Rapid Auto-Planning for Cervical Brachytherapy: Direct Dwell Time vs. Dose Prediction with Deep Learning Networks

Lance Moore[1], Karoline Kallis[1], Kelly Kisling[1], Nuno Vasconcelos[2] and Sandra Meyers[1]

[1]Department of Radiation Medicine and Applied Sciences, University of California, San Diego, La Jolla, USA
[2]Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, USA

Abstract Treatment planning for cervical brachytherapy is often performed manually and can take over an hour. Automated planning leveraging deep learning models could significantly reduce planning time, improving the patient experience. The goal of this work was to assess a new approach to automated planning: using a neural network to directly predict dwell times (VGG model). This model was compared to an automated planning method that uses a neural network to predict 3D dose (OPT model), followed by an optimizer to convert dose to dwell times. Models were developed using 1803 prior cervical brachytherapy treatment plans, with a 68%/16%/16% train/test/validation split. The dataset included 7 different types of implants, featuring some combination of tandem, ovoids, ring, and/or needles. The VGG model used a 16-layer 3D VGG regression network, while the OPT model employed a Cascade U-Net. Automated plans were compared to clinical plans using the mean absolute error (MAE) and mean error (ME) in dose and dwell times. The resulting automated plans were very similar to clinical plans for both models, as evidenced by low MAE (OPT=3.4%, VGG=4.7%) and ME in dose (OPT=0.1%, VGG=0.6%), and low MAE (OPT=7.3s, VGG=8.0s) and ME (OPT=0.2s, VGG=0.7s) in dwell times. While the OPT model featured slightly superior performance, the direct dwell time prediction of the VGG model reduced computation time. Both methods of automated planning could provide high clinical value for brachytherapy procedures.

1 Introduction

Brachytherapy is an important component of the standard of care for cervical cancer. Treatments are customized to patient anatomy on imaging, but this treatment planning is often performed manually and can take over an hour [1]. Automated treatment planning could significantly reduce treatment time, decrease the use of hospital resources, standardize treatment, and improve patients' experience.

Automated, knowledge-based planning uses models trained on prior patient data to generate an optimal treatment plan for a new patient's anatomy. Three solutions have been proposed for knowledge-based brachytherapy planning of cervical cancer [2–4]. Two groups have trained deep reinforcement learning models to autonomously adjust parameters such as inverse optimization objectives and weights [3] or dwell times [2]. These methods rely on objective functions that incorporate dose metrics for target and organ contours. However, cervical brachytherapy often delivers prescription dose to additional regions that are not contoured, such as the uterus and parametrium, and therefore this dosing could not be included in such an objective function.

An alternative is automated planning based on 3D dose prediction from a 3D U-Net [4,5]. One group has shown that the 3D dose can be predicted using 3D inputs of standard anatomical contours and applicator geometry for tandem-and-ovoids (TO) [5]. An optimizer can then iteratively determine the dwell times for each dwell position that replicate this 3D predicted dose, generating an automated treatment plan in a few minutes [4]. These U-Nets can learn common dose shapes, avoiding the need for additional target contours beyond the high-risk CTV (HRCTV).

The above method is limited by the accuracy of the CNN, which may predict physically impossible dose distributions. Cortes at al reported reduced accuracy in high dose regions near applicators [5]. Further, it requires a 3D predicted dose intermediate and a separate optimizer, which increases computation time. The purpose of this work was to develop a deep learning model for direct dwell time prediction. Automated plans produced using these two methods were compared to clinical plans to assess performance.

2 Materials and Methods

Deep learning models were developed using 1803 prior high dose-rate brachytherapy treatment plans for cervical cancer for 475 patients (from 2009-2023). The train/test/validation split was 68%/16%/16%. The approximate applicator breakdown in each dataset was 40% TO, 6% TO with 4-10 needles (TON), 14% tandem-and-ring (TR), 16% TR with 3-6 needles (TRN), 3% tandem only (T), 18% T and 3-21 needles (TN) and 2% pure interstitial with 3-18 needles (N). Two automated planning techniques based on deep learning models are described below.

