Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quality Assurance of Brachytherapy Treatment Planning Based on Template, Classification, and Segmentation Methods

Tianlong ji
Northeastern University
Wenjun Tan
Northeastern University
Dazhe zhao
DazheZNEU@126.com

Northeastern University

Research Article

Keywords: brachytherapy quality assurance, deep learning, applicator segmentation, template registration, progressive weighted Kabsch algorithm

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10916871/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background and purpose

Accurate applicator reconstruction is essential for dose calculation in CT-based brachytherapy. Reconstruction errors—driven by limited CT slice thickness and metal artifacts—can produce clinically relevant dose deviations. We developed a fully automated quality assurance (QA) pipeline that combines deep learning classification and segmentation with high-resolution template registration to verify the geometric accuracy of planned source dwell positions.

Materials and methods

We retrospectively collected 3,702 treatment plans from 742 gynecological cancer patients. A lightweight 3D convolutional neural network classified applicators into five categories, and an nnU-Net (v2) model segmented the three channels of Fletcher applicators. High-resolution (0.6-mm) template centerlines were registered to 3-mm clinical CT images using a progressive weighted arc-segment Kabsch algorithm. Geometric consistency was quantified with point-to-curve Euclidean distances. Error-detection performance was assessed in a simulation study in which one randomly selected dwell position per plan was displaced by 1, 3, or 5 mm (50 plans per magnitude), with 50 error-free plans as controls.

Results

The classification model achieved 99.28% accuracy and a weighted F1 score of 0.99. The segmentation model yielded a mean Dice coefficient of 90.7% and a 95% Hausdorff distance of approximately 1 mm. Mean distances between auto-segmented centerlines and the registered template curve ranged from 0.39 to 0.70 mm; mean dwell-point-to-template distances were 0.34–0.51 mm. In the simulation study, specificity was 94.0%; sensitivity was 36.0% for 1-mm displacements and 100% for both 3-mm and 5-mm displacements.

Conclusion

The proposed pipeline enables automated, high-precision QA of brachytherapy dwell positions and reliably detects displacements of 3 mm or larger, providing an objective tool for clinical treatment-plan verification that mitigates thick-slice artifacts through high-resolution template registration.

Highlights

Automated four-module QA pipeline for Fletcher brachytherapy applicator verification.

3D-CNN classifier achieves > 99% accuracy; nnU-Net segmentation yields ~ 91% Dice.

High-resolution template registration suppresses thick-slice CT artifacts.

Submillimetre dwell-position accuracy with 100% sensitivity for ≥ 3 mm errors.

1. Introduction

Radiotherapy is an effective treatment for malignant tumors, and brachytherapy (BT) is an essential component of definitive care.[1–3] BT is particularly valuable in resource-limited settings because it delivers high therapeutic dose with relatively low infrastructure costs. [4, 5] For definitive cervical cancer treatment, BT escalates the central tumor dose while sparing adjacent organs at risk, such as the bladder, rectum, small bowel, and sigmoid colon, thereby improving local control and survival outcomes. [6, 7] Compared with two-dimensional planning, image-guided brachytherapy (IGBT) based on three-dimensional imaging (MRI or CT) has been demonstrated to improve local control and reduce grade 3–4 toxicity rates.[8–10] Current evidence indicates that intensity-modulated radiotherapy (IMRT) and stereotactic body radiotherapy (SBRT) cannot fully replace BT in this context.[11]

Effective quality assurance (QA) is a prerequisite for achieving the expected clinical efficacy of BT.[12] Modern 3D BT is time-critical: applicator insertion, imaging, planning, and delivery must often be completed within a single session lasting under an hour. This complex workflow demands that every step be safe and controllable. Nevertheless, internal audits have revealed that errors may originate from target and organ-at-risk delineation, applicator insertion, applicator reconstruction, and applicator deformation.[13] Existing QA measures rely heavily on manual safety checklists [14, 15] and technical training.[16–18] which are subjective and do not support automated verification.

In CT-based 3D BT planning, applicator reconstruction—defining the 3D centerline of the source channel-is a critical determinant of dose-calculation accuracy. Reconstruction errors can be classified as systematic and random: systematic errors arise primarily from CT slice thickness and metal artifacts, whereas random errors are associated with applicator visibility and operator-dependent subjective judgment. Tanderup et al. demonstrated that the rectum and bladder are the most sensitive organs to reconstruction errors, with mean dose-volume histogram (DVH) parameter shifts of 5%-6% per millimeter of applicator reconstruction error in the anterior-posterior direction.[19] Although artificial intelligence (AI)-based auto-segmentation of organs and applicator contours has emerged as an active research area,[20–22] existing studies focus predominantly on contour accuracy itself, and a systematic solution for verifying the positional accuracy of planned dwell points against reconstructed geometries remains lacking.

We therefore built an automated QA framework that integrates deep learning, a high-resolution applicator template library, and geometric registration. The framework comprises four sequentially connected modules: an applicator classification module, a Fletcher applicator auto-segmentation module, a template library registration and mapping module, and a treatment plan evaluation module.

The objective is to enable automated verification of the positional accuracy of radioactive source dwell points.

2. Methods

2.1 Patient Data and Clinical Dataset

We retrospectively collected 3,702 treatment plans from 742 gynecological cancer patients treated with BT. The cohort comprised 411 cervical, 48 vaginal, and 283 endometrial cancer cases. Applicator usage was distributed as follows: tandem alone (1,632), tandem with a single ovoid (28), Fletcher applicator (three channels, 898), interstitial needle combinations (375), and vaginal applicators (769). The median patient age was 63 years (range: 48–70 years). All patients were treated with an 18-channel Nucletron afterloader (Nucletron, an Elekta company, Veenendaal, The Netherlands). CT images were acquired on a Philips MX 16 Slice scanner (Philips Healthcare, Best, The Netherlands); a 5-mm slice thickness was used for vaginal applicators, whereas a 3-mm thickness was used for all others (120 kV, 320 mAs). All plans were designed and exported from Oncentra MasterPlan (OMP) v4.3, and the dataset included CT series and corresponding RT Plan files.

This retrospective study was approved by the Institutional Review Board of Institution name blinded for peer review (protocol no. [20260712]). Given the retrospective design and the use of fully anonymized data, the requirement for written informed consent was waived by the review board. All CT images and treatment plan files were de-identified before analysis.

