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

Edward Wang[1], Ryan Au[1], Pencilla Lang[2], and Sarah A. Mattonen[1]

[1] Western University
{ewang225,rau23,sarah.mattonen}@uwo.ca
[2] London Health Sciences Centre
pencilla.lang@lhsc.on.ca

Abstract. Evidence is accumulating in favour of using stereotactic ablative body radiotherapy (SABR) to treat multiple cancer lesions in the lung. Multi-lesion lung SABR plans are complex and require significant resources to create. In this work, we propose a novel two-stage latent transformer framework (LDFormer) for dose prediction of lung SABR plans with varying numbers of lesions. In the first stage, patient anatomical information and the dose distribution are encoded into a latent space. In the second stage, a transformer learns to predict the dose latent from the anatomical latents. Causal attention is modified to adapt to different numbers of lesions. LDFormer outperforms a state-of-the-art generative adversarial network on dose conformality in and around lesions, and the performance gap widens when considering overlapping lesions. LDFormer generates predictions of 3-D dose distributions in under 30s on consumer hardware, and has the potential to assist physicians with clinical decision making, reduce resource costs, and accelerate treatment planning.

Keywords: Stereotactic Ablative Body Radiotherapy · Transformers · Dose Prediction · Oligometastatic · Deep Learning · Cancer · Lung

1 Introduction

Radiation therapy (RT) is a mainstay of cancer treatment. Approximately 50% of cancer patients worldwide will require RT over the course of their disease, although due to infrastructure and resource constraints, many patients lack access to this effective treatment [1]. The central challenge in RT is delivering sufficient radiation to treat disease while minimizing radiation toxicity. A comprehensive treatment planning and quality assurance pipeline is necessary to facilitate safe and effective treatment [10]. Stereotactic ablative body radiotherapy (SABR) is a treatment involving the delivery of very high and conformal doses of radiation to the tumour that preserves nearby healthy organs at risk (OARs) [15]. SABR is being increasingly used to treat multiple cancer lesions simultaneously, including in the lungs [21, 25]. However, creating a single multi-lesion lung SABR plan is a laborious and time-consuming process taking on average 7.5 hours at our institution. Although there are many variables affecting multi-lesion SABR,

including what dose to deliver to each lesion, or how many lesions to treat, planning resource constraints prevent radiation oncologists (ROs) from comparing different treatment options. The purpose of this study is to create a tool for real-time (<60s) prediction of multi-lesion lung SABR dose distributions, thereby allowing ROs to compare and select the optimal radiation prescription.

Most existing literature on RT dose prediction focuses on single lesions [13, 24, 4]. In 2020, Babier et al. hosted the OpenKBP challenge, and released a dataset of head and neck RT plans that included multiple planning targets [3]. However, to our knowledge, besides our own previous work [27], there has not been any research into the multi-lesion lung domain. Planning multi-lesion lung treatments are challenging due to the heterogeneity in the size, shape, location and number of metastatic lesions. Lesions can be treated with a wide array of prescriptions, and potential interactions between radiation delivered to nearby lesions must be accounted for. Additionally, overlapping lesions may occur due to multiple close lesions, or treatment for recurrence. Existing models commonly use a single channel to provide PTV information [4, 11, 13, 24, 28] and are therefore unable to account for overlapping lesions, limiting their applicability to the multi-lesion lung setting.

Transformers are a family of autoregressive sequence prediction models, originally developed for natural language tasks, that rely on the attention mechanism [26]. Through attention, transformers learn which sections of the input sequence should be more heavily weighted when predicting the next token, allowing them to capture complex relationships in the input data [7, 26]. We hypothesize that this property will allow them to better account for the dose interactions between multiple lesions. To utilize a transformer for dose prediction, spatial image data must be first encoded into sequences, and then decoded back into images. Existing implementations of transformers for dose prediction [12, 28, 11] sandwich the transformer components between encoding and decoding components. The entire network is trained end-to-end, which prevents the network from adapting to variable sequence lengths and therefore varying numbers of lesions.

