Experimental investigation of real-time 3D beam's eye view image-guided radiotherapy for prostate SBRT

Danielle Chrystall[1,2], Maegan Stewart[1,3], Freeman Jin[4], Chandrima Sengupta[4], Maria D'Oliveira[2], Alannah Kejda[1], Levi Madden[1], Doan Trang Nguyen[5], Paul Keall[4], Jeremy Booth[1,2]

1 Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards, New South Wales, Australia

2 Institute of Medical Physics, University of Sydney, Camperdown, New South Wales, Australia

3 Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales, Australia

4 Image X Institute, University of Sydney, Eveleigh, New South Wales, Australia

5 SeeTreat Medical, Sydney, New South Wales, Australia

Correspondence

Danielle Chrystall, Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards, NSW, Australia.

Email: Danielle.Chrystall@health.nsw.gov.au

Abstract

Background

Real-time image-guided radiotherapy (IGRT) is critical for accurate dose delivery during stereotactic body radiotherapy (SBRT). Beam's eye view (BEV) imaging offers the unique advantage of reporting motion in the most dosimetrically relevant frame of reference without additional imaging dose. However, its clinical use is limited by poor contrast-to-noise ratio and marker occlusion by treatment beam apertures. Deep learning enables fast identification of indistinct marker features even in low contrast images, facilitating real-time BEV-IGRT. To support integration with standard-equipped linear accelerators, accurate 3D localization is essential—necessitating the development of a 3D BEV-IGRT system.

Purpose

This study aimed to develop and experimentally evaluate a novel real-time 3D BEV-IGRT system for potential clinical implementation during prostate SBRT.

Methods

A real-time 3D BEV-IGRT system was developed by integrating a deep learning-based 2D MV marker segmentation method with a 3D IGRT framework. Marker positions were segmented on MV images using a convolutional neural network (CNN) and used to predict 3D motion via a Gaussian maximum likelihood estimation method. A failure mode and effects analysis (FMEA) was performed by a multidisciplinary team. Mitigation strategies were implemented for high-risk failure modes, and risk priority numbers (RPN) were recalculated.

Experimental system evaluation was guided by failure modes identified through the FMEA. An anthropomorphic pelvic phantom with three implanted gold markers was mounted on a 3D motion-programmable platform. System performance was assessed under static and dynamic conditions, using treatment plans of increasing complexity, ranging from open fields to patient-representative volumetric modulated arc therapy plans. Dynamic performance was evaluated using four patient-derived prostate motion traces. Localization accuracy (mean error ± 1 SD) was assessed by comparing system-reported positions to ground truth derived from known motion trajectories or static displacements. 5^{th}/95^{th} error percentiles were calculated. System latency was measured as the time delay between motion initiation and system-reported displacement. Clinically acceptable accuracy was defined as within ± 2 mm in the superior-inferior (SI), anterior-posterior (AP) and left-right (LR) directions, and latency ≤ 500 ms.

Results

Forty-six failure modes were identified through the FMEA. High-risk failure causes included algorithmic limitations, algorithmic errors, human error, and marker occlusion. Incorporation of mitigation strategies—including eligibility screening, staff training, and workflow formalization—resulted in an average RPN reduction of 43% across the top ten high-risk failure modes. A risk-informed quality assurance program was designed to support clinical implementation.

Overall 3D BEV-IGRT system accuracy was 0.1 ± 0.7 mm (SI), -0.1 ± 0.8 mm (AP), and 0.1 ± 0.7 mm (LR). Accuracy remained within ± 2 mm in all directions across all individual tests. Overall 5^{th}/95^{th} percentile errors were [-1.0, 1.3] mm (SI), [-1.2, 0.9] mm (AP), and [-0.9, 1.0] mm (LR). System latency was 300 ± 100 ms.

Conclusions

The 3D BEV-IGRT system was experimentally validated, demonstrating clinically acceptable localization accuracy and latency, supporting its feasibility for clinical implementation. Integrated risk mitigation strategies effectively reduced workflow risk and promoted understanding of system vulnerabilities. Deployment is planned for a prostate SBRT clinical trial.

1. Introduction

Stereotactic body radiotherapy (SBRT) is a well-established treatment for patients with localized low-intermediate risk prostate cancer, offering favorable tumor control with significantly fewer treatment sessions compared to conventional fractionation.[1] Real-time image-guided radiotherapy (IGRT) enables monitoring of patient anatomy or implanted surrogates (e.g., fiducial markers) during treatment delivery, promoting margin reduction through tracking or gating. Kishan et al. reported a 43.8% relative reduction in the incidence of genitourinary toxicity (grade ≥ 2) in patients treated with aggressive margin reduction under MRI guidance; real-time IGRT was employed for intrafraction target localization to support planning target volume (PTV) margin reduction from 4 to 2 mm.[2] Similarly, Zhang et al. reported a 45.9% relative reduction in the incidence of urinary toxicity (grade ≥ 2) in patients treated with real-time 3D IGRT using kV/MV-guided gating (10.7%; n = 150) versus those treated with real-time 2D IGRT using kV-guided gating alone (19.8%; n = 121).[3] Collectively, these studies support the role of real-time 3D IGRT in improving clinical outcomes for prostate SBRT.

Commercial real-time 3D IGRT approaches typically require additional costly hardware or dedicated systems (e.g., electromagnetic transponders and MRI-linac), highlighting the need for accessible alternatives that can be implemented with standard radiotherapy systems. Onboard imaging systems integrated into modern linacs offer an accessible and cost-effective solution, which several research groups have leveraged for marker-based real-time 3D IGRT in prostate SBRT using in-house developed software. One method, kV intrafraction monitoring (KIM; Image-X Institute, University of

Sydney) enables near-continuous (5.5 to 11 Hz) monitoring of prostate displacement using kV imaging.[4-6] However, it delivers additional imaging dose and is limited by its orthogonal imaging plane, which directly resolves motion in only one direction of the steep treatment dose gradient. In contrast, MV-based IGRT enables motion monitoring directly in the treatment beam-view. Combined MV/kV triangulation-based image guidance (Memorial Sloan Kettering Cancer Center) uses the dosimetrically significant views of the MV imager through MV short-arc digital tomosynthesis reconstructions at 20° gantry intervals, synchronized with triggered kV image acquisition.[7-10] Given the unpredictable nature of intrafraction prostate motion,[11] transitioning towards continuous (cine) MV monitoring may offer clinical advantages during prostate SBRT.

