

# Application of deep neural networks for automatic planning in radiation oncology treatments

A.M. Barragán-Montero<sup>1\*2</sup>, D. Nguyen<sup>2</sup>, W. Lu<sup>2</sup>, M. Lin<sup>2</sup>, X. Geets<sup>3</sup>, E. Sterpin<sup>1,4</sup>, S. Jiang<sup>2</sup>

1- UCLouvain - Center of Molecular Imaging, Radiotherapy and Oncology - Brussels, Belgium

2- UTSouthwestern - Medical Artificial Intelligence and Automation Laboratory - Dallas, USA

3- Cliniques universitaires Saint-Luc - Department of Radiation Oncology - Brussels, Belgium.

4- KU Leuven - Laboratory of Experimental Radiotherapy - Leuven, Belgium.

**Abstract.** Treatment planning for radiotherapy patients is a time-consuming and manual process. In this work, we investigate the use of deep neural networks to learn from previous clinical cases and directly predict the optimal dose distribution for a new patient. The proposed model combines two architectures, UNet and DenseNet, and used mean squared error as loss function. Ten input channels were used to include dosimetric and anatomical information. A set of 100 patients was used for training/validation and 29 for testing. Dice similarity coefficients  $\geq 0.9$  for the isodose-lines in the predicted versus the clinical dose showed the excellent accuracy of the model.

## 1 Introduction

Treatment planning for radiotherapy patients is a time-consuming and heuristic process, where the medical physicist manually adjusts the radiation dose objectives for the target volume and organs at risk, until the desired dose distribution is achieved. Depending on the complexity of the case, generating a good treatment plan might require a high level of human expertise and may take from few hours to days. This entails a plan quality variability that, together with the extended planning time, may lead to suboptimal plans and delayed treatment delivery, which can compromise the final treatment outcome.

In order to overcome these problems, we propose to use deep neural networks to learn from previous clinical cases with high quality plans and directly predict the optimal dose distribution associated to a new patient. The predicted dose can later be used as a voxel-wise objective to automatically generate the corresponding treatment plan. This removes all human intervention and brings a two-fold improvement: 1) reduced plan variability by ensuring high quality plans, and 2) decreased planning time.

An extra challenge in the application of deep neural networks to this domain is the high variability in the patient dose distribution, given that the radiation beam configuration (i.e. the number of beams used and their angles) is different from patient to patient. Recent studies have already reported promising results by using convolutional neural networks (CNN) to predict patient specific voxel-wise dose distributions from anatomical information (i.e., contours of the involved organs and/or

---

\* Ana M. Barragán Montero is supported by Fonds Baillet-Latour.

computed tomography (CT) images), either in a slice-by-slice manner (2D) [1-3] or directly as a 3D matrix [4]. However, these models rely on consistent beam configuration across the patient database used for training, which ensures an accurate dose prediction for cases with similar beam settings, but impedes the generalization of the model to more realistic clinical cases with variable beam configuration.

In this work, we investigate the use of a CNN with an architecture that combines U-net [5] and DenseNet [6], for the prediction of 3D dose distributions from a very heterogeneous database in terms of beam configuration. In addition, we quantify the added value of including both anatomical and dosimetric information as input to the CNN, in order to build a single model that is robust to variable beam configuration, achieving a more comprehensive automatic planning with a potentially easier clinical implementation.