In this work we develop a novel Latent Dose transFormer (LDFormer) framework for RT dose prediction that operates on latent representations of patient anatomy. LDFormer fully decouples image-to-sequence encoding/decoding from sequence prediction, allowing the model to adapt to multiple lesions. We validate LDFormer on a large collection of multi-lesion lung SABR plans, and compare it to a state-of-the-art (SOTA) generative adversarial network (GAN) [27].

2 Methods

2.1 Data

This study was approved by our institutional ethics review board. The dataset contains treatment plans of patients who were treated with SABR to 2-5 lung lesions (metastases ± primary) from 2010 to 2023 at the London Health Sciences Centre in Ontario, Canada. Patients were excluded if they received non-SABR

thoracic RT prior to or in between SABR treatments. All patients received 4-D CT simulation via a Canon Aquilion LB scanner or a Philips BigBore CT scanner, with motion management performed by free breathing, gating, or deep inspiration breath hold. Treatment planning was performed with Pinnacle3 (Version 9.10, Philips Canada) or RayStation (Version 7.0, RaySearch Laboratories). Each plan consists of the planning CT scan, contours of the OARs, planning target volumes (PTVs) and internal gross tumour volumes (IGTVs), and the delivered dose distribution. An IGTV is the region of space that a lesion moves through during respiratory motion, and is contoured by the treating RO. PTVs are 5 mm expansions of IGTVs, and are the regions that radiation is prescribed to. The OARs (lungs, heart, esophagus, chest wall, great vessels, and airways) were automatically contoured using Limbus Contour (Version 1.7.0, Limbus AI). Multiple plans were collected from a single patient if they received >1 multi-lesion lung SABR treatment. For example, if a patient was initially treated to three lesions, and then received additional treatment to a single lesion (new metastasis or retreatment) one year later, both a three-lesion and four-lesion plan would be collected. To combine serial treatments, the earlier treatment was non-rigidly registered to the later treatment in MIM (Version 7.2.8, MIM Software Inc) based on the planning CT scans. The PTVs and IGTVs were transferred over to the later plan, and the doses of the two treatments were combined. There were 20 dose-fractionation schemes used across all PTVs. To account for the heterogeneity in fractionations, all doses were converted to their equivalent dose in 2 Gray (Gy) fractions (EQD2) using the linear quadratic model ($\frac{\alpha}{\beta}$=3) [17]. Plans were randomly divided into training (∼70%), validation (∼15%) and testing (∼15%) sets, stratified by the number of lesions. Patients with multiple treatment plans were confined to a single set. Plan characteristics are shown in Table 1.

Table 1. Plan Characteristics

	Total	**Training**	**Validation**	**Testing**
Number of Plans	234	171	32	31
Number of Patients	198	157	23	18
Total Number of Lesions	611	406	106	99
2 Lesions	145	129	7	9
3 Lesions	51	27	13	11
4 Lesions	22	8	7	7
5 Lesions	16	7	5	4
Total PTV Vol. (cc) - Median[Range]	43[10-226]	43[11-169]	39[13-152]	52[10-226]
Number of Overlapping PTVs	62	27	19	16
Prescriptions				
60Gy in 8 Fractions	241	172	36	33
55Gy in 5 Fractions	127	97	16	14
35Gy in 5 Fractions	50	21	12	17
54Gy in 3 Fractions	49	44	2	3
30Gy in 5 Fractions	33	17	8	8
Other	108	52	32	24

2.2 Data preprocessing

OARs were voxelized to a single volume with each OAR represented by an integer and each PTV was voxelized to a separate binary volume. Voxelization was performed with the *rt-utils* Python package [23]. Following our previous work [27], we created an initial dose estimate (IDE) based on exponential dose decay [18] to help condition the transformer. All volumes were resampled to 3mm^3 spacing (linear interpolation for dose volumes, and nearest-neighbour interpolation for OAR and PTV volumes) and center cropped to 96x128x128 voxels around the lungs. The 96 voxels are in the superior-inferior direction.