Clinical use of MV-based IGRT has been limited by poor image contrast and marker occlusion, both of which limit marker visualization and motion monitoring capabilities. Several strategies to reduce marker occlusion during treatment have been proposed, including the use of imaging control points during planning,[7] recommendations for marker placement,[12] and plan optimization techniques.[13] Pre-processing image filtering is typically employed to enhance marker visibility in low contrast MV images.[14] In addition, technological advancements aimed at improving MV imaging performance have been proposed, such as multi-layer detector design[15] and pixelated scintillation.[16]

To overcome persisting limitations, a convolutional neural network (CNN) was previously trained to segment markers in cine MV images,[17] with the goal of overcoming clinical challenges in MV-based IGRT. This deep learning approach enhances segmentation robustness in challenging treatment scenarios (e.g., low-contrast and partial occlusion) by learning abstract visual features such as shape, edge, and contrast. Retrospective evaluation on a prostate SBRT patient cohort demonstrated sub-millimeter 2D accuracy in treatment beam-view co-ordinates.[17] However, 2D tracking alone is limited since it cannot support advanced tracking applications (e.g., multi-leaf collimator (MLC) or couch tracking) and requires additional 3D imaging to enable couch corrections, thereby prolonging treatment time. To address these limitations, the 2D MV CNN-based marker detection method was integrated into a real-time 3D IGRT framework, enabling 3D localization and facilitating deployment on standard-equipped linacs.

The purpose of this study was to develop and experimentally evaluate the novel real-time 3D beam's eye view (BEV)-IGRT system for potential clinical implementation during prostate SBRT. To support safe clinical translation, the evaluation included a failure mode and effects analysis (FMEA) to identify and mitigate potential risks prior to clinical deployment.

2. Methods

2.1 3D BEV-IGRT system

A previously developed deep learning-based 2D MV marker localization method[17] was integrated into a real-time 3D IGRT framework.[18] Until now, this framework has relied exclusively on kV imaging to enable real-time 3D IGRT. This study represents the first implementation of the system for real-time 3D IGRT using MV images, enabled by a CNN trained and tested on retrospective patient MV image datasets.[17] The CNN was comprised of four convolutional layers with batch normalization and rectified linear unit activations, three max-pooling layers, and a fully connected layer for binary classification (i.e., "marker" vs. "not marker"). Training was performed in MATLAB using stochastic gradient descent with cross entropy loss, and overfitting was mitigated through data augmentation, dropout, and early stopping. Figure 1 outlines the daily clinical workflow, with two distinct software stages: pre-treatment and treatment.

The pre-treatment stage comprised five key steps: (1) patient set-up, (2) real-time 3D BEV-IGRT software activation, (3) pre-treatment arc acquisition (low dose CBCT scan), (4) automatic marker segmentation in the CBCT images,19 and (5) construction of a 3D probability distribution function (PDF) to model marker positions. The PDF estimates the spatial distribution of individual markers, and a maximum likelihood estimation (MLE) method was used to determine the most probable 3D marker locations. This model captures correlations between motion in the resolved and unresolved imaging directions to infer 3D coordinates from 2D projections, following the method described by Poulsen et al.[20]

The during treatment stage comprised four key steps for each acquired MV frame. First, MV image acquisition and pre-processing were performed, including frame averaging to reduce noise and synchronization artifacts,[21] and application of a level-set masking technique to enhance marker visibility within the field of view (FOV). Second, 2D marker segmentation, using a CNN-based sliding window classification method. Tracking windows for each marker were initialized using either the planned CT position or the most recently reported 3D position. A 55 × 55 pixel window size was selected to accommodate potential marker motion between consecutive frames, while maintaining computational efficiency. In the event of complete occlusion or overlapping markers, real-time monitoring was temporarily suspended, and the system reverted to the last known 3D position. To support clinical decision-making, a countdown timer was included on the user interface (UI) to indicate the elapsed time that real-time monitoring was suspended. Further details on these steps are available in our retrospective study.[17] Third, 2D centroid estimation was performed using a marker template derived from the planning CT to infer the positions of occluded markers. Finally, the 3D centroid position was computed as a rolling average of the last 10 positions (i.e., 1 s window for a 10 Hz frame rate) and displayed on the UI enabling motion monitoring integrated with a preset gating tolerance (defined during pre-

treatment software activation stage). Manual intervention was triggered if the gating tolerance threshold was exceeded for a predefined duration.

2.2 Workflow risk assessment

A failure mode and effects analysis (FMEA), based on the methodology outlined in AAPM Task Group 100,[22] was performed to support the safe clinical implementation of 3D BEV-IGRT. The process involved a multi-disciplinary team, comprised of software developers, researchers, and qualified medical physicists.

A 3D BEV-IGRT process map was developed, outlining major processes and workflow steps. At each step, potential failure modes and their underlying causes were identified, and pre-mitigation numerical scores (1-10) were assigned for the likelihood of occurrence (O), severity (S) and lack of detectability (D), in accordance with AAPM TG 100 definitions. The risk priority number (RPN = O × S × D) was calculated for each failure mode, with higher values indicating greater clinical risk.

These initial scores, based on a preliminary understanding of the system, were used to identify system vulnerabilities and inform the design of a targeted quality assurance (QA) program. Where feasible, risk mitigation strategies and corrective actions were implemented to address root causes of high-risk failure modes, including system testing, software refinement, formalization of clinical workflows, and staff training. The FMEA was repeated, and post-mitigation scores (O', S', D') were reassessed to reflect the risk associated with the clinically deployable system. Updated risk priority numbers (RPN') were calculated to reflect the effectiveness of the proposed mitigation measures.

2.3 Experimental system evaluation

Figure 2 illustrates the experimental setup, including the phantom and treatment plans used for 3D BEV-IGRT evaluation. Three gold-plated cylindrical markers (1 mm diameter × 5 mm length) attached to a 3D-printed marker holder were suspended within a custom anthropomorphic pelvic phantom (Phantom Laboratory, Greenwich, NY) at the anatomical location of the prostate. Marker arrangement and inter-marker distances were based on an internal retrospective review of CT simulations from twelve prostate SBRT patients previously treated at our institution. The mean maximum marker separation between the three implanted markers was 1.8 ± 1.0 cm (SI), 1.0 ± 0.7 cm (AP), and 3.1 ± 0.8 cm (LR). Phantom marker arrangement was chosen to approximate these values within the confines of the marker holder geometry, resulting in maximum marker separations of 2.3 cm (SI), 2.5 cm (AP), and 2.7 cm (LR). The phantom was filled with water prior to planning CT acquisition to replicate soft tissue characteristics. Experimental tests were developed based on FMEA outcomes (Section III.A) and expanded upon previous target tracking implementations.[23-25] Three test types were

conducted: (1) static localization, (2) dynamic localization, and (3) system latency measurement.

2.3.1 Treatment planning

3D-CT images were acquired on a Brilliance CT Big Bore scanner (Philips Healthcare, Cleveland, OH). Three plan types were created in Eclipse v16.1 (Varian Medical Systems, Palo Alto, CA) using the Acuros External Beam v15.606 algorithm (AXB), reporting dose-to-medium:

1. Open: MLCs configured with a 7.4 cm × 10.4 cm (pan × tilt) aperture to ensure all markers remained within the FOV across all control points; delivered using a 10 MV flattening filter free (FFF) beam.

2. Partially Blocked: Controlled marker occlusions applied in predefined sub-arcs; delivered using a 10 MV beam and dynamic MLCs.

