DLMI Challenge
Radiation Dose Prediction for Cancer Patients

Rajae Sebai RAJAE.SEBAI@ENS-PARIS-SACLAY.FR
Imane Elbacha IMANE.ELBACHA@ENS-PARIS-SACLAY.FR
CodaLab team : jey

Abstract

The Radiation Dose Prediction for Cancer Patients project is a competition hosted on Codalab, as part of the medical imaging MVA class, that seeks to predict the radiation dose cancer patients will receive during radiation therapy. The project's primary objective is to develop a model that can accurately predict radiation doses based on CT scans, dose masks and segmentation masks of organs. In this report, we present the models and techniques we applied to achieve an accurate prediction of radiation doses. Our best performant model scored an MAE loss of 0.34 on the test set and used the DCNN model on augmented input data.

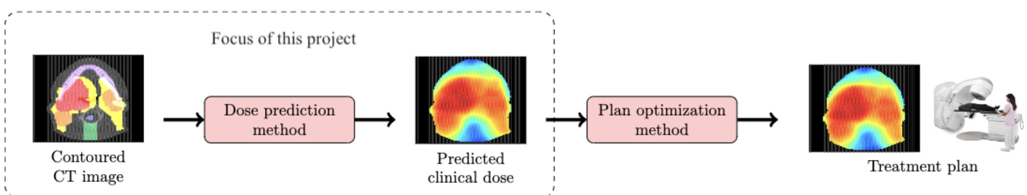


Figure 1: *source:* (A. Babier, 2021)

1. Introduction

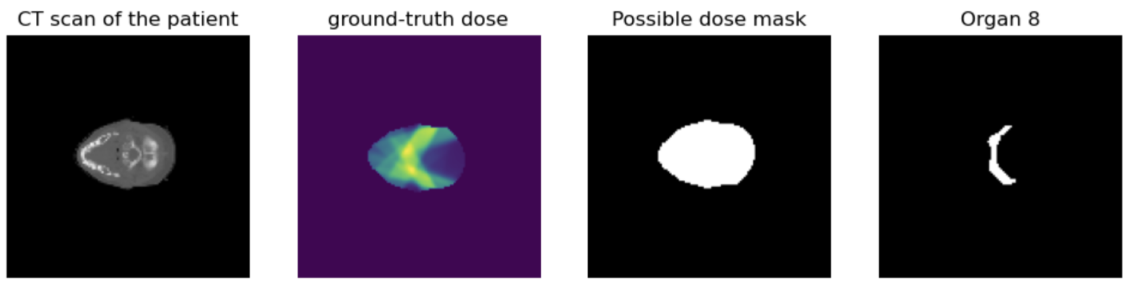
Radiation therapy is a common treatment for cancer patients. Despite its effectiveness, it has the potential to cause harmful side effects. The amount of radiation that a patient receives is a critical factor that can impact the effectiveness of the treatment as well as the patient's health. Therefore, the radiation dose should be precisely calculated and monitored to ensure that it targets cancer cells while minimizing the risk of damaging healthy tissues.

In this project, we aim to explore different models, preprocessing techniques, and data augmentation methods to achieve a fairly accurate prediction of the radiation dose. The models we will be exploring include DCNNs (densely connected network with dilated convolutions) and UNet-like models. We will also be using various preprocessing techniques to enhance the quality and the content of the data and improve the performance of the models.

2. Methodological components

2.1. Dataset

First of all, the dataset used is adapted from the Open-KPB challenge(A. Babier, 2021). The data was pre-processed to obtain 2D data from the 3D scans. We note that the provided 2D data doesn't contain patient informations and the 2D slices are assumed to be totally independant. The provided dataset is a collection of CT scans, dose masks and segmentation masks of organs involved in the treatment such as as the liver, spleen, and kidneys. It includes 7,800 train samples, 1200 validation samples and 1200 test samples. We assume that this repartition was done patient-wise to exclude any data leakage between the different sets. The dataset also includes the targets which are the maximum possible irradiation for each patient.



In this project, we focused on 2D slices, to achieve lower computational complexity and faster processing time. However, it is important to mention that there are some improvement areas in tackling the data. For instance, by performing some feature engineering(Hira, 2021), we can leverage medical research and domain expertise to add potential features such as electron density distance maps from PTV (Planning Target Volume), cone-shaped pseudo beam lines representing beam positions and angles, and radiological depth for the beams. Such additional features may improve the accuracy of the model and provide better predictions of the radiation dose.

2.2. Models

2.2.1. UNet

For this task, as a starting area, we have chosen to use a classic Unet architecture, which has shown promising results in medical image analysis tasks. For context purposes, the Unet architecture is a type of fully convolutional neural network that was originally designed for biomedical image segmentation. It is characterized by a symmetric encoder-decoder structure, where the encoder downsamples the input image to extract high-level features, while the decoder upsamples the feature maps to produce a segmentation mask. The skip connections between the encoder and decoder layers allow for the retention of spatial information and enable the model to better localize features. The UNet architecture is particularly well-suited for the task of predicting radiation doses because it can effectively capture the complex relationships between the CT scans and the corresponding segmentation masks of organs. The use of skip connections also helps to improve the accuracy of the model by allowing it to better distinguish between healthy and cancerous tissues.