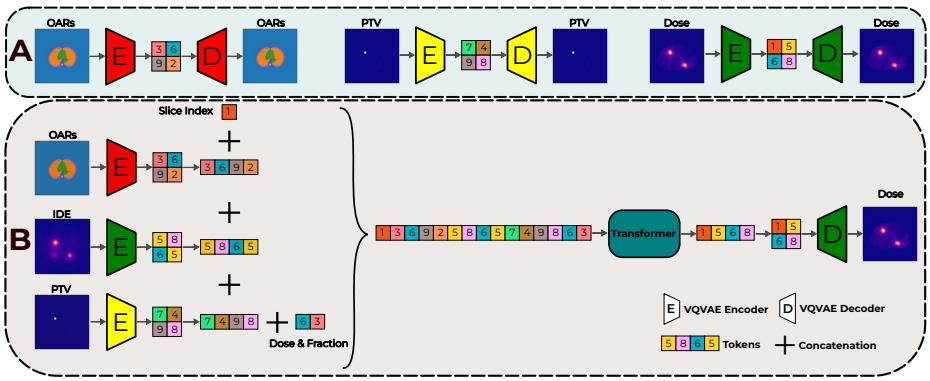

Fig. 1. The overall workflow is shown. **A:** Vector-quantized variational autoencoders (VQVAEs) are trained to encode organs at risk (OARs), planning target volumes (PTVs), and dose into latent representations (LRs). **B:** The transformer is trained to predict the dose LR from LRs of the OARs, initial dose estimate and PTVs concatenated with the slice index and prescription. The dose LR is then decoded into a dose distribution. For simplicity, LRs are depicted as 2x2, and only one PTV is shown.

2.3 Encoding spatial data into sequences

Prior to training the transformer, it was necessary to first encode the volumetric data describing patient anatomy and dose distributions into sequences of integer tokens (Figure 1A). We followed the work in [7, 29] and used a vector-quantized variational autoencoder (VQVAE) for this task. VQVAEs are a variant of autoencoders that map their input into a discrete latent space [19]. A visual representation of VQVAEs is provided in Figure S1. For 3-D spatial data, the encoder E of the VQVAE compresses the input data $x \in \mathbb{R}^{L \times W \times H \times c}$ into a learned latent representation of vectors $z_v \in \mathbb{R}^{l \times w \times h \times n_z}$. Then, in the vector quantization step, the vectors in z_v are replaced by their nearest vectors in a learned codebook $Z \in \mathbb{R}^{K \times n_z}$ to form z_v^q. The decoder D of the VQVAE uses z_v^q to create reconstruction $\hat{x}$. The VQVAE loss function [19] is

$$\mathcal{L}_{VQVAE} = L_{Rec}(x, \hat{x}) + \lambda \|sg[z_v^q] - E(x)\|^2 + \|z_v^q - sg[E(x)]\|^2. \quad (1)$$

L_{Rec} is the reconstruction error between x and $\hat{x}$, and varies per VQVAE. It is computed by mean squared error, binary cross entropy and categorical cross entropy for the dose, PTV and OAR cases respectively. λ is a weighting factor set to 2 in this work. sg represents the stop gradient operator which disables gradient backpropagation [5, 19]. Z is updated via exponential moving average. Further theoretical details about VQVAEs can be found in the original paper [19]. During training of the OAR and dose VQVAEs, the input data was augmented by flipping across the vertical and horizontal axes with 50% probability.

