Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

**ACCEPTED MANUSCRIPT**

# Deep learning enables MV-based real-time image guided radiation therapy for prostate cancer patients

To cite this article before publication: Danielle Maria Chrystall *et al* 2023 *Phys. Med. Biol.* in press https://doi.org/10.1088/1361-6560/acc77c

# Deep learning enables MV-based real-time image guided radiation therapy for prostate cancer patients

Danielle Chrystall[1, 2], Adam Mylonas[3], Emily Hewson[3], Jarad Martin[4], Jeremy Booth[1, 2], Paul Keall[3], Doan Trang Nguyen[3, 5]

[1] Northern Sydney Cancer Centre, Royal North Shore Hospital, Sydney, NSW, Australia
[2] School of Physics, University of Sydney, Sydney, NSW, Australia
[3] ACRF Image X Institute, University of Sydney, Sydney, NSW, Australia
[4] The Department of Radiation Oncology, Calvary Mater Newcastle, Waratah, NSW, Australia
[5] SeeTreat Medical, Sydney, NSW, Australia

E-mail: Danielle.Chrystall@health.nsw.gov.au

## Abstract

**Objective**: Using MV images for real-time image guided radiation therapy (IGRT) is ideal as it does not require additional imaging equipment, adds no additional imaging dose and provides motion data in the treatment beam frame of reference. However, accurate tracking using MV images is challenging due to low contrast and modulated fields. Here, a novel real-time marker tracking system based on a convolutional neural network (CNN) classifier was developed and evaluated on retrospectively acquired patient data for MV-based IGRT for prostate cancer patients.

**Approach**: MV images, acquired from 29 VMAT prostate cancer patients treated in a multi-institutional clinical trial, were used to train and evaluate a CNN-based marker tracking system. The CNN was trained using labelled MV images from 9 prostate cancer patients (35 fractions) with implanted markers. CNN performance was evaluated on an independent cohort of unseen MV images from 20 patients (78 fractions), using a Precision-Recall curve (PRC), area under the PRC plot (AUC) and sensitivity and specificity. The accuracy of the tracking system was evaluated on the same unseen dataset and quantified by calculating mean absolute (± 1 SD) and [1$^{st}$, 99$^{th}$] percentiles of the geometric tracking error in treatment beam co-ordinates using manual identification as the ground truth.

**Main results**: The CNN had an AUC of 0.99, sensitivity of 98.31% and specificity of 99.87%. The mean absolute geometric tracking error was 0.30 ± 0.27 and 0.35 ± 0.31 mm in the lateral and superior-inferior directions of the MV images, respectively. The [1$^{st}$, 99$^{th}$] percentiles of the error were [-1.03, 0.90] and [-1.12, 1.12] mm in the lateral and SI directions, respectively.

**Significance:** The high classification performance on unseen MV images demonstrates the CNN can successfully identify implanted prostate markers. Furthermore, the sub-millimetre accuracy and precision of the marker tracking system demonstrates potential for adaptation to real-time applications.

Keywords: radiotherapy, deep learning, MV imaging, EPID, prostate cancer, motion management, marker tracking

## 1. Introduction

Stereotactic body radiation therapy (SBRT) is increasingly becoming the preferred method of radiation treatment for patients with prostate cancer [1, 2]. SBRT plans have steep dose gradients at target boundaries that require a high level of treatment accuracy to achieve tumour control while avoiding toxicity to nearby radiosensitive organs (rectum, bladder and urethra). Intrafraction prostate motion can cause misalignment between the planned target volume and the delivered treatment beam, which can simultaneously reduce target dose and overdose surrounding healthy tissue. Kotte *et al* (2007) performed an analysis of 427 prostate cancer patients and reported intrafraction prostate motion exceeding 2 mm in 66% of treatment fractions, and exceeding 3 mm in 28% of treatment fractions [3]. Therefore, real-time image-guided radiation therapy (IGRT) is necessary to account for intrafraction motion and deliver accurate prostate SBRT treatment.

Current commercial real-time IGRT, including the MRI-linac [4], CyberKnife [5] and Calypso [6], require specialized systems and/or dedicated hardware that can be expensive to implement. A focus on more accessible real-time IGRT approaches have been investigated, leveraging images acquired from the onboard kV [7, 8] and MV [9-12] imagers. The unique advantages of using MV over kV imaging are that: (1) no additional imaging dose is delivered to the patient, as the therapeutic beam doubles as the imaging beam, and (2) images are acquired in real-time in the view of the treatment beam, displaying patient anatomy as it is being irradiated. Despite these advantages, MV imaging has been severely under-utilized due to the low contrast of the MV images and beam modulation, whereby MLC leaves temporarily obstruct the markers, limiting visibility. Previously developed MV-based marker tracking methods are based on Laplacian of Gaussian blob-detection filters [9, 12], template matching [10, 13, 14] and machine learning methods [11]. Limitations from these methods include an inability to track irregular shaped markers, marker misidentification, high computational cost of large templates and manual labelling requirements.

We propose a deep learning approach to automatically track markers in MV images that provides fast segmentation facilitating real-time applications and does not rely on knowledge of prior marker properties, a learning period or any manual intervention. The novel contributions of this work include: (1) Development of an automatic, deep learning based marker tracking system using MV images (Figure 1), and (2) Quantitative evaluation of the performance of this novel tracking method on an unseen patient cohort from a separate institution to the training dataset. A convolutional neural network (CNN) classifier was trained to track markers implanted in the prostate. A CNN was chosen over traditional machine learning classification methods (e.g. support vector machines (SVMs) and Naïve-Bayes) due to the higher level of achievable accuracy in image classification tasks, particularly when trained on a large dataset [15]. Mylonas *et al* (2019) developed a deep learning framework for tracking markers in kV fluoroscopic projections and reported sub-millimetre tracking accuracy and precision. Here, we extend this framework to automatically track markers in low contrast MV images.