3. Volumetric modulated arc therapy (VMAT): Six previously treated prostate SBRT patient plans were adapted to the phantom using a mix of 10 MV FFF (n = 3) and 10 MV (n = 3) beams. Each plan prescribed 36.25 Gy to the PTV over five fractions, with a simultaneous boost to the intraprostatic lesion (45 Gy) and urethral sparing.[26]

Plan types 1 to 3 were designed to systematically increase test complexity when evaluating the 3D BEV-IGRT system. The Open plan was used to assess 2D marker segmentation and 3D localization accuracy without marker occlusion. The Partially Blocked plan was used to assess handling of marker occlusion in a controlled manner. VMAT plans were used to simulate clinical conditions. All plans were delivered using full 360° arcs, initiated at a gantry angle of 180°.

An in-house Eclipse script plug-in was developed to estimate marker occlusion for given MLC configurations. Each marker was modelled as a 3D point cloud and projected relative to MLC positions at 2° control point intervals. A marker was considered visible if ≥ 90% of its projected points were within the FOV (assuming static alignment to the planning CT) and inter-marker distances were ≥ 10 mm to avoid overlap.

2.3.2 Experimental setup

The phantom was positioned on a 3D programmable motion platform (HexaMotion, ScandiDos, Sweden) and aligned to the planning CT positions using a half-fan CBCT. For each test, the 3D BEV-IGRT system was initialized and a 200° pretreatment arc acquired (frame rate = 7 Hz; source-to-isocenter distance (SID) = 100 cm; source-to-detector distance (SDD) = 150 cm).

Accuracy was defined in accordance with ISO 5725-4,27 reporting trueness (mean error) and precision (1 SD) between the BEV-IGRT measured motion and the ground truth. Accuracy and 5th/95th percentile errors were reported in the SI, AP, and LR

directions. For static localization evaluations, the ground truth consisted of intentional positional offsets introduced to the marker centroid location. For dynamic localization evaluations, the ground truth was the applied motion trajectory, retrospectively synchronized to 3D BEV-IGRT output via correlation maximization. Synchronization-related uncertainties were minimal and have been characterized.[24] Minor inaccuracies due to mechanical limitations of the programming platform (e.g., occasional jitters) may contribute to residual error.

2.3.3 Treatment delivery

Plans were delivered using a TrueBeam v4.1 linear accelerator (Varian Medical Systems) with high-definition MLCs. During treatment delivery, MV images were acquired at 10 Hz in cine acquisition mode using an electronic portal imaging device (EPID; 1024 × 768 pixels, 0.394 × 0.394 mm/pixel; SID = 100 cm; SDD = 150 cm).

2.3.3.1 Static localization accuracy

Static localization accuracy was evaluated for all three plan types. The Open plan assessed the system's ability to detect known 3D positional shifts between the pretreatment and treatment arcs of 0 mm (origin) and ± 5 mm in SI, AP, and LR directions. The Partially Blocked and six VMAT plans assessed centroid localization in the presence of MLC-induced occlusion, without applied shifts.

2.3.3.2 Dynamic localization accuracy

Dynamic accuracy, defined as the system's ability to detect realistic 3D prostate motion during treatment delivery, was evaluated using the HexaMotion platform. Four clinically observed prostate motion trajectories from Langen et al. were used: stable, continuous drift, transient excursion, and erratic behavior. No gating tolerance was applied, ensuring uninterrupted motion. For each trajectory, two single-arc treatment plans (Open and VMAT; 10 MV FFF) were delivered.

2.3.3.3 System latency

The 3D BEV-IGRT system was installed on a workstation equipped with a GeForce RTX 3090 GPU (NVIDIA, Santa Clara, CA) to support real-time processing. System latency was measured using an independent video recording (10 fps) to simultaneously capture system output and applied phantom motion. A sinusoidal motion trace (± 5 mm amplitude; 4 s period) was applied in the SI direction. The recorded video was analyzed in MATLAB (VideoReader). Five consecutive motion peaks were manually identified on both data streams. Latency was defined as the mean ± 1 SD of the time offsets between corresponding peaks.

3. Results

3.1 Workflow risk assessment

Figure 3 illustrates the 3D BEV-IGRT process map, comprising three major processes and 16 distinct steps. Forty-six failure modes were identified through the FMEA. Initial scoring of these pre-mitigation failure modes was based on a preliminary understanding of the system (i.e., "naïve" scoring). These scores were instrumental in guiding software development, prompting targeted updates to address algorithmic limitations and eliminate software-related errors. Table 1 summarizes the ten highest ranked pre-mitigation failure modes, with RPNs ranging from 168 to 490. Eight of these received a severity score ≥ 5, indicating the potential for significant positional errors and incorrect dose delivery. An additional 18 failure modes also received high severity scores, but lower associated occurrence scores (O ≤ 2), leading to lower RPNs (≤ 70). The top-ranked failure modes were selected for further investigation.

Causes of failure were broadly categorized into algorithmic limitations, algorithmic errors, human errors, and marker occlusion. Mitigation strategies—including procedural formalization, staff training, and patient screening—were implemented where feasible to reduce occurrence and improve detectability in failure modes associated with human error and marker occlusion. These interventions led to an average RPN reduction of 43% (range = 0%–86%).

One of the most critical failures identified was "no markers visible in MV FOV," which had the second-highest pre-mitigation RPN of 350. To address this, an in-house Eclipse script plug-in was developed to estimate marker occlusion for given MLC configurations. This tool is intended to be integrated into the clinical workflow (steps 1.3 and 1.4, Figure 3) to proactively identify (and correct) cases with high rates of marker occlusion, hence improving detectability of this failure mode (D = 7, D' = 1). Post-mitigation occurrence of this failure mode remained at O' = 10, reflecting that some degree of marker occlusion is unavoidable in MV motion monitoring.

Algorithmic limitations or errors were the primary cause of six of the top-ranked failure modes. Following implementation of mitigation measures, an algorithm limitation relating to inaccurate marker segmentation became the failure mode with the highest RPN.' Most limitations stem from occlusion-related inaccuracies. In cases where < 3 markers were within the FOV, the centroid position was estimated using the visible markers position(s) combined with a template based on planned marker geometry. However, deviations in expected inter-marker distances (e.g., setup variation, marker migration, or anatomical deformation) could lead to inaccuracies in the estimated centroid position. Additional limitations arose from the inherent constraint of using a monoscopic 2D imaging system to infer 3D motion. However, since the inferred motion is along the depth axis of the treatment beam, any undetected motion from this limitation is expected to have minor dosimetric impact.

Algorithmic errors were considered separately from system limitations, as they reflected more serious cases of the system failing to operate as intended. These included failures in the 2D to 3D model parameters and CNN-based 2D localization errors (e.g., false positives). A risk-informed QA program was developed to support clinical implementation (Figure S1) and formed the foundation for initial system testing to validate expected system behavior (Section III.B).

3.2 Experimental system evaluation

Figure 4 shows violin plots of 3D BEV-IGRT localization errors, separated for static and dynamic tests. Overall system localization accuracy was 0.1 ± 0.7 mm (SI), -0.1 ± 0.8 mm (AP), and 0.1 ± 0.7 mm (LR), with 5th/95th percentile errors of [-1.0, 1.3] mm (SI), [-1.2, 0.9] mm (AP), and [-0.9, 1.0] mm (LR). Localization accuracy and percentile errors for individual tests can be found in Table S1.