Using the VQVAEs, we encoded the input spatial data into integer sequences s by replacing each vector in z_v^q with its corresponding index in Z, and flattening the result into $s \in \{0, 1, 2, \ldots, K-1\}^N$ where $N = l \times w \times h$. The 2-D VQVAE formulation simply excludes the height dimension. Hyperparameters and shapes of latent representations ($l \times w \times h$) are provided in Table S1. We trained a 3-D VQVAE for the PTV masks, and 2-D VQVAEs for the OAR maps and dose distributions (sliced axially). Encoding PTVs in 3-D allows every sequence to contain information on the location of all PTVs. VQVAEs were used to encode the sequences s_{ptv}, s_{oars}, s_{ide} and s_{dose}. Both s_{dose} and s_{ide} were encoded using the dose VQVAE. s_{ptv} was modified by appending the prescribed dose and fraction to the end, where dose is represented as an index in a lookup table. All sequences were concatenated to form a combined sequence $s_c = \{s_{ax}, s_{oars}, s_{ptv1}, s_{ptv2}, s_{ptv3}, s_{ptv4}, s_{ptv5}, s_{dose}\}$ to feed into the transformer. s_{ax} is the axial index of the 2-D slice. The lengths of s_{oars}, s_{ide} and s_{dose} are 100. The lengths of s_{ptv1-5} are 14. The total length of s_c is 371. We increment all values in s_c by 1 to reserve 0 as the padding token, for empty PTV sequences.

2.4 Sequence prediction with transformers

We adapted the decoder stack from the seminal transformer paper [26] to predict s_{dose} (Figure 1B). The decoder-only transformer generates new tokens in an autoregressive manner, in which the probability of the next token in the dose sequence $s_{\mathrm{dose},i}$ depends on all previous dose tokens $s_{\mathrm{dose}<i}$, as well as conditioning sequences s_{ax}, s_{oars}, s_{ide} and s_{ptv1-5}. The objective is to maximize the likelihood of $p(s_{\mathrm{dose},i})$, and therefore the transformer loss is the negative log-likelihood

$$L_{TF} = -\prod_i \log p(s_{\mathrm{dose},i} | s_{\mathrm{dose}<i}, s_{\mathrm{ax}}, s_{\mathrm{oars}}, s_{\mathrm{ide}}, s_{\mathrm{ptv1\text{-}5}}). \tag{2}$$

L_{TF} is only computed over the positions corresponding to tokens in s_{dose}. To account for input sequences that have varying numbers of PTVs, we extended causal attention masking [26] to also mask out positions of empty PTVs based on the padding token. Token position was encoded sinusoidally [26]. We utilized a transformer with 4 layers, 2 heads and an embedding dimension of 128. The full model configuration is presented in Table S2. During training, we augmented the data by creating sequences from spatial data flipped along the sagittal, coronal and both planes as well as randomly selecting two permutations of PTV ordering in s_c, as PTV order is arbitrary. Greedy sampling was performed by simply

choosing the most likely token for $s_{\text{dose},i}$, therefore allowing for reproducible predictions. We used LDFormer to generate dose sequences for every axial slice, which were decoded to 2-D dose slices using the decoder of the dose VQVAE. Then, all 2-D dose slices were stacked to form a 3-D distribution.

2.5 Implementation details

Models were implemented in PyTorch 2.0.1 (Python 3.9). Hyperparameters were tuned via grid search on the validation set. Models were trained with the AdamW optimizer [14]. VQVAEs were trained with a learning rate of 3×10^{-4}. The learning rate for the transformer consisted of a linear warmup for 200 epochs to 1×10^{-4} followed by cosine decay [22]. The batch sizes for the 3-D VQVAE, 2-D VQVAEs, and the transformer were 16, 512 and 512 respectively. The 3-D VQVAE was trained for 1000 epochs. The 2-D VQVAEs were trained for 5000 epochs. The transformer was trained for 1000 epochs. VQVAEs were trained on a NVIDIA V100 32GB GPU with training times ranging from ∼4 hours to ∼1.5 days. The transformer was trained on a NVIDIA 3090 24GB GPU for ∼1 day. Training and evaluation code is available at https://github.com/edwardwang1/LDFormer.

