

Supplementary Material for Latent Spaces Enable Transformer-Based Dose Prediction in Complex Radiotherapy Plans

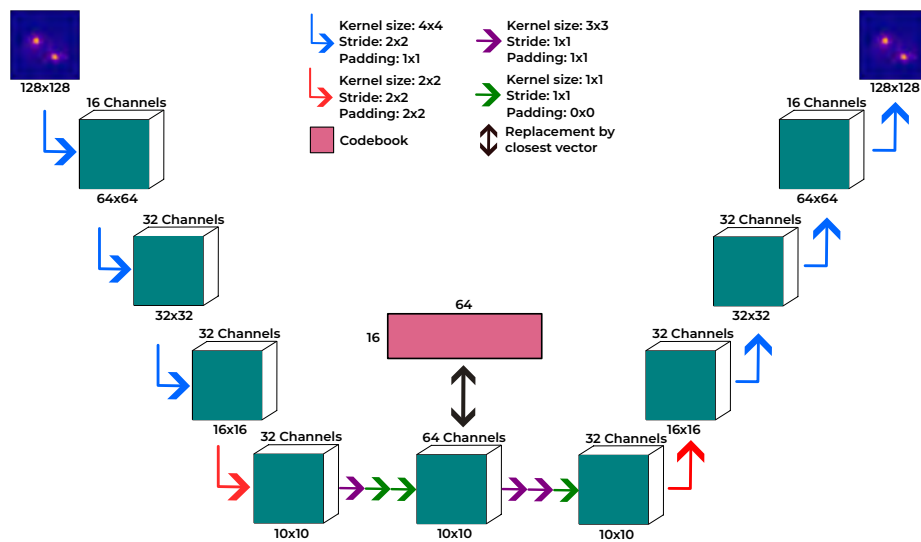


Fig. S1. The architecture of the 2-D vector-quantized variational autoencoders (VQVAE) is shown. Arrows pointing to the right are convolution layers, and arrows pointing up are transpose convolution layers. The codebook is a 2-D matrix of embedding vectors. The 3-D VQVAE has 6 downsampling convolutions instead of 4, with the properties of the last downsampling convolution provided in Table S1.

Table S1. Hyperparameters of the 3-D and 2-D vector-quantized variational autoencoders are shown, along with input and latent space dimensions. The last column displays the searched hyperparameter configurations. DS = Downsampling

	3-D	2-D	HP Search
Input Dimensions	96x128x128	128x128	-
Latent Space Dimensions	3x2x2	10x10	-
Sequence Length	12	100	-
Number of DS Convolutions	6	4	-
Last DS Convolution Kernel Size	1x4x4	2x2	-
Last DS Convolution Stride	1x2x2	2x2	-
Last DS Convolution Padding	0x1x1	2x2	-
Convolution Channels	4,8,16,32,32,32	16,32,32,32	[16, 32, 64]
Number of Embedding Vectors (K)	64	64	[32, 64, 128]
Dimensions of Embedding Vectors (n_z)	16	16	[16, 32]

Table S2. Hyperparameters of transformer model are shown. The last column displays the searched hyperparameter configurations. The vocabulary size is the number of axial slices plus 1 for padding, and block size is equal to sequence length minus 1.

Parameter	Value	HP Search
Number of Layers	4	[2, 3, 4]
Overall Model Dimension	128	-
Number of Attention Heads	2	-
Attention Head Dimension	64	-
Vocabulary Size	97	-
Block Size	370	-

Table S3. The mean absolute difference (MAD) between the predicted and ground truth dose distributions in the test set are reported for LDFormer and the GAN as mean \pm SD (Gray). There was no significant difference in MAD between the two models across any organs at risk or the planning target volumes (PTVs). Gv=Great Vessels, Cw=Chest Wall

Model	Lung	Esophagus	Heart	Airways	Gv	Cw	PTV
LDFormer	3.7 \pm 1.9	2.4 \pm 2.5	1.8 \pm 2.1	2.9 \pm 2.5	2.7 \pm 2.1	2.1 \pm 1.0	33.9 \pm 13.7
GAN	3.2 \pm 1.4	2.2 \pm 2.1	1.7 \pm 1.7	2.3 \pm 1.8	2.7 \pm 2.0	2.0 \pm 0.9	28.5 \pm 9.3

Table S4. Conformality metrics of the test set ground truth doses are shown for all planning target volumes (PTVs) and overlapping PTVs.

Metric	All PTVs	Overlapping PTVs
HI	1.58 \pm 0.35	1.67 \pm 0.32
D _{1cm} (Gy)	113 \pm 52	126 \pm 55
D _{2cm} (Gy)	80 \pm 49	119 \pm 46