

# Experimental comparison of linear regression and LSTM motion prediction models for MLC-tracking on an MRI-linac

Elia Lombardo<sup>1</sup> | Paul Z. Y. Liu<sup>2,3</sup> | David E. J. Waddington<sup>2,3</sup> | James Grover<sup>2,3</sup> |  
 Brendan Whelan<sup>2,3</sup> | Esther Wong<sup>3</sup> | Michael Reiner<sup>1</sup> | Stefanie Corradini<sup>1</sup> |  
 Claus Belka<sup>1,4,5</sup> | Marco Riboldi<sup>6</sup> | Christopher Kurz<sup>1</sup> | Guillaume Landry<sup>1</sup> |  
 Paul J. Keall<sup>2,3</sup>

<sup>1</sup>Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, Germany

<sup>2</sup>Image X Institute, University of Sydney Central Clinical School, Sydney, New South Wales, Australia

<sup>3</sup>Department of Medical Physics, Ingham Institute of Applied Medical Research, Liverpool, New South Wales, Australia

<sup>4</sup>German Cancer Consortium (DKTK), partner site Munich, a partnership between DKFZ and LMU University Hospital Munich, Munich, Germany

<sup>5</sup>Bavarian Cancer Research Center (BZKF), Munich, Germany

<sup>6</sup>Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Garching, Germany

## Correspondence

Elia Lombardo, Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, Germany.  
 Email: [elia.lombardo@med.uni-muenchen.de](mailto:elia.lombardo@med.uni-muenchen.de)

Guillaume Landry and Paul J. Keall contributed equally as senior author.

## Funding information

Deutsche Forschungsgemeinschaft; Cancer Institute NSW

## Abstract

**Background:** Magnetic resonance imaging (MRI)-guided radiotherapy with multileaf collimator (MLC)-tracking is a promising technique for intra-fractional motion management, achieving high dose conformity without prolonging treatment times. To improve beam-target alignment, the geometric error due to system latency should be reduced by using temporal prediction.

**Purpose:** To experimentally compare linear regression (LR) and long-short-term memory (LSTM) motion prediction models for MLC-tracking on an MRI-linac using multiple patient-derived traces with different complexities.

**Methods:** Experiments were performed on a prototype 1.0 T MRI-linac capable of MLC-tracking. A motion phantom was programmed to move a target in superior-inferior (SI) direction according to eight lung cancer patient respiratory motion traces. Target centroid positions were localized from sagittal 2D cine MRIs acquired at 4 Hz using a template matching algorithm. The centroid positions were input to one of four motion prediction models. We used (1) a LSTM network which had been optimized in a previous study on patient data from another cohort (*offline* LSTM). We also used (2) the same LSTM model as a starting point for continuous re-optimization of its weights during the experiment based on recent motion (*offline+online* LSTM). Furthermore, we implemented (3) a continuously updated LR model, which was solely based on recent motion (*online* LR). Finally, we used (4) the last available target centroid without any changes as a baseline (no-predictor). The predictions of the models were used to shift the MLC aperture in real-time. An electronic portal imaging device (EPID) was used to visualize the target and MLC aperture during the experiments. Based on the EPID frames, the root-mean-square error (RMSE) between the target and the MLC aperture positions was used to assess the performance of the different motion predictors. Each combination of motion trace and prediction model was repeated twice to test stability, for a total of 64 experiments.

**Results:** The end-to-end latency of the system was measured to be  $(389 \pm 15)$  ms and was successfully mitigated by both LR and LSTM models. The *offline+online* LSTM was found to outperform the other models for all investigated motion traces. It obtained a median RMSE over all traces of  $(2.8 \pm 1.3)$  mm, compared to the  $(3.2 \pm 1.9)$  mm of the *offline* LSTM, the  $(3.3 \pm 1.4)$  mm

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

of the *online* LR and the  $(4.4 \pm 2.4)$  mm when using the no-predictor. According to statistical tests, differences were significant ( $p$ -value  $< 0.05$ ) among all models in a pair-wise comparison, but for the *offline* LSTM and *online* LR pair. The *offline+online* LSTM was found to be more reproducible than the *offline* LSTM and the *online* LR with a maximum deviation in RMSE between two measurements of 10%.

**Conclusions:** This study represents the first experimental comparison of different prediction models for MRI-guided MLC-tracking using several patient-derived respiratory motion traces. We have shown that among the investigated models, continuously re-optimized LSTM networks are the most promising to account for the end-to-end system latency in MRI-guided radiotherapy with MLC-tracking.

#### KEYWORDS

linear regression, long short-term memory, MLC-tracking, motion prediction, MRI-linac, respiratory motion

## 1 | INTRODUCTION

Magnetic resonance imaging (MRI) guided radiotherapy (MRIgRT) offers high soft-tissue contrast visualization and the opportunity to adapt to changes in patient anatomy prior and during irradiation.<sup>1,2</sup> MRI-linac systems, which are linear accelerators with an embedded MRI unit, are increasingly being used clinically over the past years.<sup>3</sup> To adapt to intra-fractional changes, for instance due to respiratory motion, current clinical systems rely on motion monitoring<sup>4</sup> with gated beam delivery.<sup>5</sup> In this type of treatment, the irradiation target is visualized in real-time using cine MRI and the beam is automatically stopped if the target exits a pre-defined area, thus recovering the conformality of static treatments and avoiding an increase of dose to healthy tissues surrounding the tumor. A disadvantage of this approach are the increased treatment times, with duty cycle efficiencies of 20% or 50% having been reported in clinics, depending on patient compliance.<sup>6,7</sup> An alternative approach with comparable dose conformality but increased treatment efficiency is multileaf-collimator (MLC)-tracking, during which the MLC aperture is continuously shifted to follow the target motion.<sup>8</sup> However, it has been shown that a factor which is critical to the accuracy of MRIgRT with MLC-tracking is the system latency.<sup>9</sup>

