

Evaluation of Transformer Architecture for Deepdose Prediction in Radiotherapy

1/2

Paul RF Dubois^{1,2}, Gizem Temiz³, Nikos Paragios⁴, Paul-Henry Cournède², Pascal Fenoglietto⁵

¹Advanced research, Therapanacea, Paris, France. ²Biomathematics, MICSMICS, CentraleSupélec, Paris, France. ³Clinical Team, Therapanacea, Paris, France. ⁴CEO, TheraPanacea, Paris, France. ⁵Physics, ICM, Montpellier, France

Purpose/Objective:

Radiotherapy treatment planning seeks to maximize tumor control while sparing healthy tissue. Traditionally, treatment plans are developed through iterative, manual processes involving dose constraints. With advances in deep learning, automated dose prediction based on patient scans has gained traction, typically using architectures like 3D U-nets. However, transformers, initially popularized in natural language processing [1], have recently demonstrated robust performance on image [2] and audio [3] data, suggesting potential for application in medical imaging. This study investigates replacing the conventional 3D U-net with a 3D transformer-based architecture for dose prediction, assessing performance and training efficiency.

Material/Methods:

We utilized a dataset of 168 patient CT scans (split 80-10-10 for training-validation-test). We also used the contours for Principal Target Volume (PTV) and Organs at Risk (OARs) used for treatment to compare the Dose-Volume Histograms (DVH). The model input consisted of CT scans and anatomical contours, with the output as the 3D dose prediction. We compared a 3D convolutional U-net and a transformer network, with self-attention layers substituting the convolutional layers of the U-net. The training was performed with a voxel-wise Mean Absolute Error (MAE) loss function on the 3D dose output and the DVH (this second loss helped the model focus on essential regions).

Results:

The transformer model performed comparably to the baseline U-net in dose prediction (see table 1). Both models demonstrated similar MAE and DVH deviation metrics; however, the transformer required significantly more training time due to the computational demands of attention mechanisms applied to 3D data. Transformer training time was approximately twice as long as U-net while showing minimal performance gain over the convolutional approach.

Evaluation of Transformer Architecture for Deepdose Prediction in Radiotherapy

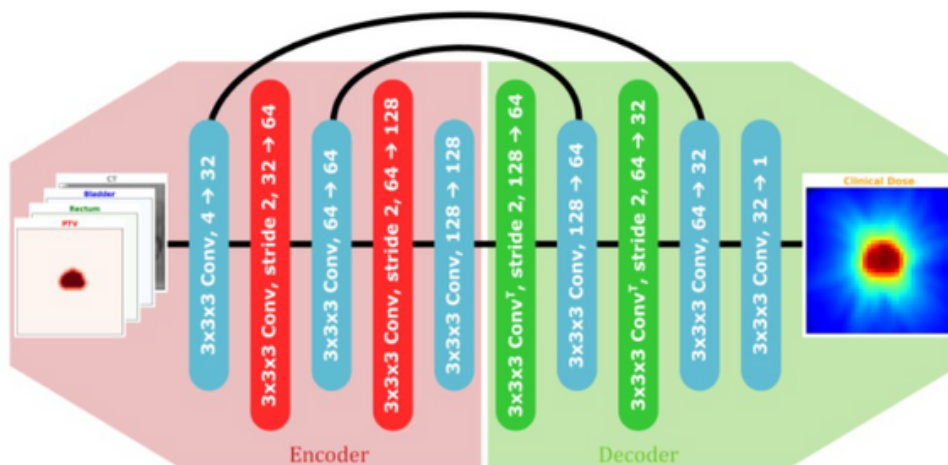


Figure 1: Convolution U-net.

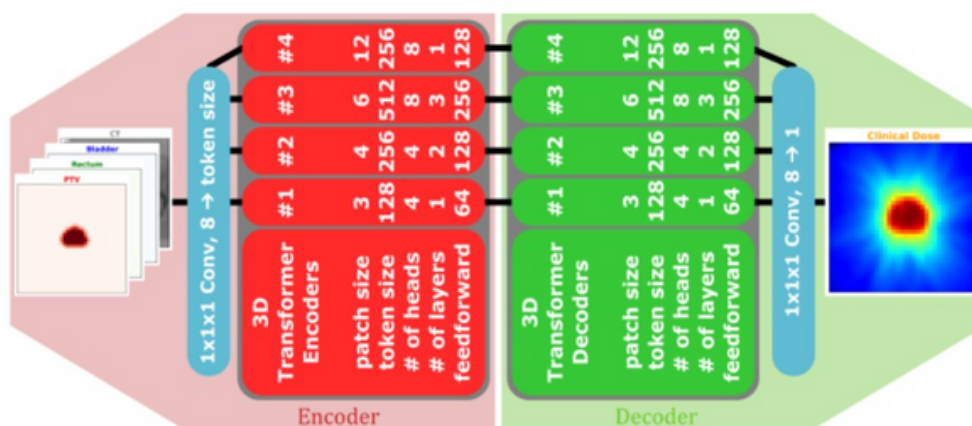


Figure 2: Transformer Network.

Conclusion:

Our findings suggest that while transformer architectures can successfully process 3D medical imaging data, their computational demands may outweigh the benefits for dose prediction in radiotherapy, given minimal performance improvement over U-net-based models. The increased training time for transformers highlights a need for further optimization or hybrid approaches before they can be practically implemented in clinical workflows. However, one potential advantage of transformers is their ability to integrate multiple data modalities. This ability allows additional inputs, such as physicians' prescriptions, directly into the model, enabling more nuanced and customized dose predictions. Future work could explore this multi-modal approach to further leverage transformers' strengths in clinical applications.

Keywords: Deep Dose, Artificial Intelligence, Transformers