lu J, Wu Y, Li C, Yang L. ConvTransformer: a convolutional transformer network for video frame synthesis. arXiv preprint arXiv:2011.10185.2020.

FIGURE 1.

Optimal DIR hyper-parameter selection via hyper-parameter optimization. The optimal hyper-parameter values for input grid size and gradient step were 32 and 0.3 for HN dataset in (a) and (b) and were 32 and 0.1 for lung dataset in (c) and (d).

FIGURE 2.

(a) Pipeline of proposed ConvLSTM model, (b) Schematic images for multi resolution ConvLSTM block. Max pooling and upsampling layers are used for multi-resolution longitudinal features analysis

FIGURE 3.

Jacobian evaluation on HN and lung datasets with prediction and DIR-derived DVFs. The Jacobian integral for predicted and DIR DVFs showed high correlation, $R^2 = 0.85$.

FIGURE 4.

Comparison of DIR and predicted Jacobian maps for the lung example in Figure 1 and HN example in Figure 2. $J < 1$ indicates shrinkage, $J > 1$ expansion, and $J = 1$ is no change.

FIGURE 5.

DVF- versus image-based sequence prediction of weeks 4, 5, and 6 given the weeks 1, 2, and 3 sequence as input. Note that our DVF representation (unlike image representation) does not contain any high-frequency components and hence, is not affected by the ConvLSTM module, resulting in sharper planning CT deformed weeks 4, 5, and 6 sequence prediction. Blue contours denote gross tumor volume (GTV) and red denote esophagus (ESO) contours. Deformed planning CT weekly predictions result in artifact/noise-free images as opposed to the CBCT artifact-ridden and blurry predictions with the image representation. *Lung sequence example 1 (Left)*, predicted week4 Dice: DVF-GTV 0.872, Image-GTV 0.89, DVF-ESO 0.73, Image-ESO 0.73; week 5 Dice: DVF-GTV 0.83, Image-GTV 0.82, DVF-ESO 0.82, Image-ESO 0.81; week 6 Dice: DVF-GTV 0.83, Image-GTV 0.79, DVF-ESO 0.76, Image-ESO 0.76. *Lung sequence example 2 (Right)*, predicted week 4 Dice: DVF-GTV 0.77, Image-GTV 0.70, DVF-ESO 0.77, Image-ESO 0.80; week 5 Dice: DVF-GTV 0.73, Image-GTV 0.65, DVF-ESO 0.80, Image-ESO 0.81; week 6 Dice: DVF-GTV 0.64, Image-GTV 0.59, DVF-ESO 0.82, Image-ESO 0.80

FIGURE 6.

HN radiotherapy dataset example illustrating DVF- versus image-based prediction results for post-treatment CT given pre- and mid-treatment CTs as input sequence. Blue contours represents GTV and red denote parotid gland, ipsilateral (iPG) and contralateral (cPG) sides. Dice, average Hausdorff (AvHD) and relative volume difference (RVD) for image prediction are as follows, GTV-DICE: 0.67, GTV-AvHD: 0.62, GTV-RVD: 26.8%, iPG-DICE: 0.59, iPG-AvHD: 1.12, iPG-RVD: 25.8%, cPG-DICE: 0.80, cPG-AvHD: 0.33, cPG-RVD: 17.9%. For DVF prediction, GTV-DICE: 0.75, GTV-AvHD: 0.39, GTV-RVD: 7.9%, iPG-DICE: 0.72, iPG-AvHD: 0.48, iPG-RVD: 24.1%, cPG-DICE: 0.83, cPG-AvHD: 0.24, cPG-RVD: 16.2%

FIGURE 7.

The absolute volume of two different lung GTV categories: Category 1 (blue) with inflammation leading to non-monotonic volume changes through the course of treatment and category 2 (red) with monotonically decreasing volume. There were 18 category 1 and 35 category 2 cases in training/validation and 3 category 1 and 7 category 2 cases in testing.

FIGURE 8.