2.2.2. Attention UNet (Oktay et al., 2018) & RR Attention UNet (Alom et al., 2018)

In addition to the classic Unet architecture, we also explored the use of Attention UNet and RR Attention UNet for predicting the radiation dose. These models build upon the original Unet architecture by incorporating attention mechanisms that selectively weight the contribution of different spatial locations in the image to the final prediction.

The Attention UNet model achieves this by adding a self-attention block to the encoder and decoder components of the Unet architecture. The self-attention block computes attention maps that highlight the most important features in the input image, which are then used to weight the feature maps of the encoder and decoder. This helps the model focus on the most relevant features while ignoring irrelevant or noisy ones. The RR Attention UNet further enhances this approach by incorporating recurrent feedback connections that allow the model to iteratively refine its predictions based on the attention maps.

2.2.3. DCNN

The choice of the deep convolutional neural network is based on the 'Predicting voxel-level dose distributions for esophageal radiotherapy using densely connected network with dilated convolutions' paper (Zhang et al., 2020) and was used by a 3rd ranking team in the original OpenKPB challenge. In fact, the DCNN model (Figure 4) is a UNet-like architecture in which dilated convolutions are incorporated, which allow for the extraction of features from multiple scales. This is particularly useful in medical imaging applications where structures of interest can vary greatly in size and shape. Additionally, the use of densely connected layers allows for better information flow between layers and helps to prevent overfitting.

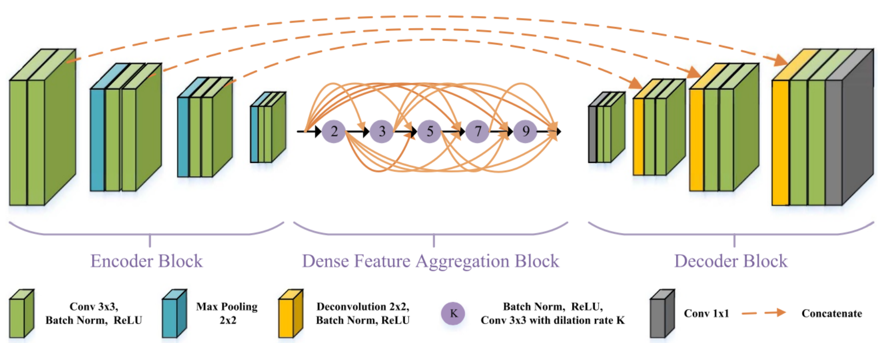


Figure 2: Schematic architecture of the DCNN, comprising an encoder block, a dense feature aggregation block, and a decoder block.(Zhang et al., 2020)

2.3. Experiments

We experimented with various combinations of models and filter settings for both UNets and DCNN models. Ultimately, we selected a configuration of 64 filters that were doubled after each layer for 4 encoder layers for the UNet and a configuration of 32 filters that were doubled after each layer for 4 encoder layers for the DCNN, the dilatation rate of the bottleneck 5 convolution layers were chosen as in 4. We trained the models over 30 or 60 epochs. We used the Adam optimizer with a learning rate of 0.004 along with an StepLR scheduler that decreased the learning rate by 0.1 each 10 epochs. We trained our models on the kaggle GPU T4 x 2, the duration of one epoch is approximately 2.5 minutes. Finally, after obtaining our dose predictions, we multiplied them by the masks to ensure that no radiation was emitted beyond the designated radiation zone. In this section, all models were trained on the entire 12 channels of the input data. As for data augmentation, different set of augmentation were tried. The best performing DCNN model used a random affine transform including random rotation, center translation and scale along with a random 90 degrees rotation (i.e random flips). We're experiencing with CycleGan and Pix2Pix generative models but the trainig procedure is not optimal at the time of writing this report.

Table 1: Experiments

Model	Train Loss	Val Loss	Test Loss
UNet	0.36	0.38	0.56
Attention UNet	0.37	0.39	
RR Attention UNet	0.68	0.45	0.77
DCNN	0.28	0.40	**0.34**

3. Model Tuning and Comparison

3.1. Preprocessing and Tricks

- **Inputs** : In order to optimize the performance of our model, we carried out experiments using different numbers of input channels. Initially, we used only the normalized CT scan as input. Then, we added the masks of the 10 organs involved in the treatment and the dose mask, resulting in a total of 12 channels. Subsequently, we experimented with combining the masks of the 10 organs involved in the treatment into a single channel, which we then superimposed on the CT scan (see Figure 3). This resulted in an input comprising the CT scan, dose mask, and the superimposed organs (3 channels). Interestingly, for some examples, the resulting image distribution appeared to be similar to that of the ground-truth radiation dose. We hoped that this would incorporate an a priori on the target dose of the model and improve its performance by providing additional information about the target but it was not the case according to performance results.