Model inputs

Planning CT images and DICOM-RT files were exported from the treatment planning system and used to generate model inputs. Binary contour masks of HRCTV, vaginal HRCTV (if present), bladder, rectum, and sigmoid were extracted from DICOM structure sets. Applicator masks were produced by extracting dwell position coordinates from the plan DICOM file and setting a 3x3x3 voxel box centered on each dwell position to 1. Glowing masks, which encode the distance of each image voxel to a structure with an inverse square fall-off, were also produced. Voxels inside a structure were set to 1, then for each image voxel, the summed total inverse squared distance from that voxel to every structure voxel was calculated. The intensity of voxels outside the structure were normalized by the maximum value outside, and voxels inside the structure were set to 1. Glowing masks were produced for anatomical contours and the applicator mask.

A dose-rate kernel was produced by exporting the 3D dose matrix for a single dwell position from the treatment planning system and dividing by the dwell time. This kernel was mapped to each dwell position in each applicator to produce 3D single-dwell position dose-rate kernels, which were used as an input for one model. Four applicator dose kernel inputs were produced by summing

dose-rate kernels for all dwell positions in a given applicator type (tandem, ring, ovoids and needles) with uniform dwell time weighting.

Dose prediction and optimizer model

This model (termed "OPT") used a 3D Cascade U-Net [6] with a mean square error (MSE) loss function to predict the 3D dose. The inputs included 5 binary contour masks, a binary applicator mask, and 4 applicator dose kernels. The model consisted of 2 in-series U-Nets, each containing 5 encoding layers and 5 decoding layers (Figure 1A). Each encoding layer contained 2 in-series depth-wise separable convolution blocks. Each block contained a 3D grouped convolution followed by an instance normalization, ReLU activation, a 3D point-wise convolution, a second instance normalization, and a second ReLU activation function. A 3D maxpool operation performed at the end of each encoder block served to reduce the feature map dimensions. In the decoder layers, the convolution blocks were preceded by a transposed convolution to increase the dimensions of the feature maps prior to cross concatenation with the feature maps from the corresponding encoder layer. At the termination of each U-Net, two additional convolution blocks were appended to reduce the number of feature maps and generate the final 3D dose prediction. To employ the cascade mechanism for the network, the penultimate feature maps from the first U-Net were concatenated with the inputs to the second U-Net prior to feeding the images to the model.

Model dose predictions were then converted into deliverable treatment plans using a COBYLA optimizer, which minimized the MSE between the 3D dose calculated for a given set of dwell times and the predicted dose [4].

VGG model

This model (termed "VGG") used a 16-layer 3D VGG [7] regression network to directly predict dwell times. Model inputs included the planning CT image, glowing contour and applicator masks, and single-dwell position dose-rate kernels. Due to memory limitations, a maximum of 200 dwell position kernels could be included, which meant that 4 plans were excluded. If a treatment plan lacked a contour or 1 of the 200 dwell positions, that channel was left blank.

The model followed a standard VGG archetype, with sequential convolutions in the encoder portion of the network and a multi-layer perception (MLP) decoder (Figure 1B). In the encoder, each convolution block consisted of a standard 3D convolution followed by a 3D batch norm and a ReLU activation. A 3D max pool was used after the second, fourth, seventh, tenth, and thirteenth convolution blocks to reduce the feature maps prior to flattening. Before the MLP, a length 200 dwell position token vector was concatenated to the flattened feature vector. This vector contained a one-hot encoding of the number of dwell positions in the plan. This vector injects plan information deep into the network to assist the model in predicting dwell times only for valid dwell positions. The MLP in the decoder consisted of three layers, each followed by a ReLU activation function and a dropout layer with probability 0.1. The decoder produced a vector of size 200, i.e. dwell times for the 200 possible dwell positions. The absolute value of the predictions ensured that no negative dwell times were predicted.

