THERAPANACEA

Evaluation of transformer architecture for deep dose prediction in radiotherapy

Paul Dubois [a,b,c], Pascal Fenoglietto [b], Paul-Henry Cournède [c], Nikos Paragios [a]
[a] **TheraPanacea** | [b] **Institut du Cancer de Montpellier**
[c] **Biomathematics, MICS, CentraleSupélec**

ESTRO 2025

PURPOSE & OBJECTIVE

Radiotherapy treatment planning seeks to maximize tumour 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 AND 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).

Metric \ Model	3D U-net	3D Transformer
3D dose MAE	3.141 Gy	3.363 Gy
Mean DVH deviation	1.942 Gy	2.052 Gy

Table 1. Performance Comparison of the Models.

RESULTS

The transformer model demonstrated performance metrics that were comparable to those of the baseline 3D U-net in the task of dose prediction, as illustrated in Table 1. Both architectures exhibited similar Mean Absolute Error (MAE) and Dose-Volume Histogram (DVH) deviation metrics, indicating that they achieved comparable accuracy in predicting the 3D dose distribution. However, a notable distinction emerged in terms of computational efficiency. The transformer model required significantly more training time due to the inherent complexity and computational demands of its self-attention mechanisms when applied to 3D medical imaging data. The training for the transformer was approximately twice as long as that for the U-net, despite offering only minimal performance gains over the convolutional approach.

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.

Figure 1. Convolution U-net Architecture,
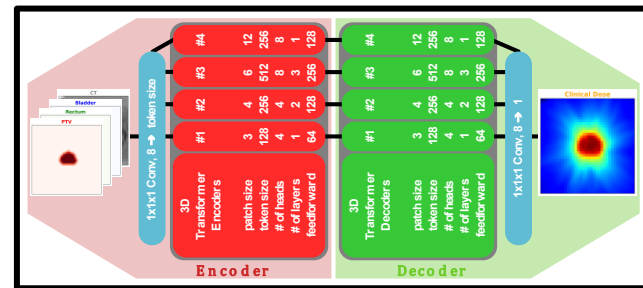
Figure 2. Transformer architecture.

ACKNOWLEDGMENTS

REFERENCES

1. Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.
2. Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A. & Tran, D.. (2018). Image Transformer. Proceedings of Machine Learning Research.
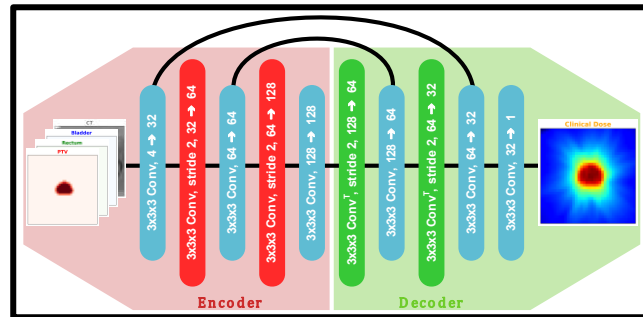3. X. Liu, H. Lu, J. Yuan and X. Li, CAT: Causal Audio Transformer for Audio Classification, International Conference on Acoustics, Speech and Signal Processing.