3.2.1 Static localization accuracy

Figure 5 shows 3D BEV-IGRT centroid displacement versus gantry angle for all plan types. In the absence of marker occlusion (Figure 5a; Open plan), 5th/95th percentile errors were within ± 1.0 mm in all directions when the marker centroid was at isocenter. For Open plan tests with a large (± 5 mm) positional shift between the pre-treatment and treatment arcs, 5th/95th percentile errors exceeded ± 2 mm in the LR and AP directions (Table S1). This was attributed to limitations in resolving 3D motion from a single 2D imager at gantry angles where displacement occurred parallel to the treatment beam.

Figure 5b-c shows centroid displacement in the presence of MLC-induced marker occlusion for the Partially Blocked plan and one representative VMAT plan, respectively. Accuracy remained within ± 1 mm for all Partially Blocked and VMAT plans, with 5th/95th percentile errors within ± 2 mm.

3.2.2 Dynamic localization accuracy

Figure 6a shows violin plots of dynamic localization errors for both Open and VMAT plans. Overall dynamic accuracy was 0.2 ± 0.6 mm (SI), -0.3 ± 1.0 mm (AP) and 0.0 ± 0.5 mm (LR), with 5th/95th percentile errors of [-0.6, 1.2] mm (SI), [-1.8, 1.1] mm (AP), and [-0.9, 0.8] mm (LR). Figure 6b compares the 3D BEV-IGRT measured motion to the ground truth for four motion trajectories with the Open plan. Mean errors for individual tests were within ± 0.7 mm in all directions, with SD within 0.8 mm in SI and LR, and 1.5 mm in AP.

Figure 7 presents a polar plot of localization error for the transient trajectory, which had the largest spread of errors, with 5th/95th percentiles of [-5.0, 0.7] mm in the AP direction. All errors exceeding 2 mm occurred between gantry angles 340° to 20°, coinciding with rapid 6 mm AP displacement. Because the AP motion was aligned with

the treatment beam (i.e., the unseen component), it was not immediately detected by the 3D BEV-IGRT system.

3.2.3 System latency

Mean system latency was 300 ± 100 ms, which was within the AAPM TG 264 recommendations of 500 ms for real-time applications.[28]

4. Discussion

Real-time 3D BEV-IGRT was experimentally evaluated on a standard linac for prostate SBRT applications. Overall, the system achieved 5th/95th percentile localization errors of [-1.0, 1.3] mm, [-1.2, 0.9] mm, and [-0.9, 1.0] mm in the SI, AP, and LR directions, respectively. All static and dynamic localization tests met the pre-defined QA accuracy requirement, with mean errors and standard deviations within ± 2 mm in all directions. Unlike existing real-time IGRT techniques compatible with standard linacs, the presented approach enables near-continuous monitoring of target motion from the treatment beams perspective. In contrast, current techniques often rely on kV imaging during treatment[23, 24, 29] or monitor motion at reduced temporal frequencies.[7]

To support safe clinical implementation a workflow risk analysis was conducted. Results of the FMEA provided a foundational understanding of system vulnerabilities and mitigation effectiveness. Although subjective, it was a valuable process in guiding software refinement and QA requirements.

Marker occlusion is a well-recognized challenge in MV-based motion monitoring. This was reflected in our risk analysis, in which “no markers visible in MV FOV” was identified as a critical failure mode. To mitigate this, we proposed integrating a scripted pre-treatment QA step to predict marker occlusion rates (i.e., proportion of the plan with ≥ 1 marker visible and maximum period of full occlusion), which reduced the RPN of this failure mode by 86%. Successful identification of plans with suboptimal tracking efficiency can inform a replanning or screening process during clinical implementation. Acceptable thresholds for marker visibility should align with institution-specific IGRT protocols and desired motion monitoring frequency. The high post-mitigation occurrence score (O’ = 10) reflects limitations of the QA tool, which does not fully account for the complex interplay between marker motion and MLC geometry that can reduce tracking efficiency. For clinical translation, these limitations may be addressed by integrating treatment planning strategies to improve marker visibility,[7, 13] or via MLC tracking.[30]

3D BEV-IGRT is subject to the same inherent limitation of all 2D monoscopic imaging systems: inability to fully resolve 3D motion from 2D image data. This limitation was reflected in the top-ranked pre-mitigation failure mode. A key distinction between kV- and MV-based approaches lies in the direction of unresolved motion. For kV-based

systems, such as KIM and Intrafraction Motion Review (Varian Medical Systems), motion is directly resolved along one direction of the steep dose gradient and in the beam's depth. For MV-based systems, both directions of steep dose gradient are directly resolved, while depth resolution is limited. This limitation introduces gantry- and direction-dependent errors, which was observed during experimental testing with the transient trajectory in the Open plan, when the phantom underwent rapid 6 mm AP displacement coinciding with the beam's depth direction (Figure 7).

Several failure modes were identified at the 2D marker localization process step. Associated failure modes included false positives, poor marker distinction, and added inaccuracies from partial occlusion. The CNN used for classification was previously retrospectively validated on prostate SBRT patient images (n = 20 patients), achieving sub-millimeter 2D accuracy in beam-view coordinates, with 98.3% sensitivity and 99.9% specificity relative to manually identified markers from kV/MV triangulation.[17] In this work, CNN performance was further evaluated under near-clinical conditions using VMAT plans. Use of a phantom may cause the observed performance to differ compared to clinical applications. However, since the model was trained on patient data, it is expected that the performance be equivalent or better to what was observed in this study during clinical use. Several workflow safeguards were implemented prior to clinical software release, including restricted tracking window sizes and automatic pausing of segmentation when inter-marker distances were ≤ 10 mm. Testing identified that partial marker occlusion by MLCs was the dominant source of error in the SI direction. This resulted in off-center marker localization, introducing an apparent shift in centroid position in this direction. This was expected, given that our phantom contained 1 mm × 5 mm markers aligned with the long axis in the SI direction; however, in our practice, transperineal marker insertion typically results in more oblique marker orientation, which is likely to reduce partial occlusion-related SI uncertainty.

Additional areas of future development include extending 3D BEV-IGRT to anatomical sites with greater intrafraction motion, such as the liver and pancreas, by training site-specific CNN models. The current implementation requires a second CBCT (pre-treatment arc) to build the 3D model. Efforts are underway to remove this additional imaging dose. Potential strategies include replacing the pre-treatment arc with the set up CBCT; using a population-based PDF; or building the PDF in the first 45° of the treatment field.

Finally, the clinical performance of the 3D BEV-IGRT system will be prospectively evaluated in an ethically approved Phase II trial (NCT01742403), with the goal of validating its safety, reliability and efficacy in prostate SBRT. While this phantom study provided controlled benchmarking conditions, several challenges present when translating to patient treatments, including variability in patient size, marker arrangement, organ motion, dose rate, MLC configurations, and image quality. Previous

work demonstrated transferability of the CNN through patient-based model training and retrospective testing on independent patient cohorts with different EPID models.[17] Robust clinical translation will also rely on safeguards implemented during the FMEA, and on patient-specific QA, including both geometric and dosimetric accuracy.