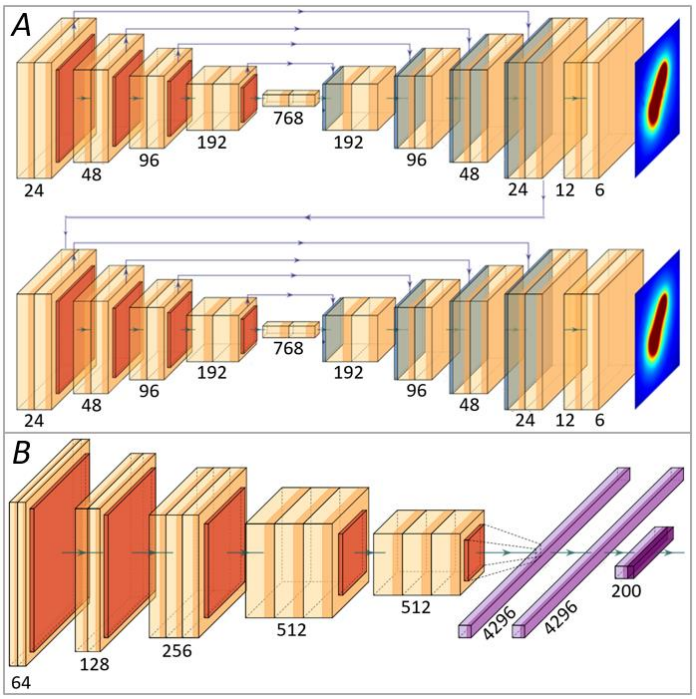

Figure 1. Model diagrams for (A) Cascade U-Net and (B) VGG network. Yellow blocks indicate a convolution block, orange indicates a maxpool, blue blocks indicate a transposed convolution, and purple indicates a fully connected layer. The numbers below the blocks depict the number of filters or channels.

Each predicted dwell time was multiplied by the corresponding dwell position dose-rate kernel and then the 3D predicted dose was computed by summing over all dwell positions. The loss function for this model compared the 3D predicted dose to the 3D clinical plan dose. A 3D squared error matrix (SE) was computed between the predicted ($Dose_{pred}$) and clinical ($Dose_{clin}$) distributions using equation (1) below.

$$SE = (Dose_{clin} - Dose_{pred})^2 \qquad (1)$$

The final MSE loss function was computed from the voxel-wise mean of the SE matrix, weighted by each of the 5 glowing contour masks ($Glow\_Mask_i$), per equation (2).

$$MSE = mean(SE + \sum_{i=1}^{5} SE \cdot Glow\_Mask_i) \qquad (2)$$

Training methods and hyperparameters

Data augmentation was employed during model training to delay overfitting. Both networks used an AdamW optimizer, a weight decay of 1e-4, and a one-cycle learning rate scheduler to anneal the learning rate over the course of training (peak learning rate 1e-3). Models were implemented using the PyTorch framework and trained for 500 epochs on a single GPU node with four 32GB NVIDIA V100 SMX2 GPUs. The validation set performance was

checked at the end of each epoch and the best performing checkpoint was selected as the final model.

Comparing model performance

Automated plans were compared between the models using the following metrics: the voxel-based mean error (ME, representing bias) between the 3D auto-plan and clinical dose, the voxel-based mean absolute error (MAE) in dose, and the ME and MAE between auto-plan and clinical plan dwell times. The ME in clinical metrics (HRCTV D90 and D2cc of organs) were also compared. A positive ME indicates the clinical dose or dwell times were higher than the auto-plan.

3 Results

Figure 2 shows example dose distributions and dwell times for the two automated plans compared to the clinical plan for three patients. While there were some differences in dwell times, the general trends in the dwell time distributions were similar, and the resulting dose distributions were nearly indistinguishable.

Box plots of MAE in dose over all voxels and dwell times, broken down by model and applicator type, are displayed in Figure 3. Averaged over all plans, the MAE and ME in dose were 23.2 cGy (3.4%) and 0.8 cGy (0.1%) for the OPT plans, and 31.6 cGy (4.7%) and 4 cGy (0.6%) for the VGG plans. MAE and ME in dwell times were 7.3 s (1.4% of the total plan dwell time) and 0.2 s (0.0%) for the OPT plans, and 8 s (1.6%) and 0.7 s (0.1%) for VGG plans.