2.6 Model Evaluation

We evaluated LDFormer on the testing set by comparing the predicted dose to the ground truth dose on the basis of dose-volume-histogram (DVH) metrics of OARs, conformality metrics of the PTVs, and mean absolute difference (MAD) across all structures. All metrics were calculated in EQD2. The DVH metrics are taken from the dose constraints used in the ongoing phase III multi-lesion SABR-SYNC clinical trial [20]. They are the maximum dose to 5 cubic centimeters (D_{5cc}) of the esophagus, chest wall and airways, D_{10cc} of the great vessels, D_{15cc} of the heart, the volume of lung receiving less than 14 Gy (CV_{14}), and the percent of lung receiving above 15 Gy (V_{15}). Both CV_{14} and V_{15} are the EQD2 equivalent of the contraints from [20]. DVH metrics were calculated from the masks of the OARs minus the IGTVs. The conformality metrics used are the heterogeneity index (HI), and the maximum dose at 1 cm and 2 cm away from the PTV (D_{1cm}, D_{2cm}). HI is the ratio of the maximum dose inside the PTV to the prescription dose [13, 9]. As dose is not guaranteed to decrease with increasing distance from the PTV due to other lesions, D_{1cm} and D_{2cm} were calculated from a 1 voxel ($27mm^3$) thick sphere at 1 cm and 2 cm from the PTV. HI, D_{1cm}, and D_{2cm} were calculated for all lesions, as well as only lesions with overlap. Ground truth test set conformality metrics are shown in Table S4. Finally, we calculated the MAD between the predictions and ground truth across all OARs and PTVs. We compared LDFormer to our previous implementation of a GAN as described in [27] on all metrics. Significance testing was performed in Python 3.9 using the T-test for normal data and Wilcoxon rank-sum test for non-normal data, with normality assessed by the Shapiro-Wilk test.

Table 2. The absolute differences in the dose-volume-histogram and conformality metrics between predicted doses and ground truth in the testing set are reported for LDFormer and the GAN as mean±SD. Conformality metrics are calculated over all lesions, as well as only lesions with overlap (Ov). Bold font indicates significantly better performance ($p<0.05$). The unit of CV_{14} is cc. The unit of V_{15} is %. The units of D_{Xcc} and D_{Xcm} are EQD2 Gy with $\frac{\alpha}{\beta} = 3$. HI is dimensionless. Ln=Lung, Es=Esophagus, Hr=Heart, Aw=Airways, Gv=Great Vessels, Cw=Chest Wall.

Model	**$LnCV_{14}$**	**LnV_{15}**	**EsD_{5cc}**	**HrD_{15cc}**	**AwD_{5cc}**	**GvD_{10cc}**	**CwD_{5cc}**
LDFormer	122±120	3.3±4.1	4.5±8.2	5.8±8.8	5.8±7.1	5.3±4.5	22±20
GAN	70±53	1.7±1.5	4.7±6.5	5.3±7.2	3.6±5.1	4.4±4.8	17±22
Model	**HI**	**D_{1cm}**	**D_{2cm}**	**OvHI**	**OvD_{1cm}**	**OvD_{2cm}**	
LDFormer	**0.44±0.40**	**32±28**	22±19	**0.67±0.49**	**44±32**	**39±25**	
GAN	0.74±0.69	53±40	27±28	1.71±0.66	97±35	65±40	

3 Results

Quantitative results are shown in Table 2. Across all PTVs, LDFormer significantly outperformed the GAN on HI and D_{1cm}. There were 16 lesions in the testing set that overlapped with other lesions. For these, LDFormer performed significantly better on all PTV conformality metrics and the magnitude of improvement was greater, as the GAN tended to overestimate dose due to overlap. Across all DVH and MAD metrics, LDFormer performed similarly to the GAN. MAD summary results are provided in Table S3. Figure 2 shows LDFormer and GAN predictions for two testing set plans with overlapping lesions. In the GAN dose, hotspots can be seen inside overlapping lesions which are not present on the LDFormer dose. Minor step artifact is visible on the LDFormer dose, caused by the stacking of 2D predictions. The inference time for a full 3-D dose distribution of a 5-lesion plan, including loading weights, is 8.5s for the GAN, and 28.7s for LDFormer on an NVIDIA 3090 24GB GPU.