2.2 Overview of the QA Framework

The proposed QA framework for BT treatment planning employs a multi-module sequential architecture to automate the verification of radioactive source dwell positions for Fletcher three-channel applicators (Figure. 1). The workflow is as follows: (1) Applicator classification module: a lightweight 3D-CNN operating on clinical CT (3-mm slice thickness) classifies applicators into five categories: tandem alone, dual-channel combinations, Fletcher three-channel, interstitial needle hybrids, and vaginal applicators. (2) Fletcher applicator segmentation module: for cases identified as Fletcher three-channel, an nnU-Net (v2) model automatically delineates the 3D contours of the tandem and the left and right ovoids. (3) Template registration and mapping module: high-resolution applicator centerlines derived from 0.6-mm thin-slice CT are mapped onto thick-slice clinical CT space via a progressive weighted Kabsch registration algorithm. (4) Plan evaluation module: effective dwell points (weight > 0) parsed from the RT Plan are compared geometrically against the registered template centerline to verify positional accuracy.

2.3 Applicator Template Library

A template library was constructed based on thin-slice CT scans. The library included tandem applicators with curvature angles of 15, 30, 45, and 60 degrees, as well as left and right ovoid catheters. All templates were scanned in an unassembled state on a SOMATOM Definition AS CT scanner (Siemens

Healthineers, Erlangen, Germany) at a slice thickness of 0.6 mm, immobilized on a foam board. The scan range extended from 3 cm beyond the applicator tip to the transfer tube interface. Following CT parsing, the applicator channel centerlines were extracted using computer graphics techniques and saved as CSV files containing 3D coordinates, serving as the ideal geometric skeletons. The template library contains no patient information and is provided as Supporting Information.

2.4 Clinical Data Preprocessing

Clinical CT scans comprised 33–56 slices with an in-plane resolution of 512 x 512. Prior to training, volumes were resampled to an isotropic 1-mm resolution and cropped to the maximum region of interest containing the applicator, with zero-padding applied where necessary to produce a final size of 192 x 192 x 192. The classification model input was further down-sampled to 96 x 96 x 96, whereas the segmentation model retained the 192 x 192 x 192 resolution.

2.5 Applicator Classification Model (3D-CNN)

A lightweight 3D-CNN, termed Conv3DNet, was developed with an input size of 1 x 96 x 96 x 96. The network consisted of four 3D feature-extraction blocks, a global adaptive pooling layer, and a fully connected classification head (Figure. 2). The first three blocks each comprised a 3 x 3 x 3 convolution, 3D batch normalization, ReLU activation, and 3D max pooling, with channel numbers increasing from 32 to 128. The fourth block omitted pooling and increased channels to 256 to preserve fine-grained features. An adaptive 3D global average pooling layer compressed the feature maps into a 1D vector, which was fed into two fully connected layers with dropout rates of 0.5 and 0.3, respectively, yielding a five-class output.

To prevent data leakage between subsets, the dataset was partitioned at the patient level: patients were stratified by applicator class and randomly assigned to training, validation, and test sets at a 70:15:15 ratio, and all treatment plans from a given patient were allocated exclusively to a single subset. Labels were standardized using LabelEncoder. To address class imbalance (Class 2 accounted for only 0.8% of samples), class-balanced weights were embedded into the cross-entropy loss. Training was performed in PyTorch (version 2.7.1 + cu118) with the Adam optimizer (initial learning rate 1e-3) and a ReduceLROnPlateau scheduler (decay factor 0.5, minimum learning rate 1e-7). Early stopping with a patience of 15 epochs was employed. The batch size was 4, and the maximum number of epochs was 100.

2.6 Fletcher Applicator Segmentation Model (nnU-Net)

Segmentation was performed using the nnU-Net v2 self-configuring framework (version 2.5.2) with the 3d_fullres configuration. The input was a single-channel CT volume resampled to 1.0 x 1.0 x 1.0 mm, and the output was a three-channel segmentation map (left ovoid = 1, right ovoid = 2, tandem = 3). The encoder-decoder architecture comprised six resolution stages, with feature maps increasing from 32 to 320. Each stage contained two 3 x 3 x 3 convolutions followed by InstanceNorm and LeakyReLU, and downsampling was achieved via strided convolutions.

The dataset comprised 50 three-dimensional CT volumes from cervical cancer BT patients. Ground-truth channel contours were manually delineated in 3D Slicer (version 5.6.2; https://www.slicer.org) by a medical physicist with more than 10 years of experience in gynecologic brachytherapy and reviewed by a second senior medical physicist; disagreements were resolved by consensus. A five-fold cross-validation was employed (40 cases for training, 10 for validation in each fold). Because the left and right ovoid catheters are highly symmetric, x-axis (left-right) mirroring was disabled during training; only y- and z-axis flips were retained. The random rotation range was reduced from the default +/-30 degrees to +/-15 degrees to prevent left-right label swapping. The Adam optimizer was used (initial learning rate 1e-2, weight decay 3e-5) with polynomial learning rate decay. The model was trained for 1,000 epochs with a batch size of 2, using a combined cross-entropy and Dice loss. Validation was performed every 5 epochs, and the exponential moving average (EMA) weights yielding the highest validation Dice score were selected as the final model.

2.7 Template Registration and Mapping

We propose a progressive weighted arc-segment Kabsch registration algorithm to align the high-resolution 0.6-mm thin-slice CT template centerline onto the 3-mm thick-slice clinical CT while preserving the original spatial coordinates of the clinical data. The algorithm proceeds as follows: (1) Template supersmoothing: a cubic B-spline least-squares fit with 30 control points was applied to the template centerline, generating 300 uniformly sampled points to establish an ideal geometric skeleton. (2) Arc-segment extraction: the segment of maximum curvature was automatically identified based on the curvature peak, and a 60-mm arc segment (30 mm on each side of the peak) was extracted as the region of interest for registration. (3) Progressive three-level weighting: weights were assigned as 10 for the core region (within 15 mm of the peak), 3 for the transition region (15–25 mm), and 1 for the straight region (> 25 mm), ensuring that high-curvature regions received the highest registration priority. (4) Weighted Kabsch rigid registration: both the thick-slice clinical arc segment and the thin-slice template arc segment were resampled to 100 equidistant points, and weighted Kabsch registration was performed using the averaged bilateral weights to solve for the optimal rotation matrix and translation vector. (5) Global transformation: the resulting rigid transformation was applied to the entire 300-point supersmoothed template, producing a high-fidelity centerline mapped into the thick-slice clinical CT space. All registration results were visually inspected and confirmed by a senior medical physicist with 10 years of clinical experience before quantitative evaluation.