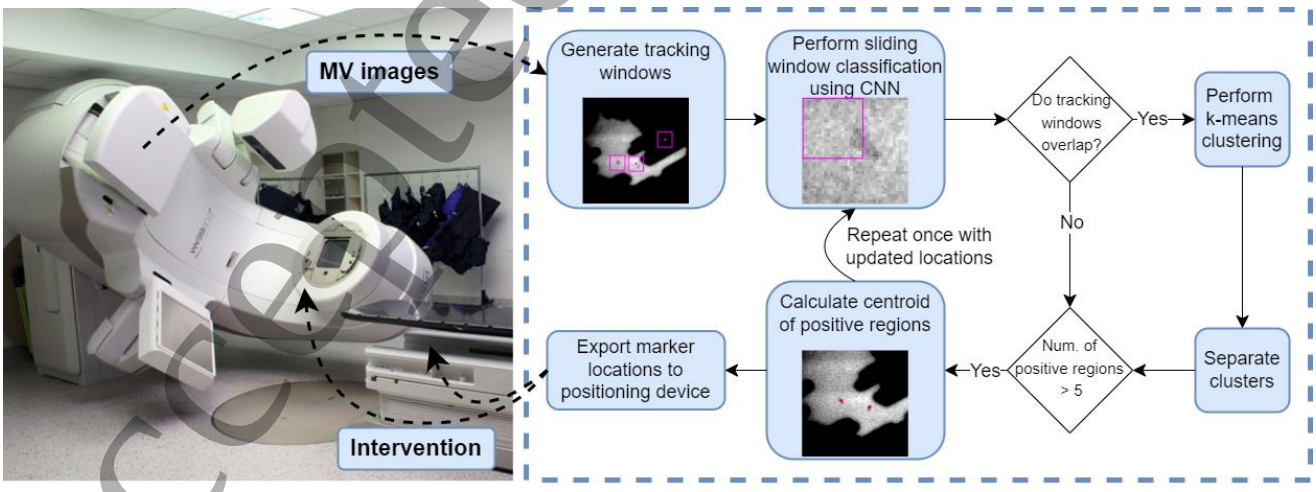


*Figure 1. A method to localize markers on cine MV images based on a trained CNN.*

## 2. Materials and methods

### 2.1. Clinical trial data

Twenty-nine patients with low to intermediate risk prostate cancer were included in this study. These patients were recruited in the ethics approved TROG 15.01 SPARK clinical trial (NCT02397317) [16, 17]. SPARK was a multi-institutional prospective phase II clinical trial that utilized Kilovoltage Intrafraction Monitoring (KIM) for real-time motion management during prostate SBRT. KIM automatically segmented the implanted markers using the kV images and estimated the 3D position [18]. Hewson *et al* (2019) reported sub-millimetre accuracy and precision of KIM target tracking [7]. Each patient had three gold cylindrical markers implanted into the prostate prior to treatment. Patients were prescribed 36.25 Gy to the planning target volume (PTV) in five fractions, using Volumetric Modulated Arc Therapy (VMAT) delivered over two arcs. Table 1 provides an overview of the linear accelerators, beam models and number of patients treated at each institution. Eleven patients were treated with a 6 MV flattened beam and 18 patients were treated with a 10 MV flattening filter free beam.

*Table 1. The linear accelerator, treatment beam and electronic portal imaging device (EPID) models and number of patients treated at the separate institutions involved in this study.*

| Institution | Linear accelerator | Beam model | EPID model (Pixel matrix size) | Num. of patients | Marker length (mm) | Use of dataset |
|---|---|---|---|---|---|---|
| 1 | Varian TrueBeam | 10 MV FFF | AS1000 (1024 × 768) | 7 | 3 | Training |
| 2 | Varian TrueBeam | 6 MV | AS1200 (1190 × 1190) | 2 | 5 | Training |
| 3 | Varian TrueBeam | 10 MV FFF<br>6 MV | AS1200 (1190 × 1190) | 11<br>9 | 3 | Testing |

### 2.2. Convolutional neural network for marker classification

#### *2.2.1. Datasets*

The CNN classifier is a key component to the marker tracking system outlined in Figure 1. It is a binary classifier that distinguishes between positive sub-images containing marker and negative sub-images containing background. Table 1 shows the distribution of patient data into training and testing datasets, and Figure 2 shows an overview of how these datasets were used to train the CNN and evaluate the performance of the CNN and marker tracking system. Manually labelled MV images from 9 patients (35 fractions) across Institutions 1 and 2 were used to train the CNN and were randomly split into 80% training and 20% validation datasets. The CNN performance was evaluated using the testing dataset, which contained manually labelled images from 20 prostate cancer patients (78 fractions) treated at Institution 3. The performance of the marker tracking system was evaluated using MV images from one representative fraction each of patients in the testing dataset.

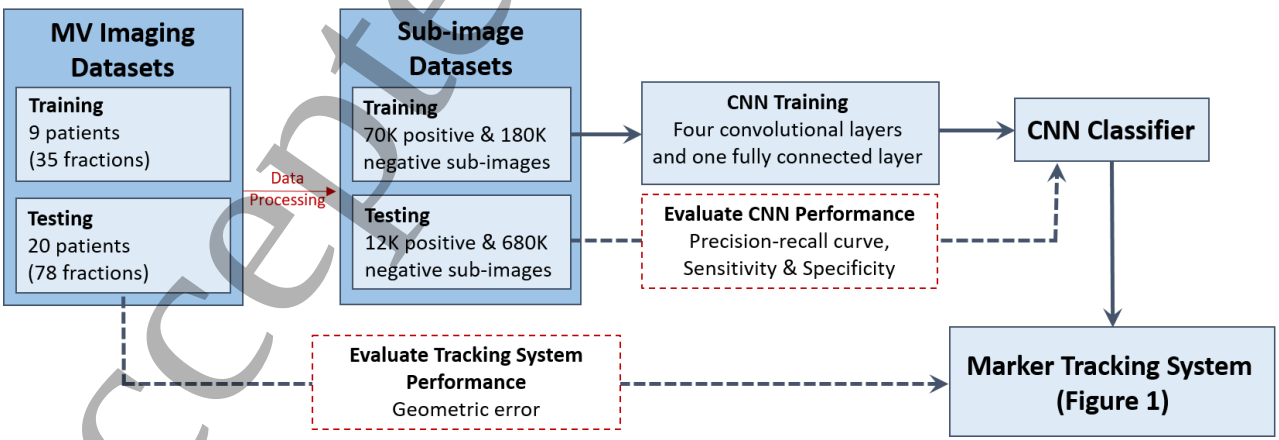


*Figure 2. Training and testing of the CNN classifier and the marker tracking system for marker classification.*

*2.2.2. Sub-image generation and pre-processing of data*

The process of extracting positive and negative sub-images from single MV images is illustrated in Figure 3. Each MV image was cropped to a 300 × 300 pixel region around the center pixel, containing markers and surrounding anatomy. Marker positions were determined by performing kV/MV triangulation between the KIM-measured marker motion on kV frames and the corresponding MV frames. The estimated marker positions from triangulation were projected on the MV images, and manual correction or exclusion was performed by a single researcher via visual inspection for inaccurate marker positions or those occluded by the MLC leaves. The marker positions were projected on the cropped region of the MV image, which was sub-divided into 21 × 21 pixel size regions to create positive and negative sub-images. These manually labelled marker positions also defined the ground truth when evaluating the performance of the marker tracking system (Section 2.4.2). Data augmentation was performed on the positive sub-images within the training imaging dataset to maximize possible marker orientations that may be encountered by the marker tracking software. Similarly, the background dataset contained a range of background situations including MLC edges, fully occluded MLC regions and unrestricted field of view (FOV) regions containing no markers. Pre-processing techniques were applied to the images to improve visibility of the markers: temporal filtering whereby three consecutive frames were averaged to reduce MV scatter noise in the images, median filtering, and normalization.