The system latency is defined as the time lag between the physical motion of the target and the execution of beam adaptation, which in the case of MLC-tracking is the time when the MLC leaves reach their desired positions. According to the AAPM Task Group 264, a latency  $\leq 500$  ms is necessary to meet the definition of real-time motion compensation.<sup>10</sup> For MRI-linacs capable of MLC-tracking it has been experimentally measured to range from 205 ms to 411 ms,<sup>9,11</sup> mainly depending on the acquisition frequency of the cine MRIs. To overcome the system latency, temporal prediction models can be

used. Over the past decade, several respiratory motion prediction models have been implemented in-silico for MRIgRT<sup>12–14</sup> or RT in general.<sup>15–17</sup> In a comparative computational study, Jöhl et al. used 93 respiratory motion traces to show that among 18 motion predictors ranging from Kalman filters to artificial neural networks, linear regression (LR) models were the best candidates for respiratory motion prediction for various time horizons and noise levels.<sup>17</sup> When computationally possible, models were retrained at every time step (i.e., *online*) as this approach had been shown to improve performance. Recently, a class of machine learning algorithms called long short-term memory (LSTM) networks, which is ideally suited to deal with sequential input data, has been shown to be very promising for motion prediction both in RT<sup>18,19</sup> and in MRIgRT.<sup>20,21</sup> Specifically, LSTMs were shown to outperform LR models for the prediction of superior-inferior (SI) target centroid positions based on patient data acquired on a 0.35 T MRI-linac.<sup>20</sup>

While all aforementioned studies were in-silico, to the best of our knowledge there are four studies which experimentally investigated motion prediction for MLC-tracking during MRIgRT. In an early phantom study by Yun et al.<sup>22</sup>, it was shown that motion prediction using artificial neural networks and sinusoidal motion led to MLC-tracking with similar dosimetric accuracy as in the static scenario. Uijtewaal et al.<sup>23</sup> showed that an *online* LR can compensate the latency for the delivery of intensity modulated radiotherapy (IMRT) plans with MLC-tracking to a phantom moving with Lujan motion ( $\cos^4$ ). In two follow up studies with the same motion predictor, they used Lujan motion and additionally one patient-derived motion trace to investigate MLC-tracking with VMAT plans<sup>24</sup> and with a hybrid 2D/4D-MRI methodology.<sup>25</sup> A limitation of all four studies is that either sinusoidal or a single patient-derived trace was used.

This study aimed at experimentally validating the in-silico comparison of LSTM and LR motion prediction models<sup>20</sup> for MLC-tracking with a prototype MRI-linac. Compared to previous MRIgRT studies with MLC-tracking and motion prediction,<sup>22–25</sup> this was the first time multiple motion traces with different complexities were used and conventional and machine learning based predictors were compared. Specifically, we compared models trained on retrospective data (i.e., *offline*) with *online* models updated in real-time during the experiments. The performance of the motion predictors with the different motion traces was evaluated in terms of geometric accuracy of the MLC-tracking.

## 2 | METHODS

### 2.1 | Experimental setup

All experiments were performed on a prototype MRI-linac system featuring a 1.0 T open bore magnet (Agilent, UK) with a control system based on a Magnetom Avanto spectrometer (Siemens, Erlangen, Germany) and a 6 MV industrial linac (Linatron, Varex Imaging, Utah, USA). The radiation beam generated by the linac is aligned with the  $B_0$  field (fixed beam line, no gantry) and shaped by a MLC with 120 leaves (Millennium, Varian, Palo Alto, California, USA).<sup>26</sup> The motion for the experiments was executed with the MRI-compatible Quasar phantom (Modus Medical Devices, Ontario, Canada), positioned at a source-to-surface distance of 2.4 m. The phantom (same as in Liu et al.<sup>9</sup>) contained a single MRI-visible target and was placed inside the bore such that the target was located at the isocenter. During the experiments, the target was moved by a motor in SI direction according to the provided motion traces (see Section 2.2). During irradiation, the target was imaged and localized as described in Section 2.3. The extracted target positions were given as input to one of the motion prediction models (Section 2.4). Prior to the motion prediction experiments, a sinusoidal trace was tracked three times to characterize the end-to-end latency of the system, analogously to Liu et al.<sup>9</sup>: A sinusoidal was fitted to the centroid positions of the targets and apertures (moved by the no-predictor, see Section 2.4) and the latency was calculated as the time difference between the two fits. We then performed motion prediction with four models on eight motion traces. Each experiment was repeated twice to test the stability of the models, for a total of 64 experiments. MLC-tracking driven by the model predictions was used to compensate for the observed motion (Section 2.5) and an electronic portal imaging device (EPID) was used to quantify the geometric accuracy of tracking when using the different models (Section 2.6). The overall experimental setup is shown in Figure 1.