2.8 Treatment Plan QA Evaluation

In the Oncentra planning system, the applicator centerline represents the potential trajectory of the radioactive source, and the first dwell position requires a tip-offset correction. In this study, effective dwell points (weight > 0) were parsed from the RT Plan and evaluated for geometric consistency against the registered template centerline. Metrics included the statistical mean, standard deviation, root mean square error (RMSE), and maximum value of the minimum Euclidean distances from each point to the curve. The deviation of the first dwell point (applicator tip) was also assessed.

2.9 Simulated Error-Detection Study

To evaluate the error-detection capability of the pipeline, a simulation study was performed on 150 treatment plans with Fletcher applicators. In each plan, one effective dwell position was selected uniformly at random and displaced by 1, 3, or 5 mm in a random direction(50 plans per magnitude). An additional 50 error-free plans served as controls. A plan was flagged as failed when the maximum deviation of any dwell position from the registered template centerline exceeded 1 mm, a threshold motivated by the reported dose sensitivity of approximately 5%-6% DVH-parameter shift per millimeter of applicator reconstruction error in the anterior-posterior direction.[19] Detection sensitivity and specificity were calculated with Wilson score 95% confidence intervals.

2.10 Statistical Analysis

Continuous variables are summarized as mean±standard deviation, with 95% confidence intervals where applicable. Distance metrics of the three applicator channels, measured on the same cases, were compared using the Friedman test with post hoc Wilcoxon signed-rank tests and Bonferroni correction. Classification performance metrics are reported with patient-level bootstrap 95% confidence intervals (1,000 resamples). For the simulated error-detection study, sensitivity and specificity are reported with Wilson score 95% confidence intervals. A two-sided p value < 0.05 was considered statistically significant. Statistical analyses were performed in Python(SciPy 1.13.1, statsmodels 0.14.4).

2.11 Use of Generative Artificial Intelligence in Manuscript Preparation

During the preparation of this manuscript, the authors used a generative artificial intelligence tool solely to polish English language expression and improve sentence logic. The tool was not used to generate scientific content, perform data analysis, or interpret results. All content was reviewed, revised, and verified by the authors, who take full responsibility for the integrity and originality of this publication.

3. Results

3.1 Applicator Classification Performance

The classification model stopped early after 42 epochs. The test set comprised 556 plans, including 245 Class-1, 4 Class-2, 135 Class-3, 56 Class-4, and 116 Class-5 cases. The overall classification accuracy was 99.28% (plan-level bootstrap 95% CI: 98.56–99.82%). Except for the small-sample Class 2 (dual-channel combination), the precision, recall, and F1-score of the remaining four classes were all close to 100%. For Class 2, precision was 0.50, recall was 0.75, and F1 was 0.60. The macro-averaged F1 was 0.92, and the weighted-averaged F1 reached 0.99. The confusion matrix revealed that misclassified samples were concentrated only between Class 2 and Classes 1 and 3, confirming the efficacy of the 3D-CNN for imbalanced 3D medical volume classification (Figure. 3). On a workstation equipped with a 12th-generation Intel Core i7-12700 CPU (2.10 GHz), 32 GB of RAM, and an NVIDIA GeForce RTX 3070 GPU (8 GB VRAM), classification of a single case was completed in less than 1 s, and the full pipeline,

including classification, segmentation, registration, and plan evaluation, required approximately 60s per case.

3.2 Applicator Segmentation Performance

Five-fold cross-validation results (Table 1) showed that the Dice similarity coefficient (DSC) was 91.3 ± 2.2% for the left ovoid catheter, 90.9 ± 2.4% for the right ovoid catheter, and 90.0 ± 2.6% for the tandem. Precision was 92.1 ± 2.4% and 91.7 ± 2.8% for the left and right ovoid catheters, respectively, and 91.5 ± 4.0% for the tandem. Recall was 94.2 ± 2.1% and 93.8 ± 2.2% for the left and right ovoid catheters, respectively, and 94.9 ± 4.2% for the tandem. The 95% Hausdorff distance (HD95) was 1.0 ± 0.3 mm and 1.1 ± 0.5 mm for the left and right ovoid catheters, and 1.0 ± 0.2 mm for the tandem. The average surface distance (ASD) was 0.2 ± 0.6 mm and 0.3 ± 0.2 mm for the left and right ovoid catheters, and 0.9 ± 0.4 mm for the tandem. A representative 3D visualization comparing the automatically segmented channel contours with manual delineations is shown in Figure. 4.

Table 1
Performance of the nnU-Net segmentation model for the three applicator channels (five-fold cross-validation, n = 50). Values are mean±standard deviation.

Channel	DSC (%)	Precision (%)	Recall (%)	HD95 (mm)	ASD (mm)
Left ovoid catheter	91.3 ± 2.2	92.1 ± 2.4	94.2 ± 2.1	1.0 ± 0.3	0.2 ± 0.6
Right ovoid catheter	90.9 ± 2.4	91.7 ± 2.8	93.8 ± 2.2	1.1 ± 0.5	0.3 ± 0.2
Tandem	90.0 ± 2.6	91.5 ± 4.0	94.9 ± 4.2	1.0 ± 0.2	0.9 ± 0.4

3.3 Consistency Between Auto-Segmented and Template Centerlines

Fifty patients treated with Fletcher three-channel applicators (one plan per patient) were randomly selected from the 898 Fletcher plans for template registration evaluation and treatment plan verification. For these 50 cases, the Euclidean distances from the auto-segmented centerlines (clinical point sets) to the registered template curves were as follows. Left ovoid catheters: mean value 0.39 ± 0.07 mm (max 0.53 mm), standard deviation 0.35 ± 0.06 mm, RMSE 0.52 ± 0.06 mm, and maximum distance 1.56 ± 0.31 mm (max 2.24 mm). Right ovoid catheters: mean value 0.70 ± 0.11 mm (max 0.91 mm), standard deviation 0.37 ± 0.07 mm, RMSE 0.79 ± 0.09 mm, and maximum distance 1.48 ± 0.22 mm (max 1.93 mm). Tandem: mean value 0.47 ± 0.09 mm (max 0.67 mm), standard deviation 0.42 ± 0.08 mm, RMSE 0.64 ± 0.10 mm, and maximum distance 1.97 ± 0.37 mm (max 2.72 mm). Overall, the auto-segmented contours demonstrated high geometric consistency with the registered template (Figure. 5).