Figure 4 shows boxplots of ME in clinical metrics for each model, all plans combined. For the OPT model, rectum D_{2cc} showed the lowest ME of -0.2%, while HRCTV D90 had the largest ME of 2.8%. The VGG model showed the best performance for bladder, with a ME of -0.7%, and the largest ME for HRCTV D90 (6.7%).

4 Discussion

In this work, we demonstrate the value of two unique, deep learning-based methods for automated planning of cervical brachytherapy. These models were trained and validated on a dataset of over 1800 treatment plans with 7 different applicator implant types. To our knowledge, this is one of the largest datasets that has been used for dose prediction or automated planning with deep learning in radiotherapy, and the greatest variety of brachytherapy implants that have been included in any single brachytherapy deep learning model thus far. Both methods produced treatment plans that were similar to clinical plans, as evidenced by the near-zero mean errors and low mean absolute errors in dose and dwell times. We believe that these facts qualify this work as a substantial and significant step forward in the field of brachytherapy. Despite the variability in the data due to the numerous applicator types included, performance was quite similar between applicator types, demonstrating that a single model was adequate.

The performance of the OPT model was slightly superior for most metrics. The tradeoff for these superior results is the time required for the optimizer, which took 30 s on average and up to 3.2 minutes (for plans with the most dwell times) on a computer with an Intel(R) Core(TM) i7-8700 3.2GHz processor. In contrast, the VGG model can directly predict dwell times in a matter of seconds, which is very valuable for real-time planning in the clinic. To our knowledge, this is the first deep learning model to directly predict dwell times for automated planning. Another study used a CNN to predict dwell positions and times for plan verification [10], but used the clinical 3D dose as input.