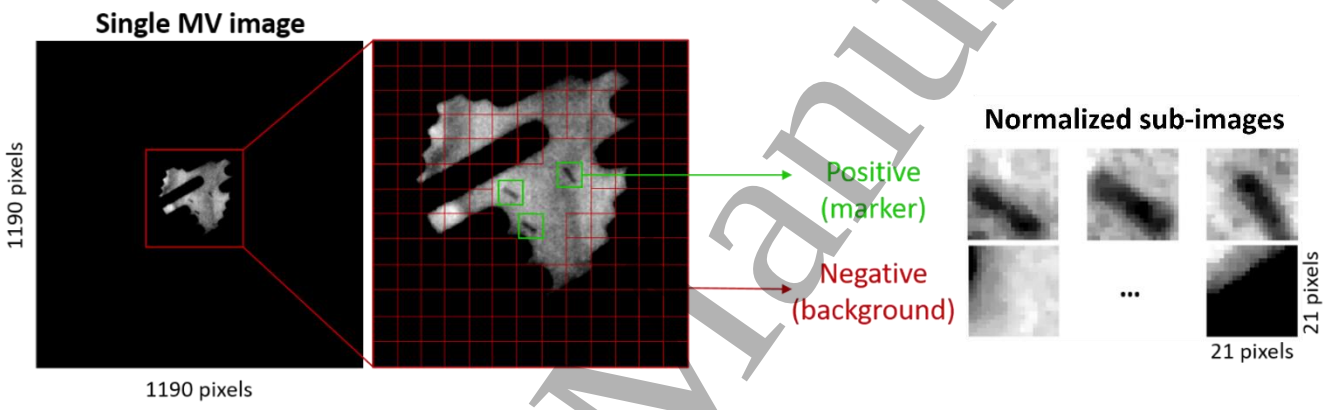


*Figure 3. Generation of normalized positive and negative sub-images in the training and testing datasets from a single MV image.*

*2.2.3. Network architecture*

The overall architecture of the CNN trained for marker detection using MV images is illustrated in Figure 4. It is comprised of 4 convolutional layers, 3 max-pooling layers and 1 fully connected layer. The first, second, third and fourth convolutional layers have 10, 16, 32 and 64 filters with size of 3 × 3, respectively. The stride and padding of all convolutional layers are equal to 1 pixel. Max-pooling is performed using a 2 × 2 window, with a stride of 2 pixels. Batch normalization layers and rectified linear units (ReLU) followed each of the convolutional layers, and were used to avoid unstable gradients, reduce initialization sensitivity on convergence and allow fast learning rates to accelerate the training of the CNN. The CNN classifier was implemented in MATLAB (Mathworks Inc., Natick, MA, USA) using the Deep Learning Toolbox. The CNN was trained using stochastic gradient descent with mini batch size, learning rate and momentum of 512, 0.0001 and 0.9, respectively. The cross entropy loss function was used to optimize the CNN during training.

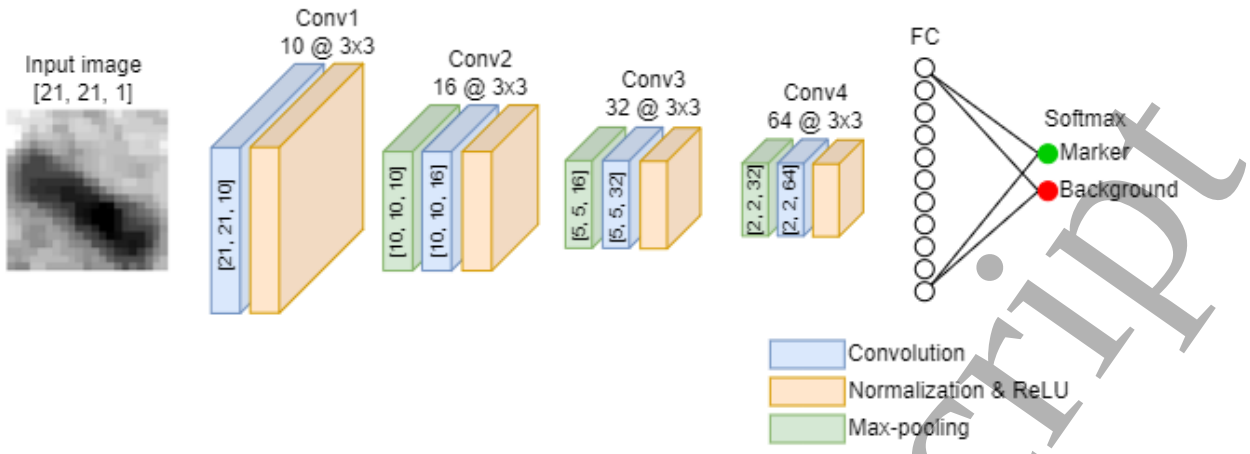


*Figure 4. Schematic of the CNN architecture, consisting of four convolutional layers (Conv) and one fully connected (FC) layer. The size of the output following each of the max-pooling and convolutional layers is provided in the square brackets.*

## 2.3. Automated marker tracking system

An automated marker tracking system was developed in MATLAB that uses the fully trained CNN classifier to track implanted markers using MV images (Figure 1). For each frame, the same pre-processing steps that were performed on the training data was performed: temporal filtering with three consecutive frames, median filtering and normalization. For each frame, tracking windows were generated for each marker, and sliding window classification was performed twice using the CNN: the first search tracking window was centered on the CT projected marker positions and the second was centered on the updated marker location from the initial classification.

Tracking windows with a size of 38 × 38 pixels centered on each individual marker were used to reduce the search area to improve the computational processing speed and efficiency of the marker tracking system. The size of each tracking window was large enough to account for deviation in marker position from the CT projected marker position, while remaining small enough that the number of sub-images required to be searched did not hinder computational processing time.

Sliding window classification was performed, using a 21 × 21 pixel sliding window and a two pixel step size, within each tracking window to localize marker positions. As the sliding window moves over the tracking window, each selected area was added to a group of sub-images and classified as either positive or negative using the CNN. Generally, the sub-image was classified as positive if more than 50% of the marker was visible since this is close to the limit that our researcher could manually identify for the ground truth observations. A logical array was created, recording the locations of the positive sub-images. For instances where two markers were visible in a single tracking window, k-means clustering was performed to identify the individual clusters, and the cluster further away from the estimated marker position was re-classified as negative (Supplementary Figure 1). The specificity of the system was improved by introducing a threshold, whereby the marker centroid of the positive regions was calculated only if more than five positive sub-images were recorded for a given marker. To further improve specificity, sliding window classification was performed twice for each MV frame, minimizing the probability of MLC edges being misclassified as positive. The x- and y-coordinates of each marker centroid were stored.