3.4 Treatment Plan Verification Results

For the same cohort of 50 patients, planned dwell points showed favourable geometric concordance with the registered template centerlines (Table 2). For the left and right ovoid catheters, mean point-to-curve distances were 0.36 ± 0.29 mm (range 1.27 mm) and 0.34 ± 0.22 mm (range 0.92 mm), respectively; RMSE was 0.41 ± 0.32 mm and 0.39 ± 0.24 mm; maximum deviations were 0.79 ± 0.55 mm and 0.76 ± 0.50 mm; and endpoint maximum distances were 0.39 ± 0.44 mm and 0.48 ± 0.38 mm. For the tandem, mean distance was 0.51 ± 0.35 mm (range 1.68 mm), RMSE was 0.53 ± 0.30 mm, maximum distance was 1.17 ± 0.55 mm, and endpoint maximum distance was 0.83 ± 0.68 mm. All three channels achieved submillimetre-to-millimetre spatial matching with stable tip anchoring, satisfying clinical requirements for dose accuracy (Fig. 6).

Table 2
Geometric consistency between planned dwell positions and registered template centerlines for the three applicator channels (n = 50). Values are mean ± standard deviation, with the observed maximum in parentheses.

Channel	Mean distance (mm)	RMSE (mm)	Maximum deviation (mm)	Endpoint maximum (mm)
Left ovoid catheter	0.36 ± 0.29 (1.27)	0.41 ± 0.32 (1.39)	0.79 ± 0.55 (2.04)	0.39 ± 0.44 (2.04)
Right ovoid catheter	0.34 ± 0.22 (0.92)	0.39 ± 0.24 (1.04)	0.76 ± 0.50 (2.13)	0.48 ± 0.38 (1.40)
Tandem	0.51 ± 0.35 (1.68)	0.53 ± 0.30 (1.23)	1.17 ± 0.55 (2.04)	0.83 ± 0.68 (2.04)

3.5 Simulated Error-Detection Results

Using the pre-defined failure criterion dwell-position deviation > 1 mm, the pipeline detected 18/50, 50/50, and 50/50 of the plans with 1-, 3-, and 5-mm simulated displacements, respectively, corresponding to a sensitivity of 36.0% (95% CI: 24.1% – 49.9%) for 1-mm errors, 100.0% (95% CI: 92.9% – 100.0%) for 3-mm errors, and 100.0% (95% CI: 92.9% – 100.0%) for 5-mm errors. Among the 50 error-free control plans, 3 were flagged, yielding a specificity of 94.0% (95% CI: 83.8% – 97.9%).

4. Discussion

Our results show that integrating a high-resolution template library with deep learning segmentation overcomes the systematic errors introduced by thick-slice CT in Fletcher applicator reconstruction.

AI-based QA tools are attractive because they offer consistent, reproducible results without the variability of manual checks.[23, 24] The BT workflow is complex and highly time-sensitive; conventional QA relying on manual checklists and operator experience struggles to meet increasing treatment throughput demands. This study first employed a lightweight 3D-CNN for rapid applicator classification (completed within one second, accuracy > 99%), which not only determines the subsequent QA pathway but also lays the technical foundation for extending QA to additional applicator types. For the minority

class (dual-channel combination, n = 28), the model maintained acceptable recall through class weighting, demonstrating a degree of robustness.

Automated applicator segmentation is a pivotal step in the QA workflow. Current manual reconstruction in BT planning is highly operator-dependent: window/level adjustment, sagittal/coronal plane selection, and visual point placement all introduce random errors. Furthermore, metal screws used in Fletcher applicator assembly generate substantial artifacts on CT, further degrading reconstruction accuracy. Deufel et al. [25] reported digitizing applicators using thresholding and HDBSCAN clustering, predicting that automation would improve plan consistency and reduce planning time. With the advent of deep learning for applicator segmentation, its accuracy, efficiency, and robustness have gained increasing clinical recognition.[26–28] The nnU-Net framework developed by Isensee et al.[29–32] is a self-configuring deep learning approach for biomedical image segmentation and has been adopted in several studies for auto-delineation in BT. In this study, the nnU-Net framework achieved a mean DSC of approximately 90.7% using only 50 annotated cases, validating its efficacy in small-sample medical image segmentation.

Nevertheless, segmentation remains affected by slice thickness and metal artifacts: the anterior portion of the ovoids is nearly perpendicular to the axial scan plane, producing a staircase contour on thick-slice CT, and metal artifacts also widen the apparent channel diameter on axial images (Figure. 5). Using the segmented contour centers directly as a reference for dwell-point evaluation would introduce systematic error. To overcome these limitations, this study innovatively introduced a 0.6-mm thin-slice CT template library and proposed a progressive weighted Kabsch registration algorithm. Scanned in an unassembled state and fitted with B-splines, the template provides a supersmoothed, high-fidelity ideal geometric skeleton. The progressive weighting strategy, assigning weights of 10 to the core region, 3 to the transition region, and 1 to the straight region, prioritizes alignment of the high-curvature arc while preventing registration imbalance caused by excessive stretching at the ends. The results showed that the mean distance between the registered template curve and the auto-segmented centerline was only 0.39–0.70 mm, markedly lower than the deviation expected when using the segmented centers directly, effectively suppressing thick-slice staircase artifacts.

Treatment plan verification demonstrated that manually reconstructed dwell points were generally well aligned with the template centerline (mean distance 0.34–0.51 mm), although the tandem tip deviation was slightly larger than that of the ovoids. This is attributable to the longitudinal slice thickness (3 mm), which is substantially larger than the in-plane submillimeter resolution, leading to increased tip-position uncertainty, a finding consistent with previous studies.[33–36] In a minority of plans, the maximum deviation exceeded 2 mm; retrospective review indicated that these cases were associated with residual registration errors in applicators with extreme curvature, highlighting a limitation when substantial elastic deformation is present.