4 Discussion

Evidence is growing in favour of treating more metastases [21, 25] with SABR. As more lesions are being treated, scaling treatment complexity and increased planning resource requirements are adding strain on hospital systems. For patients with multiple lesions, there are many permutations of which lesions are treated, and the radiation prescribed to each lesion. Creating a treatment plan involves selecting a prescription for each lesion and performing inverse planning to create a dose distribution. The dose distribution is then assessed according to clinical criteria. Due to the resources required during planning, it is not feasible for ROs to request multiple plans for comparison per patient. Therefore, although a patient is treated with a plan that passes criteria, they may not be treated with the best plan (e.g. the patient received a lower prescription while a higher prescription was possible). A real-time dose prediction tool would allow ROs to quickly compare potential treatments both visually and quantitatively,

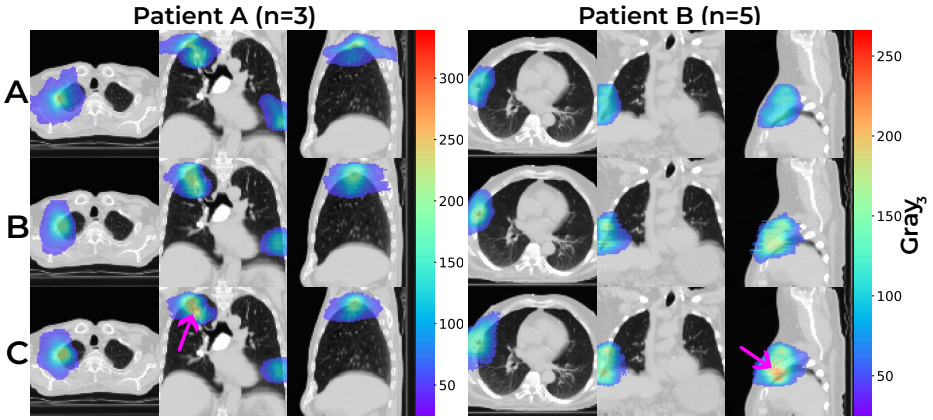

Fig. 2. Axial, coronal, and sagittal views of the (A) ground truth, (B) LDFormer and (C) GAN doses are shown for two testing set patients with overlapping lesions. Arrows indicate hotspots in overlapping lesions. The unit of the colourbar is EQD2 Gy ($\frac{\alpha}{\beta} = 3$).

ensuring that patients receive the optimal treatment. To further accelerate treatment planning, the predicted dose distributions can also be used as optimization targets during inverse planning [2, 8, 16]. In this work, we leverage the powerful modelling capabilities of the transformer [26] for dose prediction of multi-lesion lung SABR plans, and build a novel framework capable of making accurate predictions in under 30s.

LDFormer outperforms our previous GAN approach [27] on PTV conformality metrics, and is competitive with the GAN on DVH metrics. Overlap analysis demonstrates that the greatest advantage of LDFormer is in dose prediction for plans with overlapping PTVs, supporting the hypothesis that attention allows LDFormer to better account for inter-lesion dose interactions. Such cases are most common in the context of retreatment, but may also occur when multiple lesions are grouped close together. It is challenging for ROs to prescribe an appropriate treatment for these patients, and it is in these complex cases that they would benefit most from dose prediction tools like LDFormer.