Two examples of DVF, and contour prediction. The background image is week 6 CBCT. The blue and the red arrow are actual and predicted DVF, respectively. Yellow, green, blue, and red contour are planning, Week 6 ground truth, warped GTV by actual DVF and predicted DVF, respectively.

FIGURE 9.

The comparison among the planning CT, week6 CBCT, warped image by actual DVF (planning CT → week 6 CBCT), and warped image by predicted DVF. The red arrow indicated the tumor region

TABLE 1
Lung GTV prediction metrics evaluation for inflammation (1) and monotonically decreasing (2) volume categories

Structures	Image Pred.		DVF Pred.		Image Pred.		DVF Pred.		Image Pred.		DVF Pred.	
	DICE ↑	$d_{AVD}(mm) ↓$	DICE ↑	$d_{AVD}(mm) ↓$	DICE ↑	$d_{AVD}(mm) ↓$	DICE ↑	$d_{AVD}(mm) ↓$	RVD (%) ↓	RVD (%) ↓	RVD (%) ↓	RVD (%) ↓
Category 1-Week 4	0.85 ± 0.11	0.85 ± 0.12	0.85 ± 0.12	0.26 ± 0.28	0.21 ± 0.20	4.8 ± 1.5	4.7 ± 3.4					
Category 1-Week 5	0.85 ± 0.10	0.86 ± 0.10	0.86 ± 0.10	0.28 ± 0.33	0.20 ± 0.15	8.6 ± 5.6	8.1 ± 8.1					
Category 1-Week 6	0.86 ± 0.06	0.87 ± 0.08	0.87 ± 0.08	0.21 ± 0.20	0.19 ± 0.12	8.8 ± 10.6	8.6 ± 1.7					
Category 2-Week 4	0.79 ± 0.11	0.82 ± 0.09	0.82 ± 0.09	0.35 ± 0.23	0.27 ± 0.19	28.1 ± 32.4	20.1 ± 25.9					
Category 2-Week 5	0.76 ± 0.10	0.80 ± 0.07	0.80 ± 0.07	0.39 ± 0.20	0.29 ± 0.19	27.6 ± 26.5	21.9 ± 17.5					
Category 2-Week 6	0.73 ± 0.13	0.81 ± 0.10	0.81 ± 0.10	0.48 ± 0.30	0.35 ± 0.21	35.9 ± 34.5	28.0 ± 31.2					

TABLE 2

Metrics (Dice, average Hausdorff, relative volume difference) evaluation for HN and lung datasets