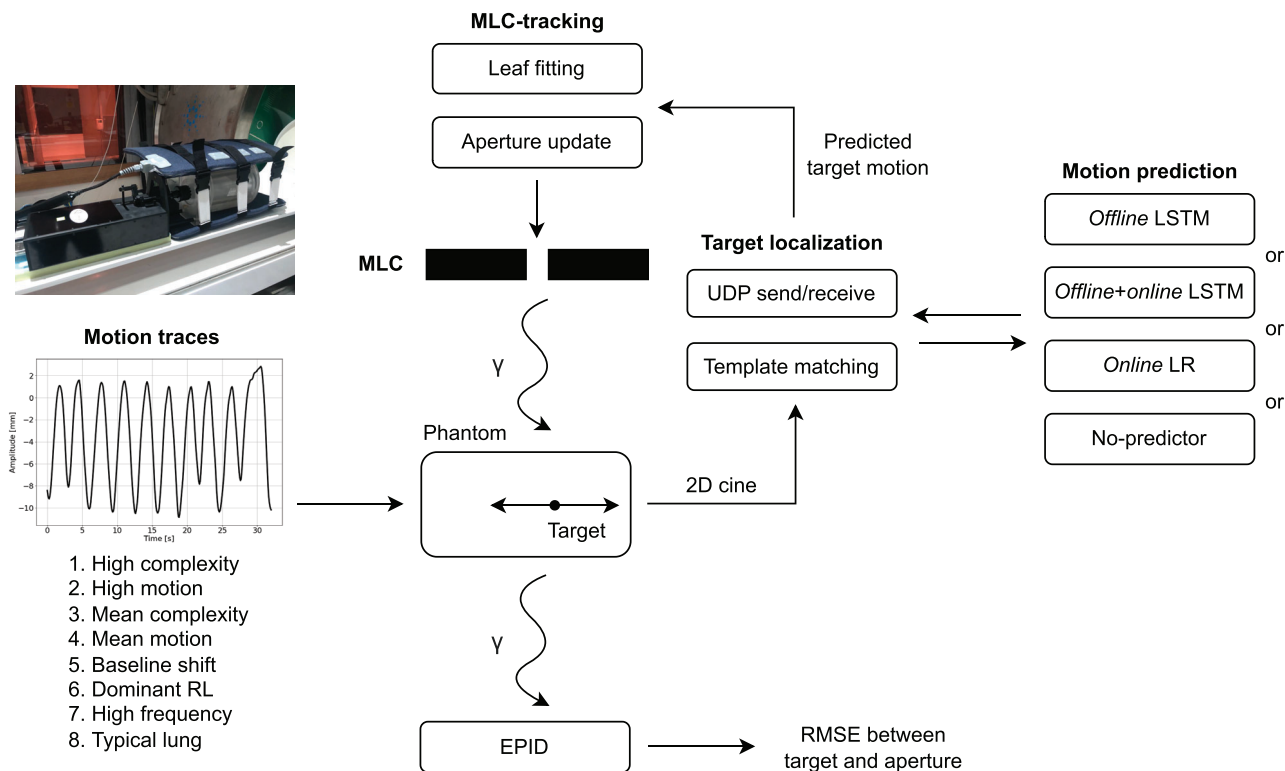
### 2.2 | Motion traces

In this study, eight publicly available motion traces previously exploited in a multi-institutional markerless lung target tracking study were used.<sup>27</sup> The traces were obtained from seven different lung tumor patients and feature different motion amplitude, complexity and frequency, all factors known to influence the accuracy of tracking. Four traces were taken from a clinical study using measurements of the centroid position of implanted Calypso beacons.<sup>28</sup> These 3D motion traces were originally acquired at 10 Hz and represent motion with high complexity, high motion amplitude, mean complexity and mean motion amplitude ([https://github.com/MarcoMueller-MCT/AAPM\\_GrandChallenge\\_MATCH/tree/master](https://github.com/MarcoMueller-MCT/AAPM_GrandChallenge_MATCH/tree/master)). The other four traces were acquired during treatments with a Cyberknife Synchrony system (Accuray Incorporated, Sunnyvale, California, USA) in 3D at a sampling rate of 25 Hz.<sup>29</sup> These traces include evident baseline shifts, right-left (RL) dominant, high frequency and typical lung target motion (<https://cloudstor.aarnet.edu.au/plus/index.php/s/iHz0aoTGBho3yu2?path=%2FLung%2F4DLungTrajectories>). Additionally to the patient traces, a sinusoidal trace with an amplitude of 20 mm and a period of 7.5 s was used to characterize the end-to-end latency of the system.

During the experiments, the SI component of the traces was used to rigidly move the target but for the “baseline shift” and the “dominant RL” traces, for which the lateral component was used. Independently of the component used, the target was always moved in SI direction. For each trace, motion was executed for about 2 min, however, during analysis, we did not use the first and last 30 s of data to exclude for instance buffering of the motion prediction models (i.e., the time needed until enough input positions are accumulated to start prediction) or the time for starting the radiation/EPID. The motion characteristics of the remaining 1 min of each trace, which was effectively used to assess the tracking accuracy, are shown in Table 1. While the name of the traces was the same as in Mueller et al.,<sup>27</sup> the period is slightly different as a different subset of each trace has been used. Also the displacement of each trace is different as in our experiments we re-scaled each trace to have a peak-to-peak amplitude over the entire trace of 30 mm to avoid very small MLC motion (arising from the large source-to-surface distance) and limitations of the EPID (spatial and temporal resolution).

### 2.3 | Imaging and localization

The moving target was visualized in real-time using cine MRI. Sagittal 2D slices were acquired using a balanced



**FIGURE 1** Experimental setup for MRI-guided MLC-tracking with motion prediction. A target inside a 1D motion phantom (photo on the top left) is moved following one of the eight patient respiratory motion traces. The moving target is imaged at 4 Hz using cine MRI. The target centroid position is then extracted with template matching and sent to the motion prediction model. The model outputs the future centroid position which is used to shift the MLC aperture in real-time. Finally, an EPID is used to characterize the geometric accuracy of MLC-tracking. EPID, electronic portal imaging device; MLC, multileaf collimator; MRI, Magnetic resonance imaging.

**TABLE 1** Motion characteristics of the eight patient traces used for the experiments.

Motion trace	IQR of displacement [mm]	Period [s]
High complexity	7.3	5.3
High motion	8.0	4.5
Mean complexity	7.4	4.0
Mean motion	10.6	2.6
Baseline shift	6.9	3.3
Dominant RL	9.8	4.2
High frequency	13.6	2.7
Typical lung	6.0	5.4

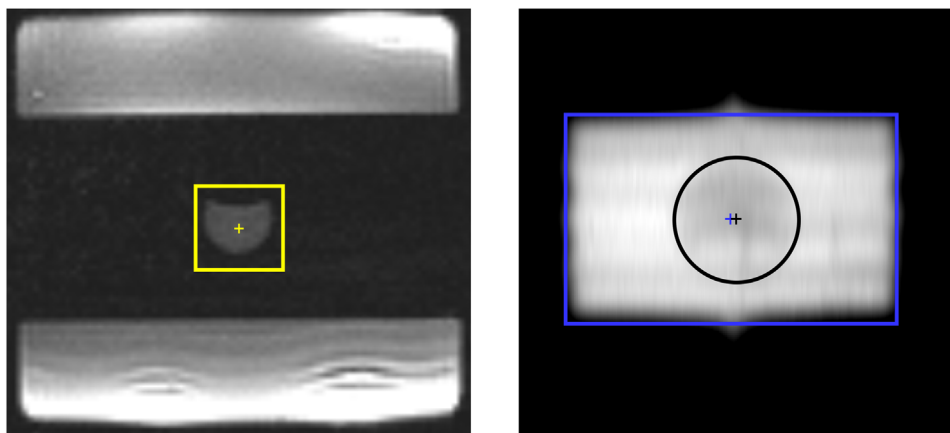
Note: As described in the main text, both the inter-quartile-range (IQR) of displacement and the period can differ from Mueller et al.<sup>27</sup> due to re-scaling and usage of a subset of each trace.