The simulated error-detection study demonstrated that the pipeline reliably detects dwell-position displacements of 3 mm or larger (sensitivity 100%), which represents the clinically actionable error

range. Sensitivity for 1-mm displacements was 36.0%, indicating that errors of this magnitude approach the detectability floor of the method. This is consistent with the intrinsic geometric uncertainty of the pipeline: in error-free plans, maximum dwell-position deviations averaged 0.76–1.17 mm and occasionally exceeded 2 mm, so a 1-mm displacement is comparable to the baseline noise level. The observed specificity was 94.0%, with 3 of 50 error-free plans flagged; such flags would prompt a targeted manual review rather than plan rejection, which is acceptable in a QA workflow. The adopted 1-mm deviation threshold is consistent with the dose sensitivity reported by Tanderup et al.[19]

This study has several limitations. First, the current QA framework is restricted to Fletcher three-channel applicators and has not yet been extended to complex scenarios such as interstitial needle combinations. Second, although metallic applicators are generally rigid, subtle elastic deformation may occur at the distal end of long tandems in individual cases; our current algorithm does not incorporate quantitative elastic deformation compensation. Third, the template library currently comprises standard applicator specifications and does not cover all clinical variants. Fourth, ground-truth segmentation was performed by a single observer, and inter-observer variability was not quantified. Fifth, the simulated error study was limited to rigid tip displacements; other failure modes, such as mid-channel elastic deformation and applicator-type mismatch, were not evaluated. Finally, the individual contributions of the progressive weighting and arc-segment extraction to registration accuracy were not isolated through ablation experiments. Future work will extend the algorithm to interstitial applicator QA, integrate an elastic deformation compensation model, and embed the pipeline into the oncology information system to achieve closed-loop management from planning to verification.

5. Conclusion

The proposed automated QA framework integrates deep learning classification and segmentation with high-resolution template registration to verify dwell-position accuracy in Fletcher applicator brachytherapy. The pipeline achieves submillimetre-precision geometric verification and reliably detects simulated displacements of 3 mm or larger, offering an objective, efficient alternative to manual QA. With further validation and clinical integration, automated deep-learning QA has the potential to become a standard safeguard in radiotherapy treatment planning.

Declarations

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. The study used retrospectively collected, fully anonymized patient data and involved no more than minimal risk. In accordance with Article 32 of the *Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects* (National Health Commission, China, 2023), the study was exempt from ethics review, as confirmed by the Ethics Committee of The First Hospital of China Medical University. Because the data were fully anonymized and the study was retrospective, the requirement for informed consent was waived.

Patient consent

The requirement for written informed consent was waived by the The First Hospital of China medical Institutional Review Board because of the retrospective design and the use of fully anonymised data.

Clinical trial number: not applicable.

Funding

None.

Author Contribution

T.J. curated the data, performed the formal analysis, and wrote the original draft. W.T. developed the methodology and supervised the study. D.Z. reviewed and edited the manuscript. All authors read and approved the final manuscript.

Data availability

The data that support the findings of this study are not publicly available due to privacy and institutional restrictions. De-identified data may be made available from the corresponding author upon reasonable request, subject to approval by the institutional review board.

References

1. Lucia F, Miranda O, Schick U, Bourbonne V, Duvergé L. Dose escalation by brachytherapy for gynecological cancers. Cancer/Radiothérapie. 2022;26(6–7):905–10.
2. Jooya A, Talla K, Wei R, Huang F, Dennis K, Gaudet M. Systematic review of brachytherapy for symptom palliation. Brachytherapy. 2022;21(6):912–32.
3. Mayadev JS, Ke G, Mahantshetty U, Pereira MD, Tarnawski R, Toita T. Global challenges of radiotherapy for the treatment of locally advanced cervical cancer. Int J Gynecol Cancer.

2022;32(3):436–45.

4. Lindegaard JC, Petric P, Tan L, Hoskin P, Schmid MP, Jürgenliemk-Schulz I, et al. Are we making progress in curing advanced cervical cancer-again? Int J Gynecol cancer: official J Int Gynecol Cancer Soc. 2024;34(12):1940–5.
5. Huang CC, Peacock DNA, Stephens SJ, Chino JP. Recent advances in gynecologic radiation oncology. Cancer. 2025;131(9).
6. Kumar R, Sherwani Z, Lopez M, Vergalasova I, Zhang X, Eckroate B, et al. Disparities in brachytherapy utilization in cervical cancer in the united states: a comprehensive literature review. Gynecol Oncol. 2023;179:79–84.
7. Chargari C, Peignaux K, Escande A, Renard S, Lafond C, Petit A, et al. Radiotherapy of cervical cancer. Cancer Radiother. 2022;26(1–2):298–308.
8. Pötter R, Tanderup K, Schmid MP, Jürgenliemk-Schulz I, Haie-Meder C, Fokdal LU, et al. MRI-guided adaptive brachytherapy in locally advanced cervical cancer (EMBRACE-i): a multicentre prospective cohort study. Lancet Oncol. 2021;22(4):538–47.
9. Chino J, Annunziata CM, Beriwal S, Bradfield L, Erickson BA, Fields EC, et al. Radiation therapy for cervical cancer: executive summary of an ASTRO clinical practice guideline. Pract Radiat Oncol. 2020;10(4):220–34.
10. Sturdza A, Pötter R, Fokdal LU, Haie-Meder C, Tan LT, Mazeron R, et al. Image guided brachytherapy in locally advanced cervical cancer: improved pelvic control and survival in RetroEMBRACE, a multicenter cohort study. Radiother Oncol. 2016;120(3):428–33.
11. Albuquerque K, Tumati V, Lea J, Ahn C, Richardson D, Miller D, et al. A phase II trial of stereotactic ablative radiation therapy as a boost for locally advanced cervical cancer. Int J Radiat Oncol Biol Phys. 2020;106(3):464–71.
12. Schubert L, Liu A, Gan G, Amini A, Hutchison R, Ernest J, et al. Practical implementation of quality improvement for high-dose-rate brachytherapy. Pract Radiat Oncol. 2016;6(1):34–43.
13. Huang YJ, Gaffney D. Brachytherapy: common pitfalls. Int J Gynecol Cancer. 2022;32(3):352–7.
14. de Los Santos EF, Evans S, Ford EC, Gaiser JE, Hayden SE, Huffman KE, et al. Medical physics practice guideline 4.a: development, implementation, use and maintenance of safety checklists. J Appl Clin Med Phys. 2015;16(3):37–59.
15. Michaud AL, Benedict S, Montemayor E, Hunt JP, Wright C, Mathai M, et al. Workflow efficiency for the treatment planning process in CT-guided high-dose-rate brachytherapy for cervical cancer. Brachytherapy. 2016;15(5):578–83.
16. Groszkiewicz A, Foust E, Palestra B, Ertel MC, Boisen M, Sukumvanich P, et al. Can an in-person hands-on applicator-based teaching session improve trainee knowledge and comfort with complex gynecologic brachytherapy? Brachytherapy. 2024;23(3):290–300.
17. Sanders JC, Showalter TN, Ouhib Z, Thomadsen BR, Jacob D, Agarwal M, et al. Safety practices and opportunities for improvement in brachytherapy: a patient safety practices survey of the american brachytherapy society membership. Brachytherapy. 2020;19(6):762–6.