Structures	Image Pred.	DVF Pred.	Image Pred.	DVF Pred.	Image Pred.	DVF Pred.
	DICE ↑		$d_{AVD}(mm) ↓$		RVD (%) ↓	
HN post-treatment image predictions without skip connections						
GTV-HN	0.82 ± 0.11	0.84 ± 0.11	0.33 ± 0.24	0.22 ± 0.25	12.6 ± 10.6	4.6 ± 2.6
iPG-HN	0.76 ± 0.08	0.78 ± 0.08	0.52 ± 0.25	0.33 ± 0.27	13.7 ± 14.1	10.9 ± 9.2
cPG-HN	0.81 ± 0.05	0.84 ± 0.05	0.31 ± 0.13	0.22 ± 0.18	6.4 ± 4.6	9.1 ± 9.5
HN post-treatment image predictions with skip connections						
GTV-HN	0.84 ± 0.10	0.87 ± 0.07	0.28 ± 0.21	0.21 ± 0.13	11.9 ± 14.1	4.5 ± 2.5
iPG-HN	0.76 ± 0.11	0.81 ± 0.06	0.52 ± 0.34	0.32 ± 0.13	13.3 ± 14.5	8.1 ± 8.2
cPG-HN	0.84 ± 0.04	0.86 ± 0.02	0.28 ± 0.21	0.21 ± 0.13	5.7 ± 6.0	4.5 ± 6.5
Lung Weeks 4–6 prediction (without skip connections)						
GTV-Lung-Week 4	0.81 ± 0.09	0.82 ± 0.10	0.31 ± 0.19	0.28 ± 0.20	24.2 ± 31.5	19.1 ± 25.5
GTV-Lung-Week 5	0.79 ± 0.07	0.81 ± 0.08	0.36 ± 0.15	0.29 ± 0.17	34.7 ± 27.6	22.3 ± 20.2
GTV-Lung-Week 6	0.77 ± 0.10	0.79 ± 0.11	0.42 ± 0.22	0.33 ± 0.22	48.0 ± 45.8	26.7 ± 34.7
Eso-Lung-Week 4	0.73 ± 0.06	0.75 ± 0.03	0.36 ± 0.11	0.29 ± 0.07	27.9 ± 23.6	27.9 ± 12.4
Eso-Lung-Week 5	0.73 ± 0.06	0.76 ± 0.05	0.42 ± 0.17	0.36 ± 0.16	26.7 ± 19.8	27.4 ± 14.6
Eso-Lung-Week 6	0.73 ± 0.05	0.77 ± 0.04	0.41 ± 0.17	0.30 ± 0.08	28.9 ± 28.4	19.1 ± 17.1
Lung Weeks 4–6 prediction (with skip connections)						
GTV-Lung-Week 4	0.81 ± 0.09	0.83 ± 0.09	0.31 ± 0.24	0.26 ± 0.19	21.9 ± 28.7	15.5 ± 22.5
GTV-Lung-Week 5	0.80 ± 0.07	0.82 ± 0.08	0.36 ± 0.23	0.29 ± 0.18	21.8 ± 23.7	17.8 ± 16.2
GTV-Lung-Week 6	0.79 ± 0.09	0.81 ± 0.10	0.40 ± 0.20	0.31 ± 0.21	27.8 ± 31.4	22.2 ± 27.1
Eso-Lung-Week 4	0.76 ± 0.04	0.77 ± 0.03	0.29 ± 0.18	0.29 ± 0.08	24.1 ± 6.8	23.6 ± 11.6
Eso-Lung-Week 5	0.76 ± 0.07	0.77 ± 0.05	0.35 ± 0.15	0.35 ± 0.16	24.0 ± 14.8	22.7 ± 13.8
Eso-Lung-Week 6	0.77 ± 0.05	0.77 ± 0.03	0.29 ± 0.08	0.29 ± 0.08	23.5 ± 14.7	15.0 ± 15.4
Lung 3–6 prediction (with skip connections)						
GTV-Lung-Week 3	0.83 ± 0.10	0.84 ± 0.13	0.26 ± 0.24	0.24 ± 0.17	20.4 ± 29.4	19.1 ± 12.0
GTV-Lung-Week 4	0.80 ± 0.12	0.81 ± 0.11	0.34 ± 0.27	0.31 ± 0.23	27.1 ± 37.5	26.7 ± 42.5
GTV-Lung-Week 5	0.76 ± 0.11	0.80 ± 0.09	0.37 ± 0.24	0.31 ± 0.21	33.3 ± 42.6	31.1 ± 31.3
GTV-Lung-Week 6	0.76 ± 0.14	0.78 ± 0.12	0.45 ± 0.34	0.36 ± 0.27	44.2 ± 58.6	36.1 ± 48.7

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Structures	Image Pred.		DVF Pred.		Image Pred.		DVF Pred.		Image Pred.		DVF Pred.	
	DICE ↑		$d_{AVD}(mm)$ ↓		RVD (%) ↓							
Eso-Lung-Week 3	0.78 ± 0.03		0.30 ± 0.04		0.30 ± 0.09		27.8 ± 20.3		26.8 ± 18.3			
Eso-Lung-Week 4	0.77 ± 0.04		0.29 ± 0.03		0.29 ± 0.06		27.9 ± 10.4		26.4 ± 10.1			
Eso-Lung-Week 5	0.76 ± 0.05		0.36 ± 0.13		0.36 ± 0.14		27.4 ± 14.6		25.9 ± 14.4			
Eso-Lung-Week 6	0.77 ± 0.04		0.30 ± 0.07		0.30 ± 0.05		24.3 ± 14.7		17.1 ± 16.0			