steady state free precession (bSSFP) sequence with a repetition time (TR) of 3.86 ms, an echo time (TE) of 1.92 ms and a flip angle of 28°. The slice thickness was 7 mm, the in-plane resolution 128 × 128 pixels and the field of view 300 mm, resulting in a voxel size of 2.34 × 2.34 × 7 mm<sup>3</sup>. Using these sequence parameters, an imaging frequency of 4 Hz was obtained. As

in Liu et al.,<sup>9</sup> a modified MRI reconstruction pipeline was used to stream raw image frames from the reconstruction computer to the target localization computer, where the images were analyzed in real-time. Using in-house software, target centroid positions in SI direction were obtained for each cine MRI frame using a cross-correlation based template matching algorithm (Figure 2, left panel). The template target for the matching process was defined on an MRI acquired prior to the motion experiments. Using the User Datagram Protocol (UDP), the target centroid positions were sent to the motion prediction computer. To avoid data loss inherent to the UDP, redundancy was introduced by sending the positions at 100 Hz.

## 2.4 | Motion prediction

The SI target centroid positions were pre-processed and used by a prediction model to obtain the future target centroid position in real-time, as described in the next sections. These tasks were implemented as different co-routines, such that they could run in an overlapping manner without blocking the main execution.



**FIGURE 2** (Left) Cine MRI frame acquired during irradiation with real-time target localization (yellow) using template matching. (Right) EPID frame acquired during irradiation with post-irradiation template matching analysis to obtain the target centroid positions (black) and the MLC aperture centroid positions (blue) needed for the evaluation. EPID, electronic portal imaging device; MLC, multileaf collimator; MRI, Magnetic resonance imaging.

### 2.4.1 | Data pre-processing

First, the data redundancy introduced by the UDP sending process was taken out by checking if the received target position differed from the previous one, thus recovering the original imaging frequency of 4 Hz. Then, target positions were accumulated until an input sequence of 8 s was formed (first buffering time). The input sequence had fixed length, that is, every time a new target position was available, the oldest target position in the input sequence was dropped and the new position was added. The input sequence was then smoothed using a moving average filter, to decrease the impact of noise arising from imaging/target localization on the prediction models. The moving average filter acted on a sliding window of three data points. To avoid boundary effects for the last two points of each sequence (the most recent and relevant for the prediction), we left the last point unchanged and set the penultimate point to the average between the unchanged penultimate point and the point one would obtain with the moving average filter. Finally, each input sequence was normalized in the range from  $-1$  to  $+1$ . The scaling factor for the normalization was temporarily saved for each sequence to later re-normalize the predictions.

### 2.4.2 | Prediction models

Three respiratory motion prediction models previously compared in-silico<sup>20</sup> and a baseline no-predictor were implemented in this study and applied to the eight unseen motion traces:

1. *Offline* LSTM: this model had been previously trained and validated using motion traces extracted from

4 Hz cine MRIs of 70 patients treated with a 0.35 T MRI-linac (13.1 h of data).<sup>20</sup> It was applied without any changes to hyper-parameters or weights to the unseen experiment traces and predicted the future target centroid position in 250 and 500 ms. Linear interpolation between these two points was used to obtain a prediction matching the end-to-end latency measured for the system. The interpolated prediction was then used for MLC-tracking.

2. *Offline+online* LSTM: this model was based on the *offline* LSTM described above. However, in this case we loaded the weights obtained from the optimization with the cohort of 70 patients and additionally re-optimized based on recent motion during the experiments. This worked by accumulating 20 s of target positions (second buffering time), which were subdivided into sets of input and output sequences. These pairs of input and output were used to iteratively train the LSTM using the mean-square-error loss between the output and predicted sequences. Every time a new target position was available (i.e., every 250 ms), the set of input/output sequences was updated and a new training was started for 250 ms, which allowed the completion of about 10 epochs. A more detailed explanation of the online optimization can be found in Lombardo et al.<sup>20</sup> As for the *offline* LSTM, the interpolated prediction was used for MLC-tracking.
3. *Online* LR: similarly to the *offline+online* LSTM, this model was continuously updated during the experiments based on recent motion. In contrast to the LSTM, the LR does not require iterative optimization, as an analytical solution exists.<sup>30</sup> For this reason, the *online* LR was solely based on the last 20 s of data and was solved from scratch on the updated set of sequences every 250 ms, that is, every time a new

target position was available, as in the previous in-silico study.<sup>20</sup> Also for the LR, linear interpolation between the 250 and 500 ms predictions provided the target position which was used for MLC-tracking.

4. No-predictor: to compare the three motion prediction models with a baseline without any prediction, we utilized the last available target centroid position for the subsequent MLC-tracking.

The LSTM models were run (and optimized) on an A5000 GPU with 24 GB of VRAM while the LR was run and solved using an Intel Xeon W-1250 CPU with 6 cores and 64 GB of RAM. A table showing the hyper-parameters taken over from the previous in-silico study can be found in the supplementary Table S1.