- **Outputs** : Before submitting the testing set radiation doses, we multiply the output of the model by the corresponding dose mask. We hereby assume that we have a prior knowledge on the targeted zone of the radiation.

- **Data augmentation** : We also applied several preprocessing steps to prevent overfitting and enhance the generalization ability of our model. These included data augmentation through combinations of random rotation, flipping, center translation, scaling, cropping and blurring as well as normalizing the pixel values of the CT scans to range between 0 and 1, CT scans being the only channel of the dataset that is not a binary mask. These preprocessing steps helped to prevent overfitting and improve the generalization ability of the model.

- **Validation** : As for the internal validation procedure, the ideal would have been to use a k-fold cross-validation on the joined train and validation sets. We used it for a little number of experiments and stopped due to limited access to GPU resources since it can be computationally expensive, especially when dealing with large datasets such as ours. Instead, we trained our model on augmented, shuffled training data and validated on the original, non-augmented validation set, assuming that the validation set is representative of the challenge test set.

3.2. Ablation Study

To understand the impact of different components of our method, we performed ablation studies on two aspects: model architecture and input channels. The best performing model was the DCNN model applied on the 12 channels (ct scans, 10 organs structure masks, and dose mask), where the 12 channels were augmented using a set of affine augmentations.

A DCNN can be seen as a UNet that employs a sequence of dilated convolutions in its bottleneck. In the first ablation study, we compared the performance of a UNet architecture with that of the DCNN model to assess the importance of the dilated convolutions in DCNN. The second ablation study was focused on the number of channels in the inputs of the model.

The results of our ablation study are summarized in Table 2. As shown in the table, the DCNN model outperformed the UNet architecture. Removing the dilated convolutions in the bottleneck of the DCNN model led to a decrease in performance. Furthermore, training the DCNN model on only CT scans (1 channel) or on CT scans, dose mask and superposed organs (3 channels) resulted in a significant decrease in performance.

Table 2: Ablation Study Results

Model	Inputs	Train Loss	Val Loss	Test Loss
UNet	12 channels	0.36	0.38	0.56
DCNN	12 channels	0.28	0.40	0.34
DCNN	1 channel	0.68	0.82	0.94
DCNN	3 channels	0.42	0.48	0.72
DCNN	12 channels	0.28	0.40	0.34

3.3. Conclusion

In conclusion, our best-performing model used a DCNN with dilated convolutions on the 2D CT-scans, masks of 10 organs, and dose masks of the region where radiation is possible. Our model achieved a score of 0.34 on the test set provided by Codalab.

References

R. Mahmood K.L. Moore T.G. Purdie A.L. McNiven T.C.Y. Chan A. Babier, B. Zhang. Openkbp: The open-access knowledge-based planning grand challenge and dataset, 2021.

Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, 2018.

Sanchit Hira. My 3rd place solution to the openkbp challenge, 2021.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.

Jingjing Zhang, Shuolin Liu, Hui Yan, Teng Li, Ronghu Mao, and Jianfei Liu. Predicting voxel-level dose distributions for esophageal radiotherapy using densely connected network with dilated convolutions. *Physics in Medicine & Biology*, oct 2020.

Appendix A.

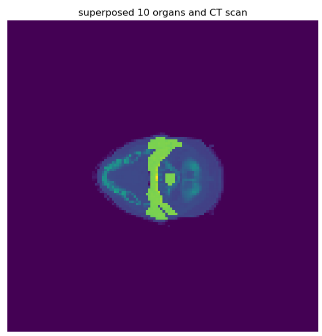


Figure 3: The 10 organs at risks superposed along with the CT scan.

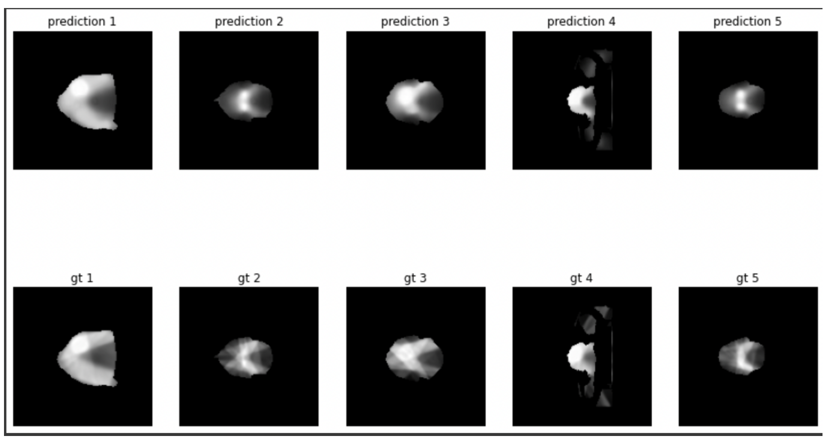


Figure 4: Prediction examples using the DCNN model.