## 2.4. Performance evaluation

### *2.4.1. CNN classification accuracy*

All sub-images in the unseen testing dataset were classified using the CNN as positive or negative. The classification performance of the CNN was evaluated using a precision-recall curve (PRC) and calculation of the sensitivity (ratio between true positive classifications and all actual positives) and specificity (ratio between true negative

classifications and all actual negatives). The PRC plots the relationship between precision (ratio between true positive classifications and all instances identified as positive) and recall (i.e. sensitivity). The PRC plot was used, rather than the receiver operating characteristic (ROC) curve, to avoid misleading results due to the imbalance between the marker and background sub-images in the testing dataset [19]. The area under the PRC curve (AUC) was used to quantify CNN performance, where an AUC of 0.5 indicates random classification and an AUC of 1.0 indicates perfect classification.

*2.4.2. Marker tracking system accuracy*

The performance of the marker tracking system was evaluated on one representative fraction of each patient from the testing dataset. The tracking system accuracy was quantified by calculating the mean absolute and [1$^{st}$, 99$^{th}$] percentiles of the geometric tracking error in treatment beam co-ordinates in the lateral and superior-inferior (SI) directions. The geometric tracking error was defined as the accuracy (mean absolute difference) and precision (1 standard deviation (SD)) between the CNN detected marker location and the manually defined ground truth marker location.

*2.4.3. Computational processing time*

The computational processing time of the marker tracking system was evaluated using a desktop computer with a 3.40 GHz AMD Ryzen 9 5950X 16-Core Processor CPU with 128 GB RAM and a NIVIDIA GeForce® RTX 3090 GPU. The reported processing speed includes the time taken for the algorithm to load three consecutive frames for temporal filtering, perform image processing and perform two rounds of sliding window classification using the CNN classifier for three markers in a single MV image. The reported processing time was evaluated for the entire duration of a single patient fraction (repeated ten times) and the mean (± 1SD) processing time per frame was reported. Similarly, the mean (± 1SD) CNN segmentation time, including two rounds of sliding window classification, was reported per frame (with three implanted markers) and per marker. Real-time performance was defined in this paper as requiring less than 100 ms of latency, enabling real-time application for a 10 Hz cine MV acquisition rate.

## 3. Results

### 3.1. CNN classification accuracy

Figure 5 shows the results of the CNN classification performance in classifying prostate markers using MV images. The AUC for the CNN was 0.9937 indicating high predictive performance when classifying regions of the image as “background” and “marker”. The CNN had high sensitivity of 98.31% and high specificity of 99.87%.

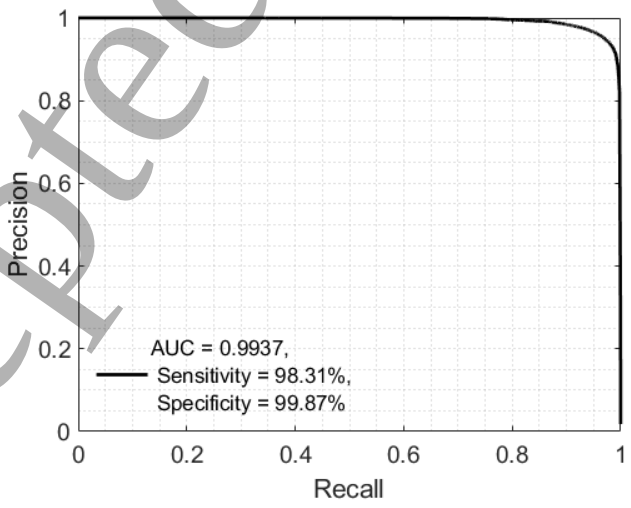


*Figure 5. Precision-Recall Curve (PRC) of the fully trained CNN for prostate marker classification in MV images.*

### 3.2. Marker tracking system accuracy

Figure 6 depicts a histogram of the mean geometric tracking error of the CNN on the unseen MV images. The mean absolute geometric tracking error was 0.30 ± 0.27 mm and 0.35 ± 0.31 mm in the lateral and SI directions, respectively. The [1st, 99th] percentiles of the error were [-1.03, 0.90] mm and [-1.12, 1.12] mm in the lateral and SI directions, respectively. Supplementary Table 1 provides patient-specific geometric tracking error.

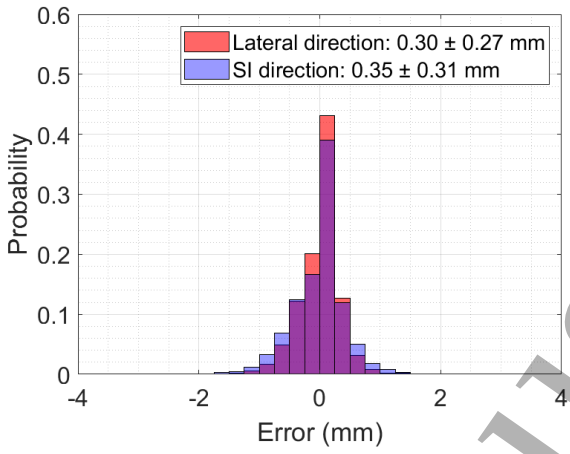


*Figure 6. Histogram showing the probability of geometric tracking error of the marker tracking system in 0.25 mm bins in the lateral (red) and SI (blue) directions.*

Figure 7 shows a comparison between the automatically tracked marker positions using the CNN classifier and the manually labelled ground truth for one patient fraction. As there were no obvious patient outliers (Supplementary Table 1), Patient 1 was chosen to provide an example of CNN tracking. The blank segments of the trace correspond to frames where all three markers were occluded by the MLC leaves, and therefore no tracking was reported. Figure 8 illustrates the CNN segmentation capabilities, including overlapping markers and markers at the MLC edges. When more than half of the marker was occluded by the MLC edge, the CNN was unable to segment it, as the algorithm was tuned for high specificity to avoid false positives.