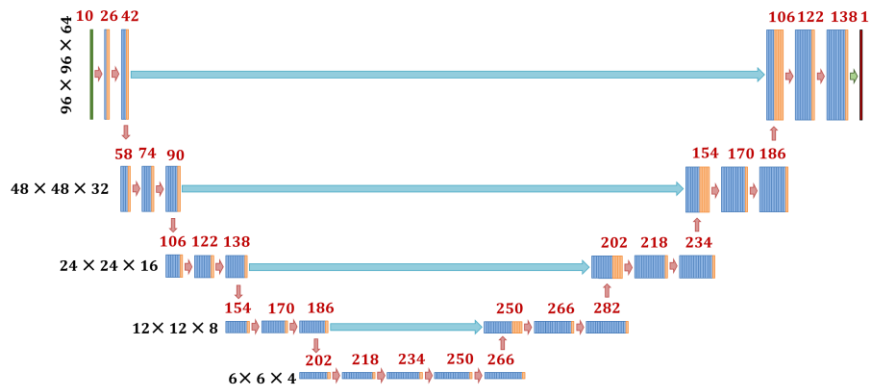
## 2 Methods

### 2.1 Model architecture

The proposed model is based on the popular U-net, published by Ronneberger et al. in 2015 [5]. The U-net is a type of fully convolutional network that it is able to include both local and global features from the input images to generate a pixel-wise (two-dimensional, 2D) prediction. To avoid prediction errors on the superior and inferior edges of the volumes, a three-dimensional (3D) variant of the classical 2D U-net was developed. In addition, the U-net was modified with the densely connected convolutional architecture used in DenseNet, proposed by Huang et al. in 2017 [6], to achieve more efficient feature propagation. The DenseNet densely connects its convolutional layers in a feed-forward fashion, using the feature-maps of all preceding layers as inputs for the current layer. This reduces the vanishing-gradient problem, enhances feature propagation and reuse, and decreases the number of trainable parameters. However, the dense connection between all layers significantly increases the memory usage. Thus, some connections between groups of layers were skipped, in order to maintain a reasonable RAM usage [7]. In particular, the convolutional layers in our model are densely connected within levels of the same resolution in the U-net, between each max pooling and up-sampling operation. We refer to each of these levels to as “hierarchy”, which motivated naming this network as Hierarchically Densely Connected U-net (HD U-net [4]). This HD U-net combines the efficient feature propagation and reuse of DenseNet, with the ability of U-net to infer the global and local image information, while keeping a reasonable memory usage. Mean squared error between the predicted and the clinical dose was used as the loss function for training our model. The detailed architecture of the HD U-net used in this study is presented in Figure 1, and the technical elements regarding the operations between layers has been previously described elsewhere [4], as well as a comparison between the HD U-net, Dense U-net, standard U-net for application to dose prediction.

Two different models were trained and compared. The first model (standard) comprised 9 input channels including only anatomical information, while the second model (robust) contained 10 input channels (Figure 1), including both dosimetric (1 channel) and anatomical information (9 channels). The dosimetric information was

provided in a 3D matrix containing the sum of the non-modulated beam's eye view ray-tracing dose distribution, computed by a fluence-convolution broad beam (FCBB) dose algorithm [8] for all involved beams. Note that the difference between this elementary dosimetric information provided in the 10<sup>th</sup> input channel and the predicted final dose distribution (i.e., the dose delivered to the patient) is mainly the beam's intensity modulation. Therefore, what the model needs to learn is the optimal modulation of the dose for a given patient, provided specific anatomical information. The anatomical information comprised 3D binary matrices, i.e. 1 inside the volume and 0 outside, for the planning target volume (PTV) and 8 relevant organs at risk (OARs) for lung treatment planning: body, heart, esophagus, spinal cord, right and left lungs, both lungs minus the target, and carina.



### Legend and Operations

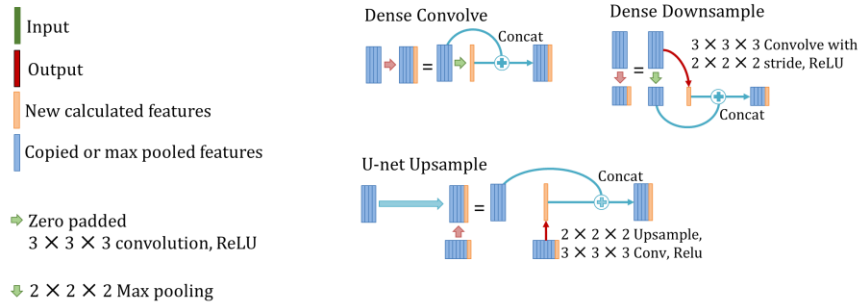


Fig. 1: HD U-net architecture. Black numbers represent the volume shape and resolution at a specific hierarchy, and red numbers are the number of feature maps.

## 2.2 Training and testing

A set of 129 lung cancer patients treated with intensity modulated radiation therapy (IMRT), with very heterogeneous beam configuration (4 to 9 beams) and orientation, was used for training/validation (100 patients) and testing (29 patients). The number

of input samples to the model was the same as the aforementioned number of patients, since no data augmentation was used. Three-dimensional images of  $5\text{mm}^3$  voxels ( $128 \times 128 \times S$ , being  $S$  the number of slides in the patient's inferior to superior direction) were used as input, although the patch size used for neural training was  $96 \times 96 \times 64$  (Figure 1). The stability of the model was evaluated by using a 5-fold cross-validation approach, where the model was randomly initialized, trained with 80 patients, and validated with the remaining 20 patients, using a different training/validation combination for each fold. The accuracy of the two methods (standard and robust) was evaluated by computing dice similarity coefficients ( $2(X \cap Y) / (X + Y)$ ) between the isodose lines from the predicted ( $D^{\text{p}_{\text{robust}}}$  and  $D^{\text{p}_{\text{standard}}}$ ) and clinical ( $D^{\text{c}}$ ) dose distributions. The models were trained on an NVIDIA TESLA K80 GPU with 12 GB dedicated RAM.

### 3 Results

The robust model, including dosimetric information as input data, outperformed the standard model, which only included anatomical information. Dice similarity coefficients for the isodose lines (Figure 2) showed that the high dose region (from the 80% isodose line) was predicted with comparable accuracy by the two models (dice around 0.9), but a major improvement was found in the medium to low dose region (from 0% to 60-70% isodose line), where the robust model achieves a dice about 10% higher than the standard model.