## 2.5 | MLC-tracking

In this study, a rigid single-target MLC-tracking software based on previous work was used.<sup>31</sup> Prior to irradiation, a rectangular MLC aperture was loaded and aligned with the target center. During irradiation, the predicted target position was used to calculate a displacement vector with respect to the target's original position. This displacement vector was then used to calculate an ideal aperture update which in turn was used by a leaf fitting algorithm to calculate the closest matching deliverable MLC aperture, taking into account physical limitations of the MLC such as finite leaf speed. The updated leaf positions were sent at 20 Hz to the MLC controller which shifted the aperture to compensate for the observed motion in real-time, as in previous MLC-tracking studies.<sup>8,32</sup>

## 2.6 | Accuracy evaluation

To evaluate the geometric accuracy of MLC-tracking with different motion predictors, EPID images were acquired at 3.5 Hz during irradiation and then analyzed after the experiments following Liu et al.<sup>9</sup> Using in-house software, we first applied a low-pass filter to the EPID frames to reduce the noise introduced by the magnetic field of the MRI-linac and then leveraged template matching to extract the target centroid positions and the MLC aperture centroid positions for each EPID frame automatically (see Figure 2, right panel). The template for the aperture and target were defined on a selected EPID frame once. All positions were scaled taking demagnification from the EPID plane to the isocenter into account. The root-mean-square error (RMSE) between the target and the MLC aperture positions (representing the motion predictions) was used to assess the performance of the different motion prediction models. As mentioned in Section 2.2, the RMSE was computed for

all prediction models on the same 1 min of each trace to enable a fair comparison.

To find out whether there was a significant difference between the RMSEs obtained by the four models for the different motion traces, a non-parametric Friedman test was used.<sup>33</sup> If the Friedman test was significant ( $p$ -value  $< 0.05$ ), a post-hoc Nemenyi test<sup>34</sup> was used to infer which model performed significantly better than another in a pair-wise fashion.

## 3 | RESULTS

### 3.1 | Latency measurements

When repeating the MLC-tracking experiments using no-predictor with the sinusoidal trace three times, an average end-to-end latency of  $(389 \pm 15)$  ms was obtained. We then computationally shifted the acquired aperture centroid curve by the calculated latency and obtained a baseline RMSE between aperture and target of  $(1.1 \pm 0.1)$  mm.

### 3.2 | Patient traces

Table 2 shows the RMSEs obtained for all prediction models and traces for each of the two measurements. MLC-tracking using the *offline+online* LSTM as motion predictor resulted in the best accuracy for all investigated motion traces. When calculating the mean and standard deviation of the RMSE over all motion traces, the *offline* LSTM led to  $(3.3 \pm 1.0)$  mm, the *offline+online* LSTM to  $(2.8 \pm 0.7)$  mm, the *online* LR to  $(3.3 \pm 0.7)$  mm and the no-predictor to  $(4.5 \pm 1.4)$  mm. The mean RMSE over all traces and for each trace (two measurements combined) for the different models are displayed as a bar plot in Figure 3.

Comparing the RMSE obtained from a measurement with its repetition revealed that the *offline+online* LSTM was also the most reproducible model with a deviation of up to 10%, compared to the *offline* LSTM with up to 14% and the *online* LR with up to 18%. Repeating the same trace using the no-predictor led to a maximum deviation of up to 6%, which can be considered the baseline. Differences between all models were significant according to the Friedman test ( $p$ -value =  $2e-9$ ). The post-hoc Nemenyi test showed that there was a significant difference between all models in a pair-wise comparison but for the *offline* LSTM and the *online* LR pair, as shown in Table 3.

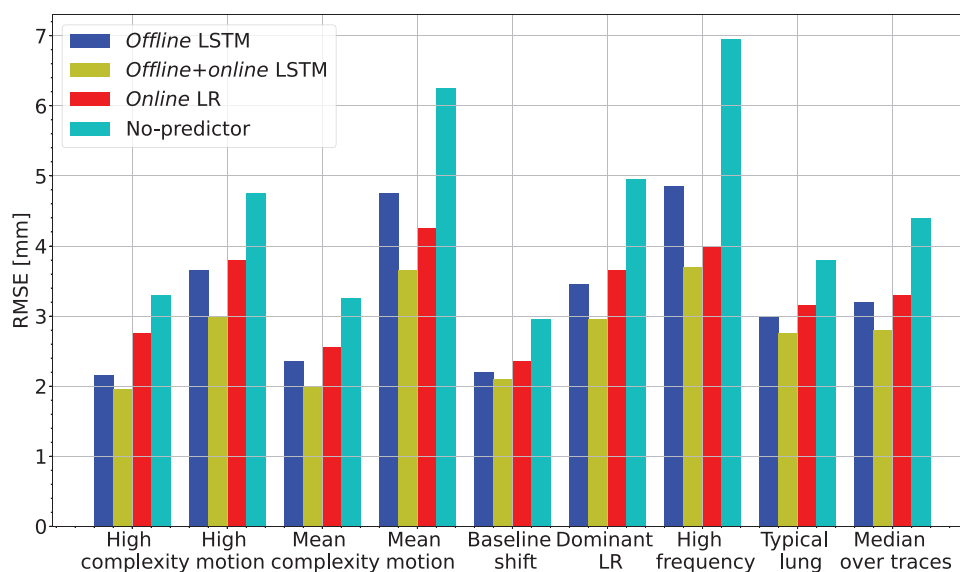
Figure 4 shows the centroid positions of the MLC aperture and the target obtained with the analysis of the EPID frames for four different models and a selected motion trace. Qualitatively, it can be noticed that the *offline* LSTM was more robust to the irregularity present at the end of the shown trace while the

**TABLE 2** MLC-tracking RMSE obtained when using different motion prediction models.

Motion trace	Offline LSTM	Offline+online LSTM	Online LR	No-predictor
	RMSE [mm]			
High complexity	2.2; 2.1	<b>1.9; 2.0</b>	2.6; 2.9	3.2; 3.4
High motion	3.5; 3.8	<b>3.0; 3.0</b>	3.6; 4.2	4.8; 4.7
Mean complexity	2.3; 2.4	<b>2.0; 2.0</b>	2.8; 2.3	3.2; 3.3
Mean motion	5.1; 4.4	<b>3.6; 3.7</b>	4.4; 4.1	6.3; 6.2
Baseline shift	2.2; 2.2	<b>2.1; 2.1</b>	2.3; 2.4	2.9; 3.0
Dominant RL	3.4; 3.5	<b>2.8; 3.1</b>	4.0; 3.3	4.9; 5.0
High frequency	4.8; 4.9	<b>3.8; 3.6</b>	4.2; 3.8	6.9; 7.0
Typical lung	3.0; 3.0	<b>2.8; 2.7</b>	3.2; 3.1	3.8; 4.0

Note: Repeated measurements are separated by a semicolon. The best RMSE on average for each trace is shown in bold.