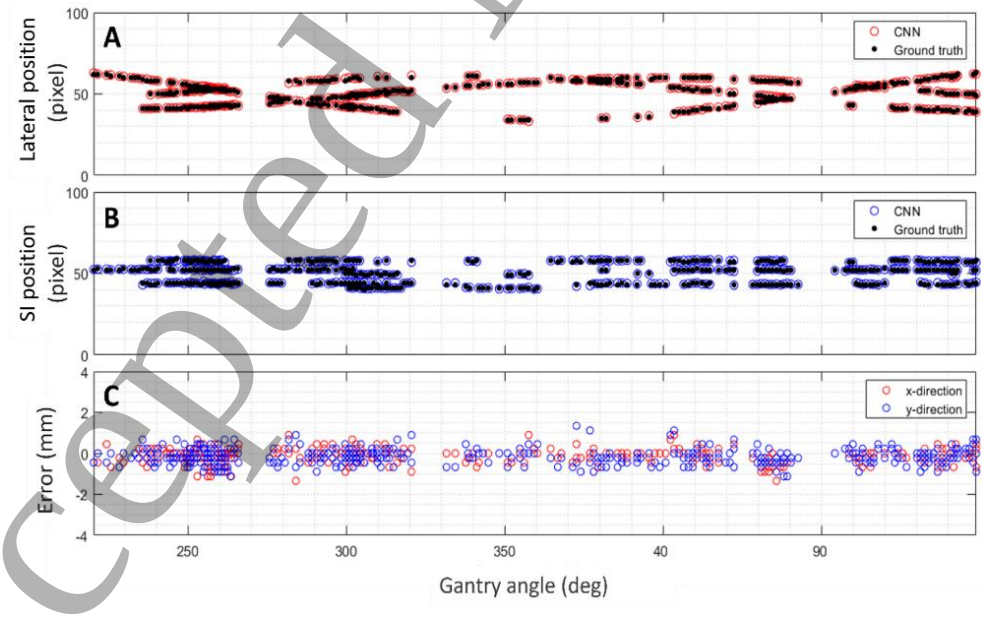


*Figure 7. Example of CNN tracking for one patient fraction. Comparison between CNN tracking and ground truth for individual markers in (A) lateral and (B) SI directions; and (C) CNN geometric tracking error in lateral (red) and SI (blue) directions.*

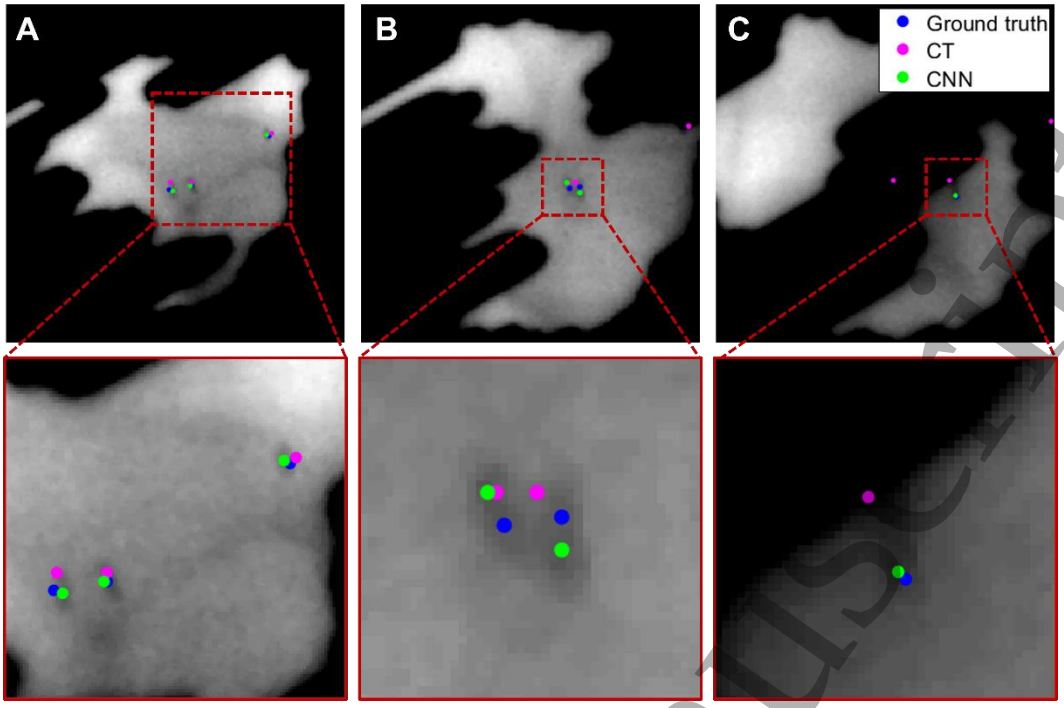


*Figure 8. Example MV frames showing CNN tracking capabilities of (A) all three markers in a single frame; (B) "overlapping markers" or two markers within a single tracking window; and (C) a partially occluded marker at the MLC edge.*

### 3.3. Computational processing time

Figure 9 shows a comparison between the processing time of the CPU and GPU systems and separates the CNN-based marker segmentation time from the total processing time. The mean computational processing time of the marker tracking algorithm evaluated on MV images with three markers was 99.3 ± 5.1 ms/frame (CPU) and 90.4 ± 0.8 ms/frame (GPU). The mean CNN segmentation time was 43.3 ± 0.5 ms/frame (CPU) and 35.9 ± 0.5 ms/frame (GPU) for three markers. The mean CNN segmentation time per marker was 14.4 ± 0.2 ms/marker (CPU) and 12.0 ± 0.2 ms/marker (GPU).

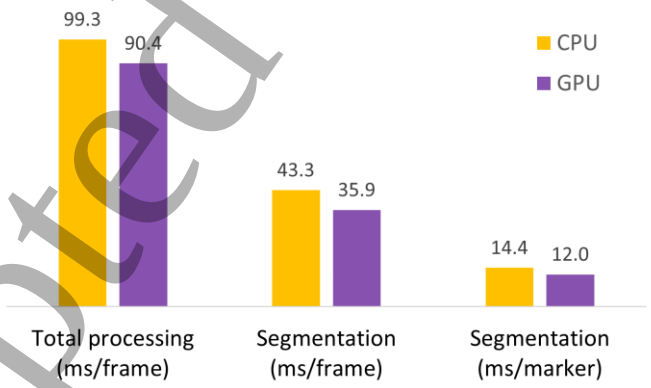


*Figure 9. Computational processing time comparison between CPU (yellow) and GPU (purple).*

## 4. Discussion

The purpose of this work was to develop and retrospectively evaluate the first MV-based real-time IGRT system that employs a deep learning framework to segment markers. Annotated datasets were created using cine MV images acquired from prostate cancer patients treated at multiple institutions and were used to train a binary CNN

classifier. This study demonstrated that this novel method performed equivalently to manual segmentation, achieving sub-millimetre accuracy and precision when evaluated on an unseen testing dataset from a separate institution. The tracking system overcame the inherent challenges associated with MV imaging that affect accurate target tracking: (1) the low contrast between the target and surrounding tissues in the MV imager, and (2) beam modulation imposed by the MLC leaves temporarily blocking radiation fields and inhibiting target visibility. Prior MV-based marker tracking studies have been somewhat successful in overcoming the low contrast challenge [9-12]; however, no studies to date have developed a robust method to overcome the challenge of beam modulation, particularly during VMAT, which is the predominant external beam delivery method for abdominal cancer sites [20]. In this work, we propose a method that overcomes limitations in previous MV-based marker tracking studies and is suitable for real-time applications.