18. Fields EC, Kahn JM, Singer L. Education in gynecological brachytherapy. Int J Gynecol Cancer. 2022;32(3):407–13.
19. Tanderup K, Hellebust TP, Lang S, Granfeldt J, Pötter R, Lindegaard JC, et al. Consequences of random and systematic reconstruction uncertainties in 3d image based brachytherapy in cervical cancer. Radiother Oncol. 2008;89(2):156–63.
20. Walter YA, Rosen L, Moncrief O, Broekhoven B, Jacobs T, Syh J, et al. Workflow efficiency in vaginal cuff high dose rate brachytherapy using artificial intelligence-based organ segmentation and multi-channel cylinder modeling. Cancers (Basel). 2025;17(17):2751.
21. Ding Y, Chen Z, Wang Z, Wang X, Hu D, Ma P, et al. Three-dimensional deep neural network for automatic delineation of cervical cancer in planning computed tomography images. J Appl Clin Med Phys. 2022;23(4):e13566.
22. Cao Y, Vassantachart A, Ragab O, Bian S, Mitra P, Xu Z, et al. Automatic segmentation of high-risk clinical target volume for tandem-and-ovoids brachytherapy patients using an asymmetric dual-path convolutional neural network. Med Phys. 2022;49(3):1712–22.
23. Jia X, Albuquerque K. Artificial intelligence and deep learning for brachytherapy. Semin Radiat Oncol. 2022;32(4):389–99.
24. Casamitjana A, Fresco AMG, Candela-Juan C, Tudela R, Sala-Llonch R, López SM, et al. Robust framework for adaptive field-of-view automatic segmentation in vaginal brachytherapy for endometrial cancer. Phys Imaging Radiat Oncol. 2026;39:101000.
25. Deufel CL, Tian S, Yan BB, Vaishnav BD, Haddock MG, Petersen IA. Automated applicator digitization for high-dose-rate cervix brachytherapy using image thresholding and density-based clustering. Brachytherapy. 2020;19(1):111–8.
26. Jung H, Gonzalez Y, Shen C, Klages P, Albuquerque K, Jia X. Deep-learning–assisted automatic digitization of applicators in 3d CT image-based high-dose-rate brachytherapy of gynecological cancer. Brachytherapy. 2019;18(6):841–51.
27. Weishaupt LL, Sayed HK, Mao X, Choo R, Stish BJ, Enger SA, et al. Approaching automated applicator digitization from a new angle: using sagittal images to improve deep learning accuracy and robustness in high-dose-rate prostate brachytherapy. Brachytherapy. 2022;21(4):520–31.
28. Zhang D, Yang Z, Jiang S, Zhou Z, Meng M, Wang W. Automatic segmentation and applicator reconstruction for CT-based brachytherapy of cervical cancer using 3d convolutional neural networks. J Appl Clin Med Phys. 2020;21(10):158–69.
29. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2020;18(2):203–11.
30. Duprez D, Trauernicht C, Simonds H, Williams O. Self-configuring nnu-net for automatic delineation of the organs at risk and target in high-dose rate cervical brachytherapy, a low/middle-income country's experience. J Appl Clin Med Phys. 2023;24(8):e13988.
31. Krupien AJ, Abdulkadir Y, Luximon DC, Charters J, Dong H, Pham J, et al. Open-source deep-learning models for segmentation of normal structures for prostatic and gynecological high-dose-rate

brachytherapy: comparison of architectures. J Appl Clin Med Phys. 2025;26(6):e70089.
32. Li Z, Zhu Q, Zhang L, Yang X, Li Z, Fu J. A deep learning-based self-adapting ensemble method for segmentation in gynecological brachytherapy. Radiat Oncol. 2022;17(1).
33. Binny D, Lancaster CM, Kairn T, Trapp JV, Back P, Cheuk R, et al. Investigating the use of image thresholding in brachytherapy catheter reconstruction. Australas Phys Eng Sci Med. 2016;39(4):913–9.
34. Hu H, Yang Q, Li J, Wang P, Tang B, Wang X, et al. Deep learning applications in automatic segmentation and reconstruction in CT-based cervix brachytherapy. J Contemp Brachytherapy. 2021;13(3):325–30.
35. Hellebust TP, Kirisits C, Berger D, Pérez-Calatayud J, De Brabandere M, De Leeuw A, et al. Recommendations from gynaecological (GYN) GEC-ESTRO working group: considerations and pitfalls in commissioning and applicator reconstruction in 3d image-based treatment planning of cervix cancer brachytherapy. Radiother Oncol. 2010;96(2):153–60.
36. Kirisits C, Rivard MJ, Baltas D, Ballester F, De Brabandere M, van der Laarse R, et al. Review of clinical brachytherapy uncertainties: analysis guidelines of GEC-ESTRO and the AAPM. Radiotherapy oncology: J Eur Soc Therapeutic Radiol Oncol. 2014;110(1):199–212.