Abbreviations: LR, linear regression; LSTM, long-short-term memory; MLC, multileaf collimator; RMSE, root-mean-square error.



**FIGURE 3** Plot with the mean (bar) and standard deviation (error bar) of the RMSE obtained from the two measurements for the four models and eight motion traces used in this study. Additionally, the mean and standard deviation of the RMSE over all traces for each model is shown. RMSE, root-mean-square error.

**TABLE 3**  $p$ -values obtained from the post-hoc Nemenyi test for all possible pairwise model comparisons.

Model 1	Model 2	$p$ -value
offline LSTM	offline+online LSTM	0.01*
offline LSTM	online LR	0.9
offline LSTM	no-predictor	0.002*
offline+online LSTM	online LR	0.002*
offline+online LSTM	no-predictor	0.001*
online LR	no-predictor	0.01*

Note: Significant  $p$ -values ( $< 0.05$ ) are denoted with an asterisk.

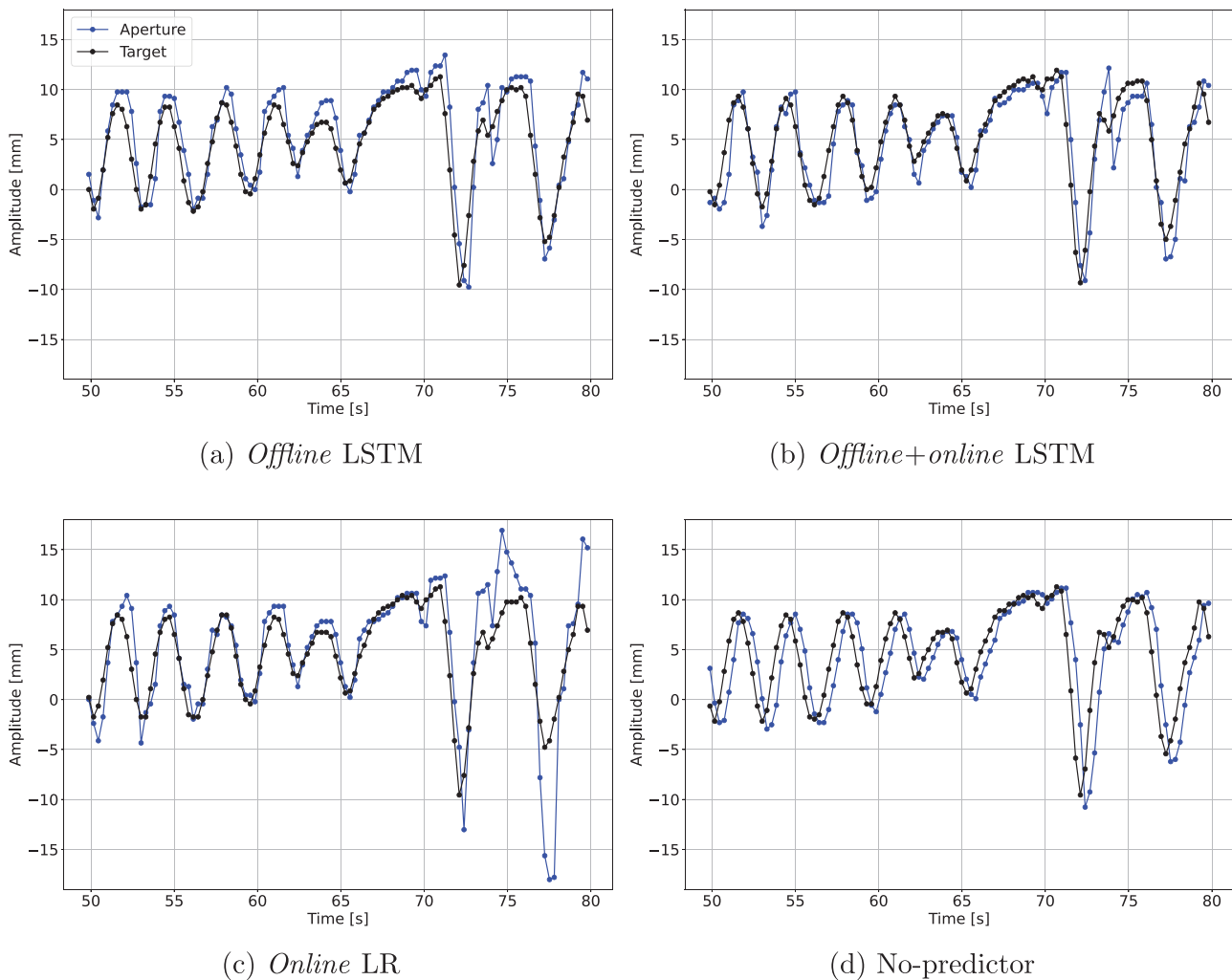
Abbreviations: LR, linear regression; LSTM, long-short-term memory.

online LR overshoot less during regular breathing. The offline+online LSTM seemed to combine the advantages of the two models while for the no-predictor

the system latency was clearly visible. For the same trace, we show in the online supplementary materials an EPID video displaying the target and the MLC aperture driven by the no-predictor ('pass\_through\_high\_complexity\_1.avi') or the offline+online LSTM ('online\_lstm\_smooth\_high\_complexity\_1.avi').

## 4 | DISCUSSION

The experiments performed in this study showed that accurate MLC-tracking using motion prediction is possible for a variety of different breathing patterns. We were able to successfully compensate for a measured end-to-end system latency of  $(389 \pm 15)$  ms, which is slightly increased compared to the latency of  $(328 \pm 44)$  ms



**FIGURE 4** MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the “high complexity” motion trace for one of the two measurements. The MLC is moved according to one of the four motion prediction models. The same four models but for the other seven traces are shown in the online supplementary ([mp16770-sup-0001-SuppMat.pdf](https://onlinelibrary.wiley.com/doi/10.1002/mp.16770)). EPID, electronic portal imaging device; MLC, multileaf collimator.

measured by Liu et al. on the same system,<sup>9</sup> but still within what is expected for MLC-tracking on MRI-linacs in other studies.<sup>11,22</sup> The latency corrected baseline RMSE of  $(1.1 \pm 0.1)$  mm is comparable with the RMSE of  $(1.2 \pm 0.1)$  found by Liu et al.<sup>9</sup> and includes experimental limitations such as imaging resolution, MLC performance and EPID accuracy/analysis.