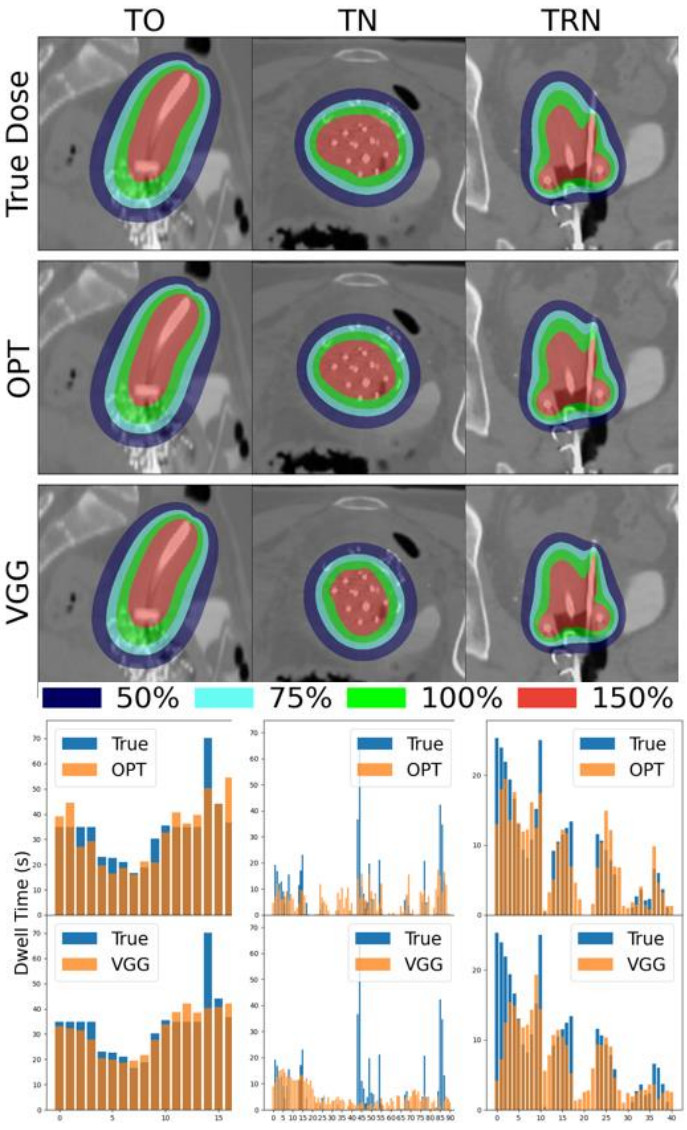

Figure 2. Comparison of dose distributions and dwell times for the two automated plans and clinical plan for three random patients. Isodose color washes (where 100% is the prescription dose) are overlaid over coronal (TRN), axial (TN), and sagittal (TO) CT images.

The VGG model directly incorporates a dose rate kernel exported from the treatment planning system. This kernel is used to compute a physically realistic dose from predicted dwell times for known dwell positions, which is compared to the clinical dose in the loss function. Thus, physical constraints for dose delivery are forced within the VGG model itself. In contrast, the OPT model was not

bound by physics and predicted dose distributions that featured larger errors at high dose. However, it appears that the optimizer was able to correct for most of those errors.

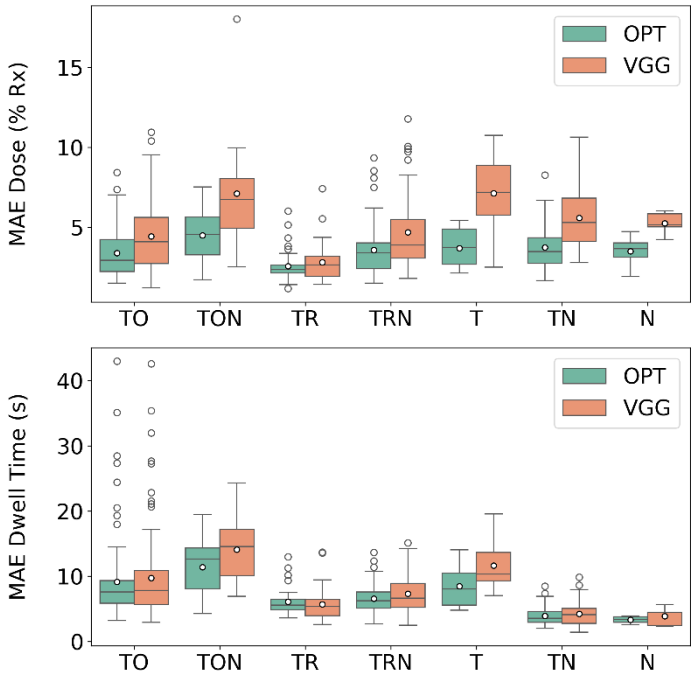

Figure 3. MAE in dose over all voxels (relative to prescription dose) and dwell times, comparing automated plans to clinical plans for each model and applicator. White circles overlaid on boxplots indicate mean values.

One limitation of the VGG method is that it can only support plans with less than 200 dwell positions. We expect that this model could improve further with a few modifications, which will be explored in future work. Currently, the model loads dwell position images into sequential channels, meaning the first dwell position channel is always loaded with the first dwell position in a plan. Shuffling these dwell positions could avoid over-emphasizing the first channels. We employed a dose-based loss function to account for the fact that slight variations in dwell times could result in approximately the same dose. However, we plan to try incorporating dwell time error in the loss function in the future.

In future work, we will continue to refine these models in pursuit of yet more accurate results, and test more sophisticated models. While the OPT model utilized only binary contour and applicator inputs, we found VGG model performance improved when incorporating glowing masks. Future work will explore the use of glowing masks in the OPT model as well. Blinded physician evaluation will be used to evaluate clinical acceptability of automated plans.

5 Conclusion

Two deep learning models were trained to produce automated treatment plans for cervical brachytherapy. The best-performing model predicted the 3D dose as an intermediate step, followed by an optimizer to convert the dose to a deliverable treatment plan. The second model directly predicted dwell times in a matter of seconds, reducing computation time. Both methods produced automated plans that were comparable to clinical plans, and thus could provide great clinical value by significantly reducing planning time.