5. Conclusion

A deep learning-based real-time 3D BEV-IGRT system was experimentally evaluated on a standard linear accelerator. 3D localization accuracy was within ± 2 mm across all individual static and dynamic tests. A comprehensive workflow risk analysis was performed to support safe clinical implementation for prostate SBRT applications, guiding clinical software deployment, informing design of a QA protocol, and integrating risk-mitigative actions into the workflow.

References

1 Jackson WC, Silva J, Hartman HE, et al. Stereotactic Body Radiation Therapy for Localized Prostate Cancer: a Systematic Review and Meta-Analysis of Over 6,000 Patients Treated On Prospective Studies. Int J Radiat Oncol Biol Phys. 2019; 104(4): 778-

2 Kishan AU, Ma TM, Lamb JM, et al. Magnetic Resonance Imaging-Guided vs Computed Tomography-Guided Stereotactic Body Radiotherapy for Prostate Cancer: the MIRAGE Randomized Clinical Trial. JAMA Oncol. 2023; 9(3): 365-373. doi:10.1001/jamaoncol.2022.6558

3 Zhang P, Happersett L, Burleson S, et al. Reduction of Postradiation Therapy Urinary Toxicity Via Intrafractional Megavoltage-Kilovoltage Prostate Location Monitoring. Int J Radiat Oncol Biol Phys. 2025; 121(1): 261-268. doi:10.1016/j.ijrobp.2024.07.2325

4 Keall P, Nguyen DT, O'Brien R, et al. Stereotactic prostate adaptive radiotherapy utilising kilovoltage intrafraction monitoring: the TROG 15.01 SPARK trial. BMC Cancer. 2017; 17(1): 180. doi:10.1186/s12885-017-3164-1

5 Ng JA, Booth JT, Poulsen PR, et al. Kilovoltage intrafraction monitoring for prostate intensity modulated arc therapy: first clinical results. Int J Radiat Oncol Biol Phys. 2012; 84(5): e655-61. doi:10.1016/j.ijrobp.2012.07.2367

6 Keall PJ, Ng JA, Juneja P, et al. Real-Time 3D Image Guidance Using a Standard LINAC: measured Motion, Accuracy, and Precision of the First Prospective Clinical Trial of Kilovoltage Intrafraction Monitoring-Guided Gating for Prostate Cancer Radiation Therapy. Int J Radiat Oncol Biol Phys. 2016; 94(5): 1015-1021. doi:10.1016/j.ijrobp.2015.10.009

7 Happersett L, Wang P, Zhang P, et al. Developing a MLC modifier program to improve fiducial detection for MV/kV imaging during hypofractionated prostate volumetric modulated arc therapy. J Appl Clin Med Phys. 2019; 20(6): 120-124.

8 Crotteau K, Lu W, Berry S, Happersett L, Burleson S, Cai W. Retrospective analysis of MV-kV imaging-based fiducial tracking in prostate SBRT treatment. J Appl Clin Med Phys. 2022; 23(6):e13593. doi:10.1002/acm2.13593

9 Zhang P, Hunt M, Telles AB, et al. Design and validation of a MV/kV imaging-based markerless tracking system for assessing real-time lung tumor motion. Med Phys 2018; 45(12): 5555-5563. doi:10.1002/mp.13259

10 Zhang P, Happersett L, Ravindranath B, Zelefsky M, Mageras G, Hunt M. Optimizing fiducial visibility on periodically acquired megavoltage and kilovoltage image pairs during prostate volumetric modulated arc therapy. Med Phys. 2016; 43(5): 2024. doi:10.1118/1.4944737

11 Langen KM, Willoughby TR, Meeks SL, et al. Observations on real-time prostate gland motion using electromagnetic tracking. Int J Radiat Oncol Biol Phys. 2008; 71(4):

12 Ma T, Kilian-Meneghin J, Kumaraswamy LK. Recommendation of fiducial marker implantation for better target tracking using MV imager in prostate radiotherapy. J Appl Clin Med Phys. 2018; 19(5): 389-397. doi:10.1002/acm2.12390

13 Omotayo A, Venkataraman S, McCurdy B. Constrained optimization towards marker-based tumor tracking in VMAT. Biomed Phys Eng Express. 2020; 7(1)doi:10.1088/2057-1976/abce0c

14 Park SJ, Ionascu D, Hacker F, Mamon H, Berbeco R. Automatic marker detection and 3D position reconstruction using cine EPID images for SBRT verification. Med Phys. 2009; 36(10): 4536-4546. doi:10.1118/1.3218845

15 Harris TC, Seco J, Ferguson D, et al. Improvements in beam's eye view fiducial tracking using a novel multilayer imager. Phys Med Biol. 2021; 66(15)doi:10.1088/1361-6560/ac1246

16 Star-Lack J, Shedlock D, Swahn D, et al. A piecewise-focused high DQE detector for MV imaging. Med Phys. 2015; 42(9): 5084-5099. doi:10.1118/1.4927786

17 Chrystall D, Mylonas A, Hewson E, et al. Deep learning enables MV-based real-time image guided radiation therapy for prostate cancer patients. Phys Med Biol. 2023; 68(9)doi:10.1088/1361-6560/acc77c

18 Hewson EA, Nguyen DT, O'Brien R, et al. The accuracy and precision of the KIM motion monitoring system used in the multi-institutional TROG 15.01 Stereotactic Prostate Ablative Radiotherapy with KIM (SPARK) trial. Med Phys. 2019; 46(11): 4725-4737. doi:10.1002/mp.13784

19 Mylonas A, Keall PJ, Booth JT, et al. A deep learning framework for automatic detection of arbitrarily shaped fiducial markers in intrafraction fluoroscopic images. Med Phys. 2019; 46(5): 2286-2297. doi:10.1002/mp.13519

20 Poulsen PR, Cho B, Keall PJ. A method to estimate mean position, motion magnitude, motion correlation, and trajectory of a tumor from cone-beam CT projections for image-guided radiotherapy. Int J Radiat Oncol Biol Phys. 2008; 72(5): 1587-1596. doi:10.1016/j.ijrobp.2008.07.037

21 Mooslechner M, Mitterlechner B, Weichenberger H, Huber S, Sedlmayer F, Deutschmann H. Analysis of a free-running synchronization artifact correction for MV-imaging with aSi:h flat panels. Med Phys. 2013; 40(3):031906. doi:10.1118/1.4790564

22 Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: application of risk analysis methods to radiation therapy quality management. Med Phys. 2016; 43(7): 4209. doi:10.1118/1.4947547

23 Ng JA, Booth JT, O'Brien RT, et al. Quality assurance for the clinical implementation of kilovoltage intrafraction monitoring for prostate cancer VMAT. Med Phys. 2014; 41(11):111712. doi:10.1118/1.4898119

24 Mueller M, Zolfaghari R, Briggs A, et al. The first prospective implementation of markerless lung target tracking in an experimental quality assurance procedure on a standard linear accelerator. Phys Med Biol. 2020; 65(2):025008. doi:10.1088/1361-6560/ab5d8b

25 Santanam L, Noel C, Willoughby TR, et al. Quality assurance for clinical implementation of an electromagnetic tracking system. Med Phys. 2009; 36(8): 3477-3486. doi:10.1118/1.3158812

26 Optimal Prostate Study. https://clinicaltrials.gov/study/NCT03386045

27 Accuracy (trueness and precision) of measurements methods and results. ISO Standard No 5725-4:2020: International Organization of Standardization; 2020.