Deep learning models (e.g. neural networks) have become well-researched in the field of image classification in recent years due to the high achievable classification accuracy. Wang *et al* (2021) performed a comparative analysis on the performance of a machine learning-based algorithm using an SVM and a deep learning-based algorithm using a CNN for image classification on different sized datasets. They concluded that, while the SVM classifier outperformed the CNN classifier on small datasets reporting accuracies of 86% (SVM) and 83% (CNN), the accuracy of the CNN classifier greatly improved to 98% when trained on a larger dataset, while the accuracy of the SVM classifier increased only slightly to 88% [15]. Previous work conducted by Mylonas *et al* (2019) involved training CNN classifiers to segment cylindrical and arbitrarily shaped markers in kV images. The cylindrical marker CNNs reported high sensitivity ranging from 98.87% to 99.42% and high specificity ranging from 97.99% to 99.32%. Similarly, the arbitrarily shaped marker CNN achieved sensitivity of 98.58% and specificity of 98.97% showing that CNN-based classification can be well-suited for tracking markers of different types and shapes [8]. In this work, we extended upon the deep learning framework using kV images developed by Mylonas *et al* to train a CNN classifier for detecting cylindrical shaped markers in MV images. Our trained CNN displayed high classification accuracy with an AUC of 0.99, high sensitivity of 98.31% and high specificity of 99.87%.

The marker tracking system presented in this work had a mean absolute geometric tracking error of 0.30 ± 0.27 mm and 0.35 ± 0.31 mm in the lateral and SI directions, respectively. The accuracy of this marker tracking method is comparable to manual marker localization and to the performance of previously developed MV-based marker tracking algorithms. Table 2 provides a summary of MV-based marker tracking algorithms reported in the literature and a comparison to the method presented in this study. Note that accuracy is not reported for these studies since they used different methods for acquiring the ground truth and different reported performance metrics, making a direct comparison difficult.

*Table 2. Summary of marker tracking approaches using clinically acquired MV images.*

| Study | Detection method | Treatment technique | Treatment site (# patients) | Marker shape | Processing time (ms/frame) |
|---|---|---|---|---|---|
| **Our approach** | **DL (CNN)** | **VMAT** | **Prostate (20)** | **Cylindrical** | **< 100 (CPU, GPU)** |
| Keall *et al* 2004 | Blob detection | Non-modulated | Phantom | Cylindrical | - |
| Park *et al* 2009 | Blob detection | Non-modulated | Phantom, Lung (5) | Cylindrical | < 100 |
| Mao *et al* 2008 | Template matching | IMRT | Phantom | Cylindrical | < 100 |
| Lin *et al* 2013 | ML | IMRT | Prostate (2) | Cylindrical | 128 (Frame 1), 23 (Subsequent) |
| Azcona *et al* 2013 | Template matching | VMAT | Prostate (5) | Cylindrical | ~1 sec/frame [21] |
| Fledelius *et al* 2014 | Template matching | Non-modulated | Liver (13) | Cylindrical | 5.5 ms/marker |

Park *et al* (2009) developed a marker tracking algorithm combining marker registration and blob filter detection. They reported 100% marker detection rate with the phantom, and sub-millimetre accuracy in the lateral and SI directions when compared to manually labelled markers for both phantom and five lung SBRT patients. While this method reported processing speeds within 100 ms/frame, it was susceptible to misidentification due to two markers being in proximity, marker occlusion and organ deformation in patient studies, excluding it's use as a real-time motion management tool [12]. Template matching allows for more complex objects to be identified compared to more rudimentary levels of object detection such as masking and blob detection. A significant number of studies have investigated the use of template matching to track markers using kV [22], MV [10, 21] and combined kV/MV [13, 14, 23] images. Azcona *et al* (2013) developed a template matching based marker tracking algorithm and employed a 3D marker position prediction model to project the expected marker locations on each MV image [10]. This approach was unique in that it leveraged CT planning positions to generate a predictive model, which was iteratively updated using each successful marker localization, enabling the algorithm to work with modulated VMAT fields. However, the algorithm was unable to accurately segment cases involving overlapping markers and was subject to false positives that the authors deemed "bad detections" and had to be manually removed. In a subsequent study they reported a processing speed (including template matching search and mask generation) of approximately 1 second/image; too slow for real-time applications [21].

Due to the large number of templates often required to cover a range of marker orientations, angles and positions, template matching-based marker tracking methods are often limited by a slow processing time, preventing them from being used in real-time applications [10, 21]. Lin *et al* (2013) somewhat improved upon earlier template matching methods by developing an algorithm that uses artificial intelligence techniques to track markers using MV images. Discriminant analysis was performed to initialize marker positions in the first frame and initialization and tracking on sequential frames was performed using a mean-shift feature space analysis. They reported an average root-mean-square-error of 1.9 and 2.1 pixels for 2 patients (0.26 mm/pixel) compared to a manually labelled ground truth. They employed a temporal correlation approach, exploiting the fact that markers move very little between frames, which enabled fast processing speeds (128 ms for the first frame and 23 ms/frame for subsequent frames). However, this limited the ability of the algorithm to adequately manage modulated fields. A further limitation was the requirement of a learning period, which was performed during the first treatment fraction for each patient. This required manual selection of training samples via visual inspection, which can be a labor intensive and time-consuming process and limited the ability to perform real-time marker tracking for the entirety of the patients treatment [11].