Figures

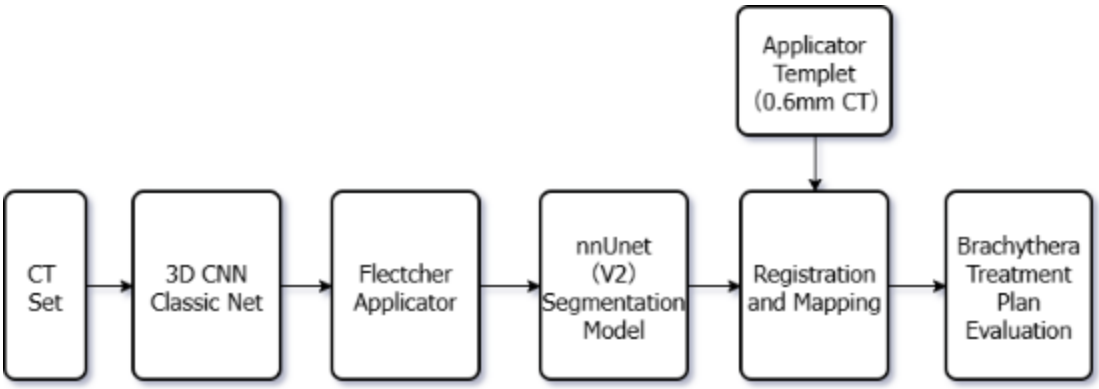


Figure 1

Workflow of the QA framework for brachytherapy treatment plans. CT series are input into a classification model to identify the applicator type. A segmentation model automatically delineates the three channels of the Fletcher applicator. The channel centerlines are registered to the corresponding applicator template in the template library, and the registered template centerline serves as the geometric reference for assessing whether planned source dwell positions fall within the acceptance threshold.



Figure 2

Architecture of the lightweight Conv3DNet for applicator classification. The network takes a single-channel 3D volume of size 1 x 96 x 96 x 96 as input. The feature-extraction backbone consists of four 3D convolutional blocks; the first three each comprise a 3 x 3 x 3 convolution, 3D batch normalization (BN), ReLU activation, and 3D max pooling, with channel depths increasing from 32 to 128. The fourth block omits pooling and expands channels to 256 to preserve fine-grained features. An adaptive 3D global average pooling layer follows the convolutional stages, feeding two fully connected (FC) layers with dropout rates of 0.5 and 0.3, respectively, yielding a five-class output.

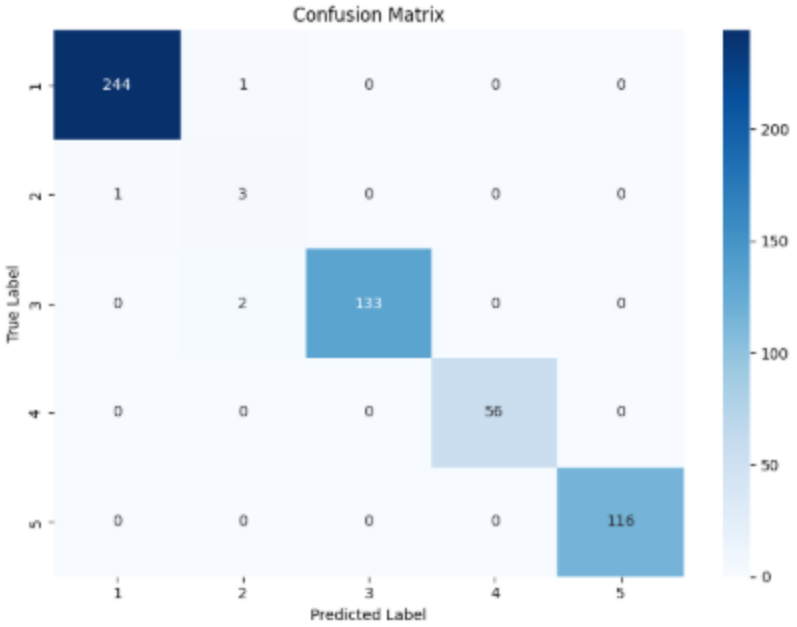


Figure 3

Confusion matrix heatmap of the applicator classification model. The overall accuracy was 99.28%. Misclassifications were confined to the minority Class 2 (dual-channel combination, n = 28) and its adjacent Classes 1 and 3, validating the robustness of the model on imbalanced 3D medical datasets.

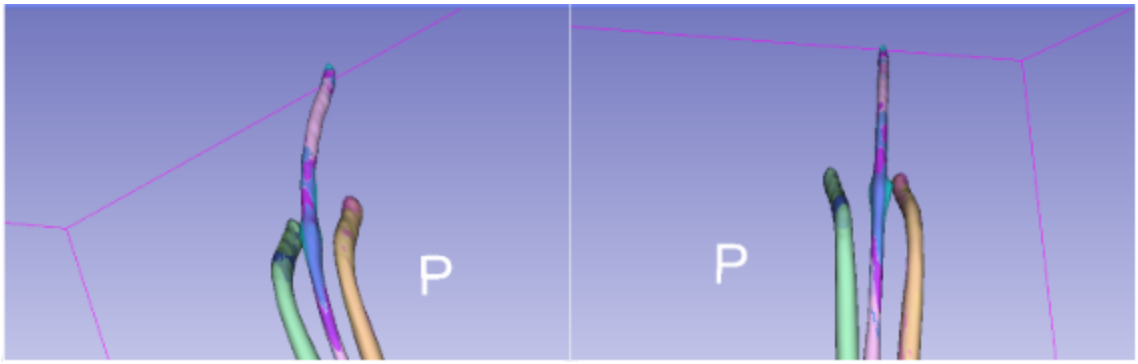


Figure 4

Three-dimensional visualization comparing the automatically segmented applicator channel contours with manual delineations for a representative patient.