28 Keall PJ, Sawant A, Berbeco RI, et al. AAPM Task Group 264: the safe clinical implementation of MLC tracking in radiotherapy. Med Phys. 2021; 48(5): e44-e64. doi:10.1002/mp.14625

29 Bertholet J, Toftegaard J, Hansen R, et al. Automatic online and real-time tumour motion monitoring during stereotactic liver treatments on a conventional linac by combined optical and sparse monoscopic imaging with kilovoltage x-rays (COSMIK). Phys Med Biol. 2018; 63(5):055012. doi:10.1088/1361-6560/aaae8b

30 Keall PJ, Colvill E, O'Brien R, et al. The first clinical implementation of electromagnetic transponder-guided MLC tracking. Med Phys. 2014; 41(2):020702. doi:10.1118/1.4862509

31 Continuous Monitoring of Prostate Position During Radiotherapy (KIM Gating). Recruiting, Royal North Shore Hospital. https://clinicaltrials.gov/study/NCT01742403

Figure Captions

Figure 1. 3D BEV-IGRT treatment workflow.

Figure 2. Overview of the phantom, setup, and plans used in experimental system evaluation.

Figure 3. 3D BEV-IGRT process map, highlighting both major processes (preparation, pre-treatment, and during treatment) and steps.

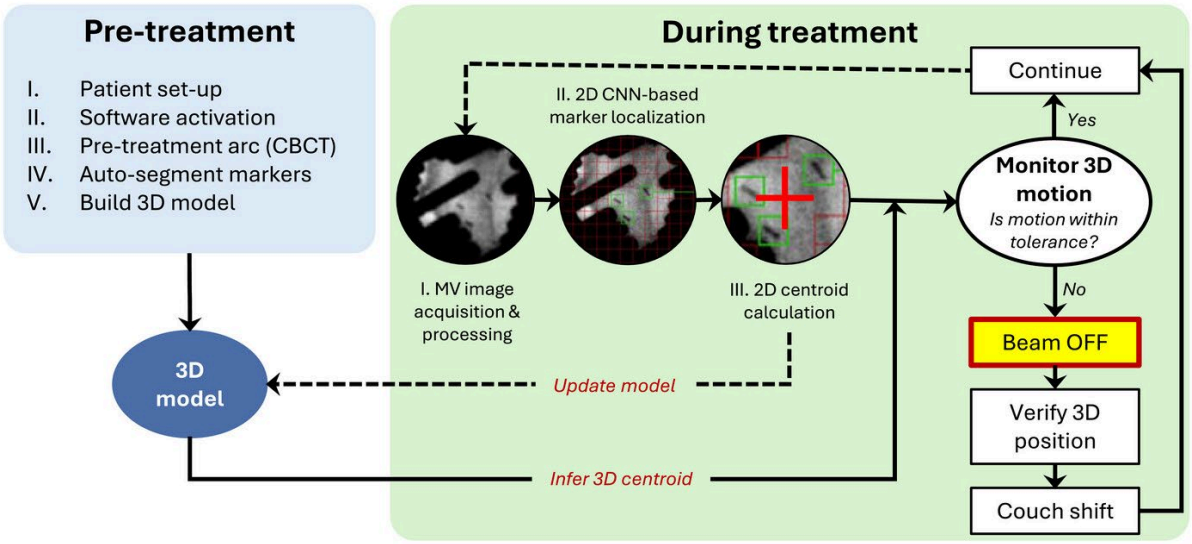
Figure 4. Violin plots of the 3D BEV-IGRT localization errors (SI = green, AP = red, LR = blue). The left half of each violin (darker shade) represents static localization errors, while the right side (lighter shade) represents dynamic localization errors. Central boxplots represent overall spread of errors, 5^{th}/95^{th} percentile errors (box bounds) and mean error (white circle).

Figure 5. 3D centroid displacement (SI = green, AP = red, LR = blue) versus gantry angle under static phantom conditions, aligned to isocenter, for (a) Open, (b) Partially Blocked, and (c) a representative VMAT plans. Shaded regions indicate Eclipse script-predicted real-time monitoring interruptions by potential marker overlap (dark gray; inter-marker distance < 10 mm) or complete marker occlusion (light gray; no markers in the FOV due to MLCs for > 10° gantry rotation).

Figure 6. Dynamic localization errors (SI = green, AP = red, LR = blue). (a) Violin plots showing dynamic localization errors, with the left half of each violin (darker shade) representing Open plan errors, and the right half (lighter shade) representing VMAT plan errors. (b) 3D BEV-IGRT measured motion overlayed on the applied motion traces for the Open plan with stable, continuous drift, transient, and erratic motion trajectories.

Figure 7. Polar plot representation of 3D BEV-IGRT measured motion (broken line) compared with the ground truth (solid line), and the localization error (dotted line). Plan = Open; Motion trajectory = Transient.