The *offline+online LSTM* model was found to be the best motion predictor for all eight investigated motion traces, being significantly better than both its offline version, that is without continuous re-optimization, and an *online LR*. The *offline LSTM* and *online LR* performed similarly, with one model being better than the other depending on the motion trace. The no-predictor was significantly worse than all other models, confirming the value of motion prediction during MRIgRT with MLC-tracking. In general, absolute performance differences were also dependent on the motion trace.

For the “typical lung” trace, which is the one with the smallest IQR of motion (Table 1), we found a maximum RMSE difference between any two experiments with the *offline+online LSTM* and the *offline LSTM/online LR/no-predictor* of 0.3 mm/0.5 mm/1.3 mm (Table 2). On the other hand, for the “high frequency” trace, which is very regular but presents high amplitude and frequency, we found an RMSE difference between the *offline+online LSTM* and the *offline LSTM/online LR/no-predictor* of 1.3 mm/0.6 mm/3.4 mm. Based on this trace, we hypothesize that the *online LR*, which is solely based on the last 20 s of motion, performs particularly well if the motion is regular, compared to the *offline LSTM* which has been trained on data which presented on average a different frequency. The *offline+online LSTM* might have outperformed all other models because it combined training on a large set of different breathing patterns with being able to adapt to for example a patient-specific breathing



frequency. We also investigated reproducibility of the models by repeating each experiment twice and found the *online* LR to be the least reproducible with a deviation of 18%. We hypothesize that this originates from the fact that this is the only model which solely relies on current data (no prior training on large datasets) and is therefore the most sensible to variations in the input centroids due to for example acquisition noise or imperfections of the template matching.

While this work investigated MLC-tracking with motion prediction using MRI-guidance, the proposed methods could also be used with x-ray guidance. The reduced soft-tissue contrast might require more advanced target localization algorithms than template matching in a markerless setting.<sup>35</sup> As the motion prediction models are based on centroid positions, only small modifications would be needed such as training on other temporal resolutions to reflect the different imaging frequency and adjustment of the prediction horizon to smaller latencies.

The current study presents a few limitations. The motion phantom allows for 1D rigid shifts only, neglecting the fact that motion occurs in all three directions and that deformation or rotations can be observed.<sup>36,37</sup> Assuming the cine MRI is acquired in a single 2D slice as in this study, MLC-tracking to compensate for in-plane displacement/deformation/rotation could be implemented in future studies by leveraging a deformable target localization algorithm,<sup>38</sup> a 2D contour prediction algorithm<sup>21</sup> and a deformable MLC-tracking algorithm.<sup>39</sup> Beam's eye view 2D cine MRI with tumor-volume projection might be used to ensure better beam conformality.<sup>40,41</sup> To fully compensate motion in all three directions, time-resolved volumetric MRIs would be needed, which are currently being investigated by several groups<sup>42–44</sup> and would lead to an increment in latency, which in turn increases the relevance of motion prediction. Another limitation consists in the fact that we had to re-scale the motion traces to 30 mm peak-to-peak amplitude due to limitations in the experimental setting, the original mean peak-to-peak amplitude over all traces being 16.4 mm (range 9.4–23.8 mm). This means that the obtained RMSEs represent in absolute terms an overestimation of the error which would have been obtained with the original traces while all relative comparisons between the models hold true. Finally, the fact that the target in the phantom is visible with high contrast on the cine MRI facilitated its localization. However, imaging a real tumor would have affected the target localization and therefore all models in the same manner, so the results of our comparison should hold true.

## 5 | CONCLUSIONS

In this study, we experimentally compared conventional and machine learning motion prediction models

for MLC-tracking in SI direction based on 4 Hz cine MRI. We showed for eight patient-derived respiratory motion traces with different complexity that all models significantly improved the MLC-tracking performance compared to a baseline no-predictor. A continuously re-optimized LSTM model was found to perform the best for all motion traces, confirming the in-silico result that this model is an ideal candidate to mitigate the latency and therefore improve the accuracy of MLC-tracking during MRIgRT.

## ACKNOWLEDGEMENTS

The authors thank Magdalena Bazalova-Carter for her help during the measurements and staff at Image X Institute and Ingham Institute for their support. EL was funded by the German Research Foundation (DFG) within the Research Training Group GRK 2274. PZYL and DEJW acknowledge funding from Cancer Institute NSW fellowships.

Open access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

The Department of Radiation Oncology of the University Hospital of LMU Munich has research agreements with Brainlab, Elekta and ViewRay. PJK is an inventor on US patents 7,469,035 and 8,971,489 that are related to MLC-tracking. Patent 7,469,035 is unlicensed; patent 8,971,489 is exclusively licensed to Asto CT.

## REFERENCES

1. Kurz C, Buizza G, Landry G, et al. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol.* 2020;15:93.
2. Keall PJ, Brighi C, Glide-Hurst C, et al. Integrated MRI-guided radiotherapy - opportunities and challenges. *Nat Rev Clin Oncol.* 2022;19:458–470.
3. Corradini S, Alongi F, Andratschke N, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. *Radiat Oncol.* 2019;14:1–12.
4. Jassar H, Tai A, Chen X, et al. Real-time motion monitoring using orthogonal cine MRI during MR-guided adaptive radiation therapy for abdominal tumors on 1.5 T MR-Linac. *Med Phys.* 2023;50(5):3103–3116.
5. Green O, Rankine LJ, Cai B, et al. First clinical implementation of real-time, real anatomy tracking and radiation beam control. *Med Phys.* 2018;45:3728–3740.
6. Eze C, Lombardo E, Nierer L, et al. MR-guided radiotherapy in node-positive non-small cell lung cancer and severely limited pulmonary reserve: a report proposing a new clinical pathway for the management of high-risk patients. *Radiat Oncol.* 2022;17:1–8.
7. Finazzi T, de Koste JRvS, Palaciosn MA, et al. Delivery of magnetic resonance-guided single-fraction stereotactic lung radiotherapy. *Phys Imaging Radiat Oncol.* 2020;14:17–23.
8. Smith RL, Sawant A, Santanam L, et al. Integration of real-time internal electromagnetic position monitoring coupled with dynamic multileaf collimator tracking: an intensity-modulated radiation therapy feasibility study. *Int J Radiat Oncol Biol Phys.* 2009;74:868–875.
9. Liu PZ, Dong B, Nguyen DT, et al. First experimental investigation of simultaneously tracking two independently moving targets on