As illustration of the 3D dose distribution predicted by the two models, the results for one of the test patients is presented in Figure 3, for an axial slice located at the center of the target, for  $D^{\text{c}}$ ,  $D^{\text{p}_{\text{robust}}}$  and  $D^{\text{p}_{\text{standard}}}$ . The rest of the patients are not presented here due to limited space in the manuscript, but the behavior is similar for all of them: the standard model predictions show a very isotropic dose gradient that uniformly decrease from the target till the edge of the body, while the robust model is able to capture the dose features along the beam path thanks to the additional dosimetric information.

The two models, standard and robust, were trained during 150 epochs, which took about 15 hours in both cases, although the final model in both cases (standard and robust) was selected as the one corresponding to the epoch with lower validation loss. Additional training for a larger number of epochs was investigated but did not result in significant improvement. The average prediction time and its standard deviation was  $11.42 \pm 0.12$  s per patient for the standard model, and  $11.66 \pm 0.14$  s for the robust model.

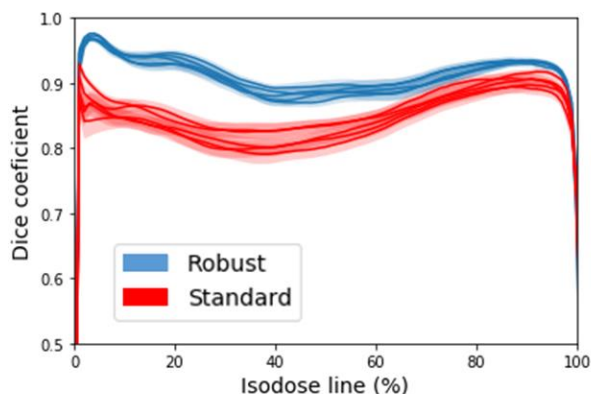


Fig 2. Dice similarity coefficients of the isodose volumes from 0 to 100% of the prescription dose for  $D^p_{\text{robust}}$  (solid blue lines) and  $D^p_{\text{standard}}$  (solid red lines) versus  $D^c$ , together with their corresponding standard deviation (color wash), for the 5 models obtained after cross-validation, evaluated on the test set.

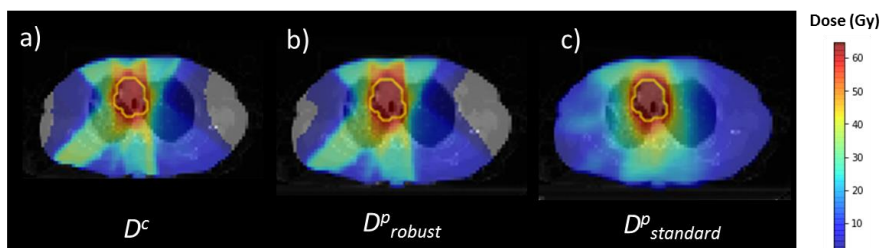


Fig 3. Illustration of an axial slice at the center of the target volume (contoured in yellow) for one of the test patients: a) clinical dose ( $D^c$ ), b) predicted dose from the robust model ( $D^p_{\text{robust}}$ ), and c) predicted dose from the standard model ( $D^p_{\text{standard}}$ ).

## 4 Conclusion

The proposed architecture was able to accurately predict patient specific 3D dose distributions for lung cancer patients. Two different models were investigated, the first one including only anatomical information as input, while the second containing an extra channel with dosimetric information. The two models presented similar accuracy for the high dose region (inside and around the target volume), but the second model outperformed the first one in the prediction of the medium to low dose region. This suggests the need of including dosimetric information as input to the network, in order to build a robust model that is robust to variable beam configuration. This represents an important step towards an easier and more robust implementation of automatic planning techniques, since it reduces the model dependence on consistent dosimetric characteristics through the training patient database.

## References

- [1] D Nguyen, T Long, X Jia, W Lu, X Gu, Z Iqbal, S Jiang , Dose Prediction with U-net: A Feasibility Study for Predicting Dose Distributions from Contours using Deep Learning on Prostate IMRT Patients, arXiv preprint arXiv:1709.09233
- [2] J Fan, J Wang, Z Chen, C Hu, Z Zhang, W Hu, Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique, *Med Phys.* 2018 Nov 1.
- [3] X Chen, K Men, Y Li, J Yi, J Dai, A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning, *Med Phys.* 2018 Oct 26.
- [4] D. Nguyen, X. Jia, D. Sher, M. Lin, Z. Iqbal, H. Liu, S. Jiang, Three-Dimensional Radiotherapy Dose Prediction on Head and Neck Cancer Patients with a Hierarchically Densely Connected U-net Deep Learning Architecture, arXiv preprint arXiv:1805.10397
- [5] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2015.
- [6] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, Densely connected convolutional networks, in *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*.
- [7] F. Milletari, N. Navab, and S.A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*
- [8] W. Lu and M. Chen, Fluence-convolution broad-beam (FCBB) dose calculation, *Phys. Med. Biol.* (2010).