582 Figure 1

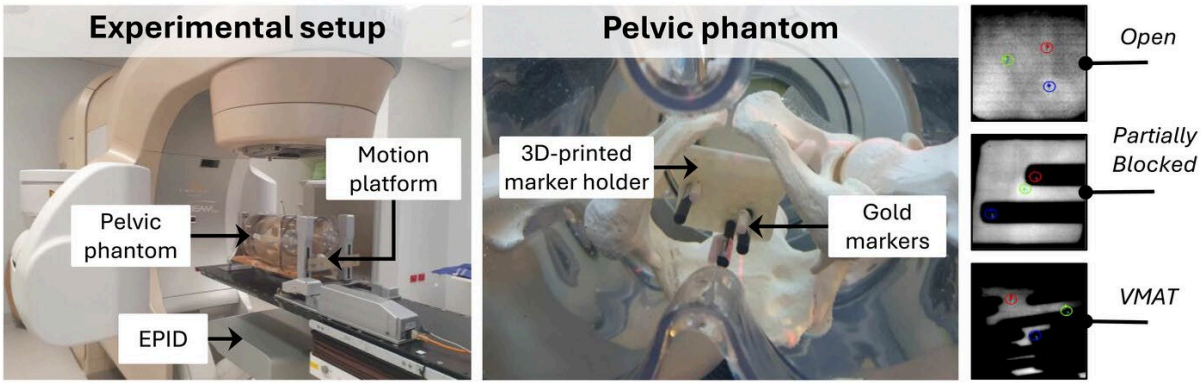


583

584

585

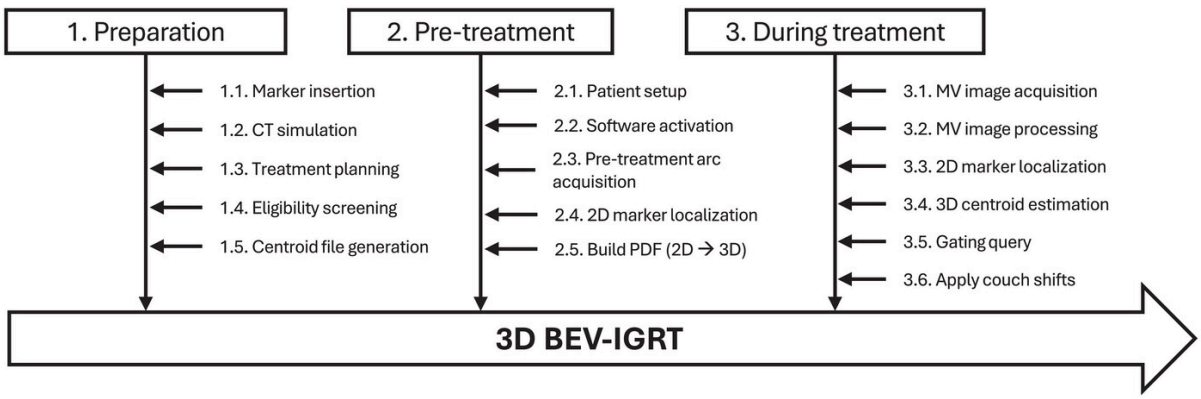
586 Figure 2



587

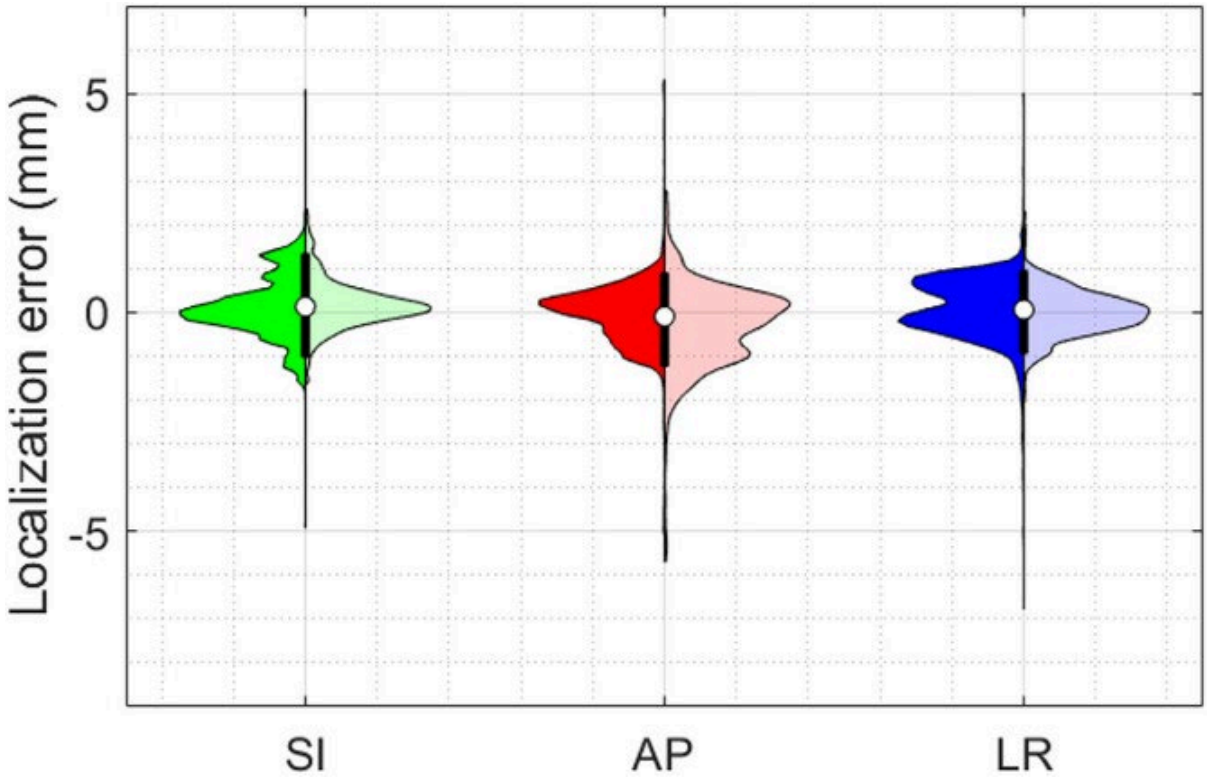
588

589 Figure 3



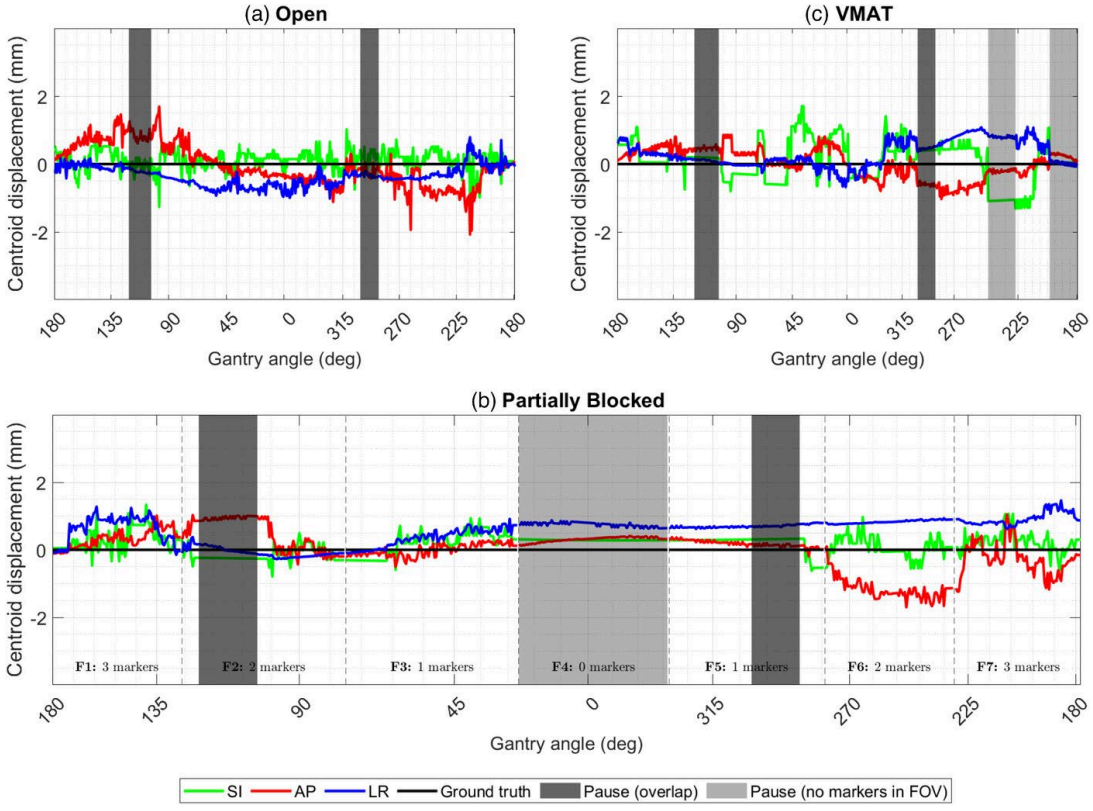
590

591 Figure 4



592

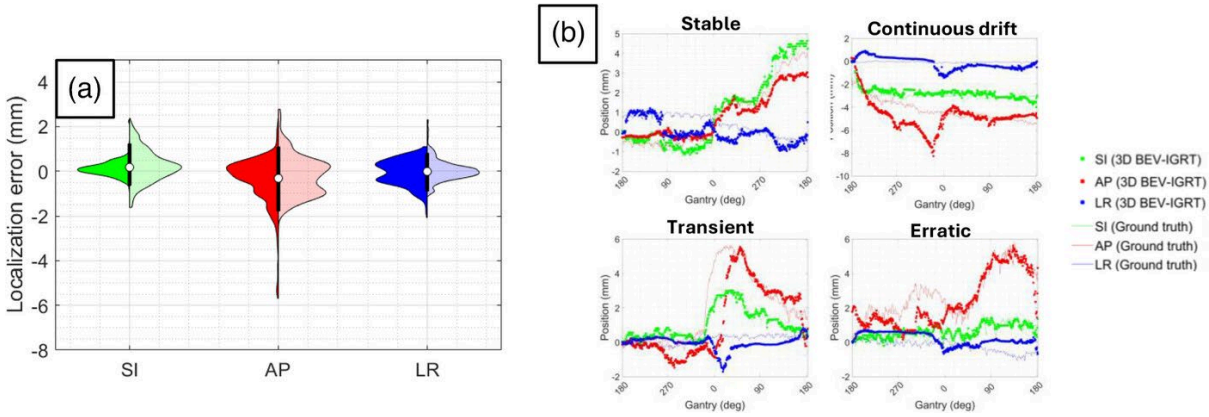
593 Figure 5



594

595

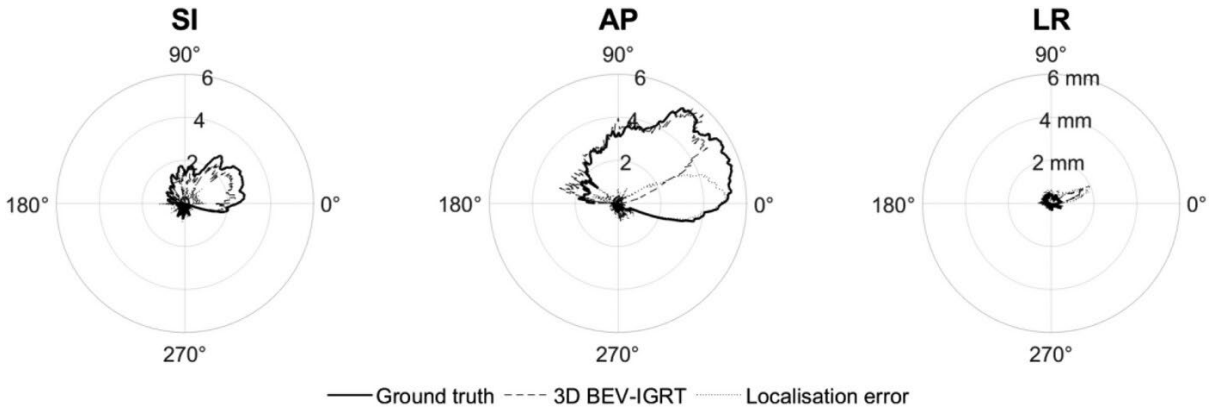
596 Figure 6



597

598

599 Figure 7



600

TABLE 1. Summary of the top ten 3D BEV-IGRT failure modes (and their causes) ranked by RPN prior to mitigation. Mitigation measures and re-scored occurrence, severity and lack of detectability estimates included. Refer to Figure 3 for 'Process step' definitions.

Rank	Process step	Potential failure mode	Cause of failure	O	S	D	RPN	Mitigation measure	O'	S'	D'	RPN'	Estimates in RPN reduction
1	3.4	Inaccurate 2D to 3D position estimation	Algorithm limitation (e.g., inaccurate 3D model parameters; direction of 3D motion unresolved in 2D imager)	7	7	10	490	Integrate workflow safeguards (3D position verification); Perform system testing	3	7	10	210	−57%
2	1.3	No markers visible in MV FOV	Occlusion of markers by MLCs	10	5	7	350	Prior assessment of marker visibility; Software update (indicate occlusion); Perform system testing	10	5	1	50	−86%
3	3.3	Poor marker distinction	Algorithm limitation (e.g.,	10	7	5	350	Software update (pause segmentation	5	7	5	175	−50%

Rank	Process step	Potential failure mode	Cause of failure	O	S	D	RPN	Mitigation measure	O’	S’	D’	RPN’	Estimates in RPN reduction
			handling overlapping markers)					during potential overlap); Perform system testing					