A limitation of this work is the modest size of the dataset. LDFormer consists of a small transformer operating on heavily compressed latent representations of dose and anatomical data. Additional training data would enable a larger model and longer sequence lengths (i.e. larger, less compressed latent representations), reducing error introduced by the VQVAEs. Further, transformers lack inductive spatial biases [7] present in convolutional neural networks (such as the GAN used for comparison [27]) and therefore require larger datasets for image-based tasks [6]. Another limitation common to dose prediction methods is that the output is not guaranteed to be physically deliverable, even if the model was exclusively trained on real plans. However, the real benefit of these dose prediction techniques is the reduction of required resources, not that the patient's real treatment exactly matches the prediction. An "undeliverable" prediction may

still benefit RO clinical decision making and during inverse planning. To study this, we are preparing a prospective validation series to quantify the resource savings of introducing LDFormer into the clinical workflow at our centre.

5 Conclusion

In this work, we present the LDFormer framework for multi-lesion lung SABR dose prediction. LDFormer outperforms a SOTA GAN in PTV conformality metrics, and is most beneficial for plans with overlapping lesions. Multi-lesion lung SABR is an effective but resource intensive treatment. Our work has the potential to reduce resource burden and increase the adoption of this technique.

Acknowledgments. This project was supported by the Gerald C. Baines Foundation, donor support through the London Health Sciences Foundation, the Keith Samitt Translational Cancer Research Grant, and trainee support from the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank Karen Eddy for assisting with the clinical data curation.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdel-Wahab, M., Gondhowiardjo, S.S., Rosa, A.A., Lievens, Y., El-Haj, N., Polo Rubio, J.A., Prajogi, G.B., Helgadottir, H., Zubizarreta, E., Meghzifene, A., et al.: Global radiotherapy: current status and future directions—white paper. JCO Global Oncology **7**, 827–842 (2021)
2. Babier, A., Mahmood, R., Zhang, B., Alves, V.G., Barragán-Montero, A.M., Beaudry, J., Cardenas, C.E., Chang, Y., Chen, Z., Chun, J., et al.: Openkbp-opt: an international and reproducible evaluation of 76 knowledge-based planning pipelines. Physics in Medicine & Biology **67**(18), 185012 (2022)
3. Babier, A., Zhang, B., Mahmood, R., Moore, K.L., Purdie, T.G., McNiven, A.L., Chan, T.C.Y.: Openkbp: The open-access knowledge-based planning grand challenge and dataset. Medical Physics **48**(9), 5549–5561 (2021)
4. Barragán-Montero, A.M., Nguyen, D., Lu, W., Lin, M.H., Norouzi-Kandalan, R., Geets, X., Sterpin, E., Jiang, S.: Three-dimensional dose prediction for lung imrt patients with deep neural networks: robust learning from heterogeneous beam configurations. Medical physics **46**(8), 3679–3691 (2019)
5. Dhariwal, P., Jun, H., Payne, C., Kim, J.W., Radford, A., Sutskever, I.: Jukebox: A generative model for music (2020)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
7. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12873–12883 (June 2021)