Our proposed deep learning marker tracking system overcame the limitations highlighted in previous studies and presents a novel motion management solution suitable for modulated radiotherapy. The TG-324 respiratory motion management survey reported the most common delivery type between 502 centers and found that 81% primarily use VMAT, 7.9% primarily use intensity modulated radiation therapy (IMRT) and the remainder primarily use non-modulated radiation therapy [20]. The biggest challenge in marker tracking using MV images acquired during VMAT treatment is beam modulation imposed by the MLCs. This includes full or partial occlusion of the markers hindering the segmentation ability of the algorithm and misidentification of the high contrast MLC edges as markers (leading to higher false positives). Supplementary Figure 2 illustrates the complex MLC configurations during the VMAT treatment of one patient fraction in this study, highlighting how the MLC leaves may block the markers during treatment. Azcona *et al* (2013) is the only study in current literature that evaluated their MV marker tracking method using images acquired during VMAT treatment delivery and reported some success. We employed a similar initialization method that was used in their work, where the CT-estimated marker positions were projected on each MV frame. This uniquely allows both our approaches to be used in VMAT treatments with highly modulated fields, as the marker can be re-segmented following a period of obstruction by the MLC leaves. Azcona *et al* (2013) reported several challenges rendering their method inappropriate for adaptation to real-time IGRT: inability to segment overlapping markers, false positives, and slow computational processing time. Our algorithm overcame each of these challenges. K-means clustering was employed for overlapping markers to separate the positive clusters into individual objects (Supplementary Figure 1). We used this additional k-means clustering method to

separate instances of markers while maintaining the fast segmentation requirements for real-time application. Direct instance segmentation (e.g. U-Net) would likely require deployment with specialized computer hardware to work in real-time and it would require a large number of training mask labels, which is a labor intensive acquisition process. The algorithm was specifically tuned for high specificity to avoid false positives by ensuring the training dataset of background sub-images contained a large variety of MLC edges and that a positive sub-image threshold was applied. The compact CNN architecture enabled a mean processing time of 99.3 ± 5.1 ms/frame (CPU) and 90.4 ± 0.8 ms/frame (GPU) for three markers, which is fast enough to be considered for real-time applications. More than half of the computational processing time is due to image loading and pre-processing steps (Figure 9). The mean CNN segmentation time per marker was 14.4 ± 0.2 ms (CPU) and 12.0 ± 0.2 ms (GPU).

A shared limitation between this study and previous MV-based marker tracking studies is that the algorithm can only report real-time motion while the markers are visible in the FOV and are not occluded by MLC leaves. This issue has been somewhat resolved by direct methods proposed by Ma *et al* (2009) and Happersatt *et al* (2019). Ma *et al* developed a 4D inverse planning strategy that focused on modifying MLC fields during planning to maximize marker visibility during treatment [24]. Happersatt *et al* developed an Eclipse plug-in that automatically modifies MLC configurations at control points to expose obscured markers during VMAT [25]. These methods were found to improve the visibility of the implanted markers without compromising the final dose distribution. In addition to MLC configuration modification strategies, Ma *et al* (2018) developed a model to recommend optimal marker insertion locations to improve visibility of markers using MV imaging during prostate treatment [25].

Previous studies have been somewhat successful in solving the challenge of low contrast in MV images [10-12]. A commonly used method of enhancing marker visualization in MV images is to apply a Laplacian of Gaussian filter in the spatial domain of each image [12]. A limitation to this approach is that non-marker regions may be highlighted resulting in false positives, therefore small regions of interest are commonly used to minimize the number of false positives reported [10]. In our study, the contrast was enhanced by applying temporal and median filtering and normalizing each frame prior to performing CNN-based classification. The temporal filtering step involved a rolling average of three consecutive frames, which reduced noise from the MV scatter. The gantry rotation between frames in this study was minimal (within 0.5 degrees), therefore we expect minimal blurring effect in the processed images. The subsequent improvements in image quality and marker visibility were sufficient for the algorithm to detect and segment visible markers. In addition, the CNN was trained using MV image data acquired from multiple institutions with different delivered beam energies (10 MV FFF and 6 MV) and different EPID models (Varian as1000 and as1200). The variety of data included in the training dataset enabled the CNN to be trained on a wide range of MV images with differing contrast, allowing the model to learn features to extract from the low contrast MV images. This improved the robustness and transferrability of the model between institutions. Furthermore, recent research has been dedicated towards the improvement of detective quantum efficiency and contrast-to-noise ratio of the MV imager, which include the multi-layer imager (MLI) design [26-29], pixelated scintillation [30] and advanced scintillation materials [31]. Harris *et al* (2021) demonstrated a 41.7% increase in marker tracking efficiency (number of frames that successfully track markers compared to maximum number of trackable frames) when using a novel multilayer MV imager [26].

The zoomed-in sub-figures depicted in Figure 8 were chosen, in part, to reflect the uncertainty in our ground truth estimation. There were many frames to manually segment, therefore the marker position is sometimes not directly in the center of the marker and our ground truth estimate has an associated uncertainty. Lin *et al* (2013) used a similar ground truth method for evaluating the performance of their approach. They averaged the manually labelled ground truth positions from six different researchers. They reported the standard deviation between the six researchers was 2.3 pixels (0.6 mm) and 2.6 pixels (0.68 mm) for the two patients included in their study. The diameter × length of the markers used in their study was 1.2 × 3.0 mm [11]. In our study, a single researcher was tasked with manually labelling the marker positions from the 20 patients in the testing dataset. With similar marker sizes and ground truth estimation methods used in this work compared to Lin *et al* (2013), we expect a similar associated uncertainty in our own ground truth estimation.

The high marker detection accuracy of markers using MV images presented in this work shows the potential for real-time marker tracking using the MV imager as a robust, cost-effective and widely accessible approach. The CNN was trained using gold cylindrical markers between 3 – 5 mm in length. This method could be expanded to train the CNN on a larger variety of inserted marker types and sizes to improve transferability between institutions. The mean segmentation time of the current system was within 100 ms/frame when measured on the CPU and GPU systems, which is fast enough to be considered for real-time applications. Future work will involve integrating the developed software with existing clinical systems and developing a quality assurance protocol, as well as more extensive testing of the algorithm with different marker types, orientations, overlapping markers, and number of implanted markers in a phantom to improve the robustness of the algorithm. Furthermore, this method has the potential to extend to marker tracking at other anatomical sites, such as the liver, lung and pancreas. These treatment sites are more mobile, so marker position will deviate from the projected CT planning position. To account for this, future work may aim to investigate variable tracking window sizes or an internal-external correlation model to improve the accuracy of the initial "estimation".

## 5. Conclusions

A real-time IGRT system has been developed, based on a fully trained CNN classifier for intrafraction monitoring of implanted markers as surrogates for tumour motion in prostate SBRT. The high classification performance on unseen images demonstrates that the CNN can successfully identify markers on MV images acquired during VMAT treatments. Furthermore, the sub-millimetre accuracy and precision of the tracking system demonstrates that it can be feasibly used for MV-based marker tracking with the potential to be considered for real-time applications.

## Acknowledgements

The work presented in this paper was supported by a Cancer Australia grant. The authors would like to thank the TROG SPARK trial (NCT02397317) patient participants and the coordinators for their management of this trial. The protocol for this clinical trial was approved by the Hunder New England Human Research Ethics Committee (HREC). All participants were over 18 years old and provided written informed consent to participate in the study and for trial outcomes to be published in journals. The authors would like to thank the following institutions for their involvement in this multi-institutional clinical trial: Westmead Hospital, NSW, Australia; Peter MacCallum Cancer Centre, Melbourne, VIC, Australia; and Calvary Mater Newcastle, NSW, Australia. The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements.