4	3.3	Markers not localized after occlusion	Algorithm limitation (e.g., Markers outside tracking window; Large patient movement)	10	7	5	350	Software update (enable manual marker re-segmentation); Perform system testing	2	7	5	70	−80%
5	3.3	Inaccurate segmentation	Algorithm limitation (e.g., marker edge)	10	3	8	240	Perform system testing	10	3	8	240	0%
6	3.4	Inaccurate centroid estimation when < 3 markers visible	Algorithm limitation (e.g., software not handling < 3 markers visible)	3	7	10	210	Perform system testing	3	7	10	210	0%
7	2.2	Incorrect trigger	Human error	3	7	10	210	Develop workflow	2	7	10	140	−33%

Rank	Process step	Potential failure mode	Cause of failure	O	S	D	RPN	Mitigation measure	O’	S’	D’	RPN’	Estimates in RPN reduction
		threshold selected						documentation; Train staff					
8	2.2	Incorrect settings	Human error (e.g., incorrect entry of data directories)	3	7	10	210	Develop workflow documentation; Train staff	1	7	10	70	−67%
9	3.1	Incorrect EPID SDD	Human error	6	3	10	180	Develop workflow documentation; Train staff	3	3	10	90	−50%
10	3.3	Incorrect segmentation	Algorithm error (e.g., CNN detected false positive); Technical error (MV image artifact)	3	7	8	168	Perform system testing	3	7	8	168	0%

Abbreviations: O, occurrence score; S, severity score; D, lack of detectability score; RPN, risk priority number (O × S × D); MV, megavoltage; FOV, field of view; MLC, multi-leaf collimator; EPID, electronic portal imaging device; SDD, source-to-detector distance; PDF, probability distribution function; O’, S’, D’, re-scored occurrence, severity and detectability with mitigation applied; RPN’, re-calculated risk priority number with mitigation applied (O’ × S’ × D’).