8. Fan, J., Wang, J., Chen, Z., Hu, C., Zhang, Z., Hu, W.: Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Medical physics **46**(1), 370–381 (2019)
9. Feuvret, L., Noël, G., Mazeron, J.J., Bey, P.: Conformity index: a review. International Journal of Radiation Oncology* Biology* Physics **64**(2), 333–342 (2006)
10. Fraass, B., Doppke, K., Hunt, M., Kutcher, G., Starkschall, G., Stern, R., Van Dyke, J.: American association of physicists in medicine radiation therapy committee task group 53: Quality assurance for clinical radiotherapy treatment planning. Medical Physics **25**(10), 1773–1829 (1998)
11. Hu, C., Wang, H., Zhang, W., Xie, Y., Jiao, L., Cui, S.: Trdosepred: A deep learning dose prediction algorithm based on transformers for head and neck cancer radiotherapy. Journal of Applied Clinical Medical Physics **24**(7), e13942 (2023)
12. Jiao, Z., Peng, X., Wang, Y., Xiao, J., Nie, D., Wu, X., Wang, X., Zhou, J., Shen, D.: Transdose: Transformer-based radiotherapy dose prediction from ct images guided by super-pixel-level gcn classification. Medical Image Analysis **89**, 102902 (2023)
13. Kearney, V., Chan, J.W., Wang, T., Perry, A., Descovich, M., Morin, O., Yom, S.S., Solberg, T.D.: Dosegan: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation. Scientific reports **10**(1), 11073 (2020)
14. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
15. Martin, A., Gaya, A.: Stereotactic body radiotherapy: a review. Clinical Oncology **22**(3), 157–172 (2010)
16. McIntosh, C., Welch, M., McNiven, A., Jaffray, D.A., Purdie, T.G.: Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. Physics in Medicine & Biology **62**(15), 5926 (2017)
17. McMahon, S.J.: The linear quadratic model: usage, interpretation and challenges. Physics in Medicine and Biology **64**(1), 01TR01 (Dec 2018)
18. Narayanasamy, G., Desai, D., Maraboyina, S., Penagaricano, J., Zwicker, R., Johnson, E.L.: A dose falloff gradient study in rapidarc planning of lung stereotactic body radiation therapy. Journal of Medical Physics **43**(3), 147 (2018)
19. van den Oord, A., Vinyals, O., Kavukcuoglu, K.: Neural discrete representation learning. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)
20. Palma, D.: A randomized phase III trial of stereotactic ablative radiotherapy for patients with up to 10 oligometastases and a synchronous primary tumor. (SABR-SYNC), Identifier: NCT05717166. Phase III. Status: Recruiting
21. Palma, D.A., Olson, R., Harrow, S., Gaede, S., Louie, A.V., Haasbeek, C., Mulroy, L., Lock, M., Rodrigues, G.B., Yaremko, B.P., et al.: Stereotactic ablative radiotherapy versus standard of care palliative treatment in patients with oligometastatic cancers (sabr-comet): a randomised, phase 2, open-label trial. The Lancet **393**(10185), 2051–2058 (2019)
22. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al.: Improving language understanding by generative pre-training (2018)
23. Shrestha, A., Watkins, A., Uribe, C.: RT-Utils: A minimal Python library for RT Struct manipulation. https://github.com/qurit/rt-utils (2020)

24. Song, Y., Hu, J., Liu, Y., Hu, H., Huang, Y., Bai, S., Yi, Z.: Dose prediction using a deep neural network for accelerated planning of rectal cancer radiotherapy. Radiotherapy and Oncology **149**, 111–116 (2020)
25. Tsai, C., Yang, J., Guttmann, D., Shaverdian, N., Eng, J., Yeh, R., Girshman, J., Das, J., Gelblum, D., Xu, A., et al.: Final analysis of consolidative use of radiotherapy to block (curb) oligoprogression trial-a randomized study of stereotactic body radiotherapy for oligoprogressive metastatic lung and breast cancers. International Journal of Radiation Oncology* Biology* Physics **114**(5), 1061 (2022)
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)
27. Wang, E., Snir, J., Chong, J., Mattonen, S.A., Lang, P.: Predicting the dose distribution of multi-lesion lung stereotactic ablative radiotherapy plans using generative adversarial networks. In: Yu, L., Fahrig, R., Sabol, J.M. (eds.) Medical Imaging 2023: Physics of Medical Imaging. vol. 12463, p. 124630N. International Society for Optics and Photonics, SPIE (2023)
28. Wen, L., Xiao, J., Tan, S., Wu, X., Zhou, J., Peng, X., Wang, Y.: A transformer-embedded multi-task model for dose distribution prediction. International Journal of Neural Systems **33**(08), 2350043 (2023)
29. Yan, W., Zhang, Y., Abbeel, P., Srinivas, A.: Videogpt: Video generation using vq-vae and transformers (2021)