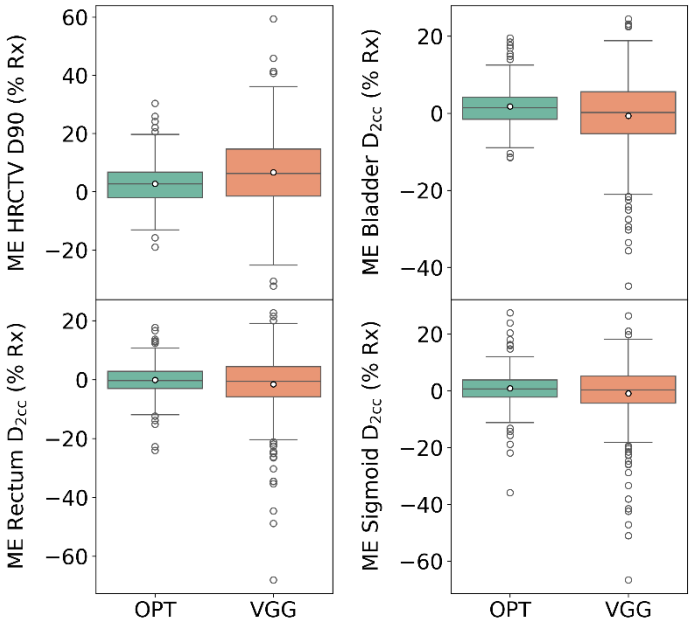

Figure 4. ME in clinical metrics for each model as a percent of the prescription dose, where a positive value indicates the clinical plan metric was higher than the automated plan. White circles overlaid on boxplots indicate mean values.

References

[1] J. Mayadev, L. Qi, S. Lentz, S. Benedict, J. Courquin, S. Dieterich, M. Mathai, R. Stern, R. Valicenti, Implant time and process efficiency for CT-guided high-dose-rate brachytherapy for cervical cancer, Brachytherapy 13 (2014) 233–239. https://doi.org/10.1016/j.brachy.2014.01.004.

[2] G. Pu, S. Jiang, Z. Yang, Y. Hu, Z. Liu, Deep reinforcement learning for treatment planning in high-dose-rate cervical brachytherapy, Physica Medica 94 (2022) 1–7. https://doi.org/10.1016/j.ejmp.2021.12.009.

[3] C. Shen, Y. Gonzalez, P. Klages, N. Qin, H. Jung, L. Chen, D. Nguyen, S.B. Jiang, X. Jia, Intelligent inverse treatment planning via deep reinforcement learning, a proof-of-principle study in high dose-rate brachytherapy for cervical cancer, Phys Med Biol 64 (2019) 115013. https://doi.org/10.1088/1361-6560/ab18bf.

[4] K. Kallis, L.C. Moore, K.G. Cortes, D. Brown, J. Mayadev, K.L. Moore, S.M. Meyers, Automated treatment planning framework for brachytherapy of cervical cancer using 3D dose predictions, Phys Med Biol 68 (2023). https://doi.org/10.1088/1361-6560/acc37c.

[5] K.G. Cortes, K. Kallis, A. Simon, J. Mayadev, S.M. Meyers, K.L. Moore, Knowledge-based three-dimensional dose prediction for tandem-and-ovoid brachytherapy, Brachytherapy 21 (2022) 532–542. https://doi.org/10.1016/j.brachy.2022.03.002.

[6] S. Liu, J. Zhang, T. Li, H. Yan, J. Liu, Technical Note: A cascade 3D U-Net for dose prediction in radiotherapy, Med Phys 48 (2021) 5574–5582. https://doi.org/10.1002/mp.15034.

[7] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, (2015). https://doi.org/10.48550/arXiv.1409.1556.

[8] J. Fan, L. Xing, Y. Yang, Independent verification of brachytherapy treatment plan by using deep learning inference modeling, Phys Med Biol 66 (2021). https://doi.org/10.1088/1361-6560/ac067f.