## Conflicts of interest

Authors A. Mylonas, D.T. Nguyen, J. Booth and P. Keall are inventors on a filed patent application based on this technology. Authors P. Keall and D.T. Nguyen are directors and shareholders, and J. Martin is the Chief Medical Officer, of the company that owns this patent application.

## References

1. Morgan, S.C., et al., *Hypofractionated Radiation Therapy for Localized Prostate Cancer: Executive Summary of an ASTRO, ASCO and AUA Evidence-Based Guideline.* J Urol, 2019. **201**(3): p. 528-534.
2. Mahase, S.S., et al., *Trends in the Use of Stereotactic Body Radiotherapy for Treatment of Prostate Cancer in the United States.* JAMA Netw Open, 2020. **3**(2): p. e1920471.
3. Kotte, A.N., et al., *Intrafraction motion of the prostate during external-beam radiation therapy: analysis of 427 patients with implanted fiducial markers.* Int J Radiat Oncol Biol Phys, 2007. **69**(2): p. 419-25.
4. Winkel, D., et al., *Adaptive radiotherapy: The Elekta Unity MR-linac concept.* Clin Transl Radiat Oncol, 2019. **18**: p. 54-59.
5. Tree, A.C., et al., *Intensity-modulated radiotherapy versus stereotactic body radiotherapy for prostate cancer (PACE-B): 2-year toxicity results from an open-label, randomised, phase 3, non-inferiority trial.* Lancet Oncol, 2022.

6. Mantz, C., *A Phase II Trial of Stereotactic Ablative Body Radiotherapy for Low-Risk Prostate Cancer Using a Non-Robotic Linear Accelerator and Real-Time Target Tracking: Report of Toxicity, Quality of Life, and Disease Control Outcomes with 5-Year Minimum Follow-Up.* Front Oncol, 2014. **4**: p. 279.
7. Hewson, E.A., 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): p. 4725-4737.
8. Mylonas, A., et al., *A deep learning framework for automatic detection of arbitrarily shaped fiducial markers in intrafraction fluoroscopic images.* Med Phys, 2019. **46**(5): p. 2286-2297.
9. Keall, P.J., et al., *On the use of EPID-based implanted marker tracking for 4D radiotherapy.* Med Phys, 2004. **31**(12): p. 3492-9.
10. Azcona, J.D., et al., *Development and clinical evaluation of automatic fiducial detection for tumor tracking in cine megavoltage images during volumetric modulated arc therapy.* Med Phys, 2013. **40**(3): p. 031708.
11. Lin, W.Y., et al., *Real-time automatic fiducial marker tracking in low contrast cine-MV images.* Med Phys, 2013. **40**(1): p. 011715.
12. Park, S.J., et al., *Automatic marker detection and 3D position reconstruction using cine EPID images for SBRT verification.* Med Phys, 2009. **36**(10): p. 4536-46.
13. Mao, W., et al., *A fiducial detection algorithm for real-time image guided IMRT based on simultaneous MV and kV imaging.* Med Phys, 2008. **35**(8): p. 3554-64.
14. Fledelius, W., et al., *Real-time segmentation of multiple implanted cylindrical liver markers in kilovoltage and megavoltage x-ray images.* Phys Med Biol, 2014. **59**(11): p. 2787-800.
15. Wang, P., E. Fan, and P. Wang, *Comparative analysis of image classification algorithms based on traditional machine learning and deep learning.* Pattern Recognition Letters, 2021. **141**: p. 61-67.
16. Keall, P., et al., *Stereotactic prostate adaptive radiotherapy utilising kilovoltage intrafraction monitoring: the TROG 15.01 SPARK trial.* BMC Cancer, 2017. **17**(1): p. 180.
17. Keall, P., et al., *Real-Time Image Guided Ablative Prostate Cancer Radiation Therapy: Results From the TROG 15.01 SPARK Trial.* Int J Radiat Oncol Biol Phys, 2020. **107**(3): p. 530-538.
18. Poulsen, P.R., et al., *Three-dimensional prostate position estimation with a single x-ray imager utilizing the spatial probability density.* Phys Med Biol, 2008. **53**(16): p. 4331-53.
19. Saito, T. and M. Rehmsmeier, *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets.* PLoS ONE, 2015. **10**(3): p. e0118432.
20. Ball, H.J., et al., *Results from the AAPM Task Group 324 respiratory motion management in radiation oncology survey.* J Appl Clin Med Phys, 2022. **23**(11): p. e13810.
21. Azcona, J.D., et al., *Automatic prostate tracking and motion assessment in volumetric modulated arc therapy with an electronic portal imaging device.* Int J Radiat Oncol Biol Phys, 2013. **86**(4): p. 762-8.
22. Bertholet, J., et al., *Fully automatic segmentation of arbitrarily shaped fiducial markers in cone-beam CT projections.* Phys Med Biol, 2017. **62**(4): p. 1327-1341.
23. Mao, W., et al., *Image-guided radiotherapy in near real time with intensity-modulated radiotherapy megavoltage treatment beam imaging.* Int J Radiat Oncol Biol Phys, 2009. **75**(2): p. 603-10.
24. Ma, Y., et al., *Four-dimensional inverse treatment planning with inclusion of implanted fiducials in IMRT segmented fields.* Med Phys, 2009. **36**(6): p. 2215-21.
25. Ma, T., J. Kilian-Meneghin, and L.K. Kumaraswamy, *Recommendation of fiducial marker implantation for better target tracking using MV imager in prostate radiotherapy.* J Appl Clin Med Phys, 2018. **19**(5): p. 389-397.
26. Harris, T.C., et al., *Improvements in beam's eye view fiducial tracking using a novel multilayer imager.* Phys Med Biol, 2021. **66**(15).
27. Harris, T.C., et al., *Clinical translation of a new flat-panel detector for beam's-eye-view imaging.* Phys Med Biol, 2020. **65**(22): p. 225004.
28. Hu, Y.H., et al., *Physics considerations in MV-CBCT multi-layer imager design.* Phys Med Biol, 2018. **63**(12): p. 125016.
29. Myronakis, M., et al., *A novel multilayer MV imager computational model for component optimization.* Med Phys, 2017. **44**(8): p. 4213-4222.
30. Star-Lack, J., et al., *A piecewise-focused high DQE detector for MV imaging.* Med Phys, 2015. **42**(9): p. 5084-99.
31. Hu, Y.H., et al., *Characterizing a novel scintillating glass for application to megavoltage cone-beam computed tomography.* Med Phys, 2019. **46**(3): p. 1323-1330.