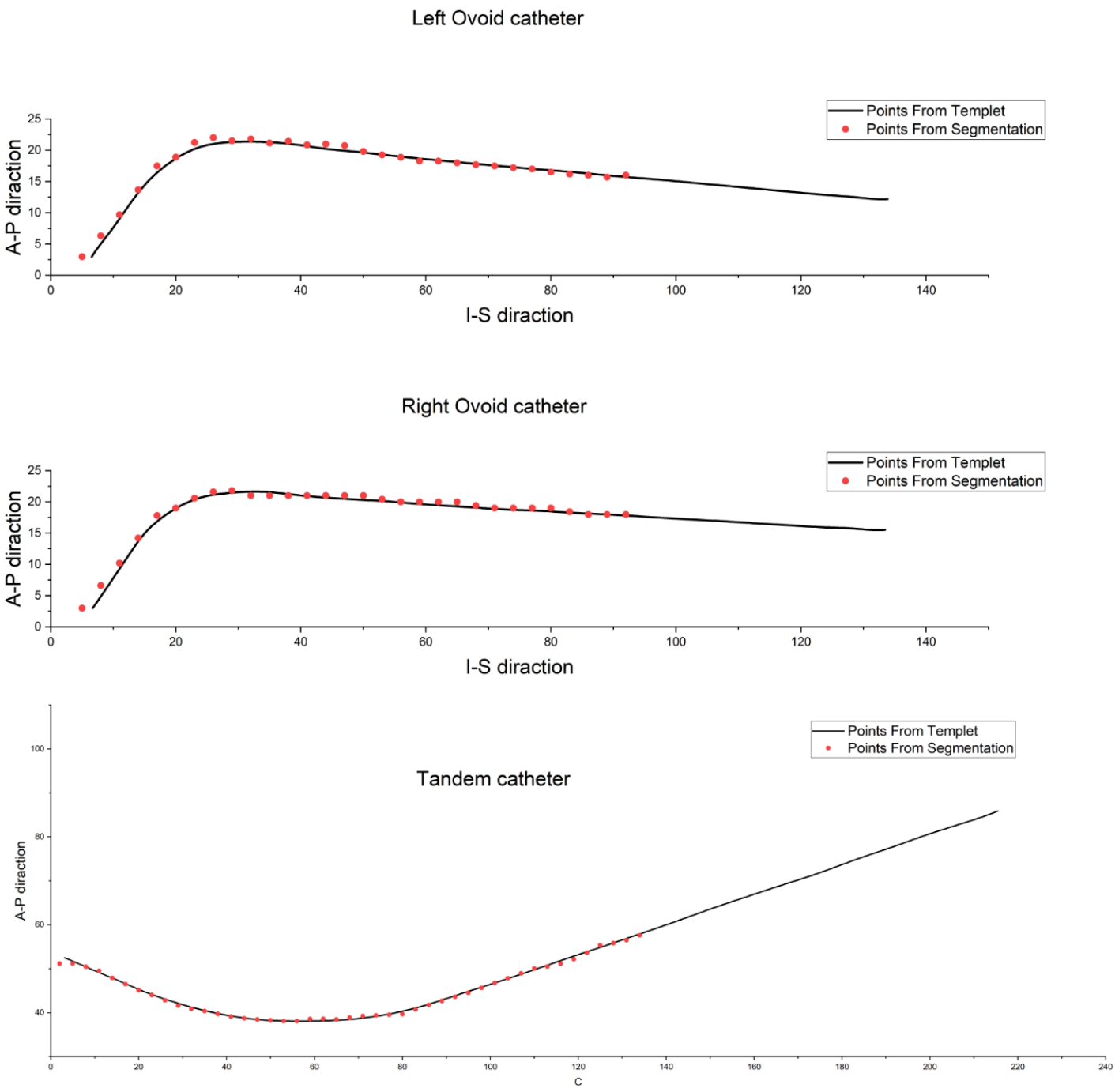


Figure 5

Sagittal projection of the auto-segmented applicator channel center points (red crosses) overlaid with the registered template applicator centerlines (black curves) for the left ovoid, right ovoid, and tandem channels. High geometric consistency is observed between the automatically segmented contours and the registered high-resolution template.

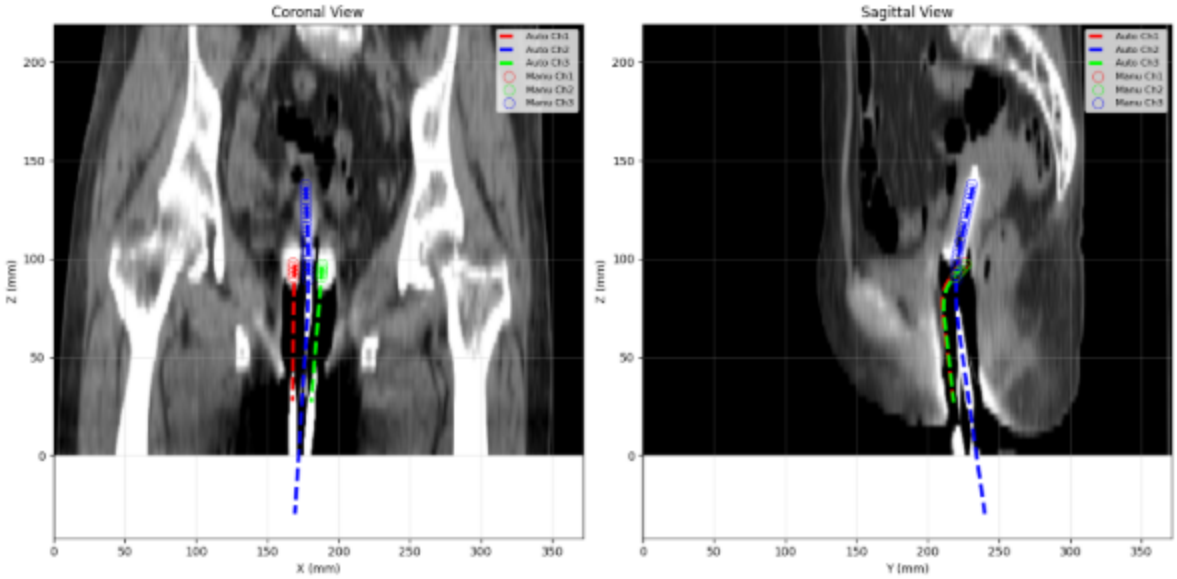


Figure 6

Comparison of manually reconstructed treatment plan dwell points (colored circles) and the registered template applicator centerlines (dashed lines) in coronal and sagittal views. The geometric deviation between the planned source positions and the template reference is maintained at the submillimeter-to-millimeter level.