- an MRI-linac using real-time MRI and MLC tracking. *Med Phys*. 2020;47:6440-6449.
10. 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:e44-e64.
  11. Glitzner M, Woodhead PL, Borman PTS, Lagendijk JJW, Raaymakers BW. Technical note: MLC-tracking performance on the Elekta Unity MRI-linac. *Phys Med Biol*. 2019;64:15NT02.
  12. Yun J, Mackenzie M, Rathee S, Robinson D, Fallone B. An artificial neural network(ANN)-based lung-tumor motion predictor for intrafractional MR tumor tracking. *Med Phys*. 2012;39:4423-4433.
  13. Bourque AE, Carrier JF, Filion É, Bedwani S. A particle filter motion prediction algorithm based on an autoregressive model for real-time MRI-guided radiotherapy of lung cancer. *Biomed Phys Eng Express*. 2017;3:035001.
  14. Ginn JS, Ruan D, Low DA, Lamb JM. An image regression motion prediction technique for MRI-guided radiotherapy evaluated in single-plane cine imaging. *Med Phys*. 2020;47:404-413.
  15. Sharp G, Jiang SB, Shimizu S, Shirato H. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys Med Biol*. 2004;49:425.
  16. Sun W, Jiang M, Ren L, Dang J, You T, Yin F. Respiratory signal prediction based on adaptive boosting and multi-layer perceptron neural network. *Phys Med Biol*. 2017;62:6822.
  17. A., Jöhl, Ehrbar S, Guckenberger M, et al. Performance comparison of prediction filters for respiratory motion tracking in radiotherapy. *Med Phys*. 2020;47:643-650.
  18. Wang R, Liang X, Zhu X, Xie Y. A feasibility of respiration prediction based on deep bi-LSTM for real-time tumor tracking. *IEEE Access*. 2018;6:51262-51268.
  19. Lin H, Shi C, Wang B, Chan MF, Tang X, Ji W. Towards real-time respiratory motion prediction based on long short-term memory neural networks. *Phys Med Biol*. 2019;64:085010.
  20. Lombardo E, Rabe M, Xiong Y, et al. Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. *Phys Med Biol*. 2022;67:095006.
  21. Lombardo E, Rabe M, Xiong Y, et al. Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy. *Radiother Oncol*. 2023;182:109555.
  22. Yun J, Wachowicz K, Mackenzie M, Rathee S, Robinson D, Fallone B. First demonstration of intrafractional tumor-tracked irradiation using 2D phantom MR images on a prototype linac-MR. *Med Phys*. 2013;40:051718.
  23. Uijtewaal P, Borman PT, Woodhead PL, Hackett SL, Raaymakers BW, Fast MF. Dosimetric evaluation of MRI-guided multi-leaf collimator tracking and trailing for lung stereotactic body radiation therapy. *Med Phys*. 2021;48:1520-1532.
  24. Uijtewaal P, Borman PT, Woodhead PL, et al. First experimental demonstration of VMAT combined with MLC tracking for single and multi fraction lung SBRT on an MR-linac. *Radiother Oncol*. 2022;174:149-157.
  25. Keijneemans K, Borman PT, Uijtewaal P, Woodhead PL, Raaymakers BW, Fast MF. A hybrid 2D/4D-MRI methodology using simultaneous multislice imaging for radiotherapy guidance. *Med Phys*. 2022;49:6068-6081.
  26. Liney GP, Dong B, Weber E, et al. Imaging performance of a dedicated radiation transparent RF coil on a 1.0 Tesla inline MRI-linac. *Phys Med Biol*. 2018;63:135005.
  27. M., Mueller et al. The markerless lung target tracking AAPM grand challenge (MATCH) results. *Med Phys*. 2022;49:1161-1180.
  28. Schmitt D, Nill S, Roeder F, Gompelmann D, Herth F, Oelfke U. Motion monitoring during a course of lung radiotherapy with anchored electromagnetic transponders: quantification of inter- and intrafraction motion and variability of relative transponder positions. *Strahlenther Onkol*. 2017;193:840.
  29. Suh Y, Dieterich S, Cho B, Keall PJ. An analysis of thoracic and abdominal tumour motion for stereotactic body radiotherapy patients. *Phys Med Biol*. 2008;53:3623.
  30. Krauss A, Nill S, Oelfke U. The comparative performance of four respiratory motion predictors for real-time tumour tracking. *Phys Med Biol*. 2011;56:5303-5317.
  31. Sawant A, Venkat R, Srivastava V, et al. Management of three-dimensional intrafraction motion through real-time DMLC tracking. *Med Phys*. 2008;35:2050-2061.
  32. Booth JT, Caillet V, Hardcastle N, et al. The first patient treatment of electromagnetic-guided real time adaptive radiotherapy using MLC tracking for lung SABR. *Radiother Oncol*. 2016;121:19-25.
  33. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc*. 1937;32:675-701.
  34. Nemenyi PB. *Distribution-free Multiple Comparisons*. PhD thesis. Princeton University; 1963.
  35. Sakata Y, Hirai R, Kobuna K, Tanizawa A, Mori S. A machine learning-based real-time tumor tracking system for fluoroscopic gating of lung radiotherapy. *Phys Med Biol*. 2020;65:085014.
  36. Pantarotto JR, Piet AH, Vincent A, de Koste JRvS, Senan S. Motion analysis of 100 mediastinal lymph nodes: potential pitfalls in treatment planning and adaptive strategies. *Int J Radiat Oncol Biol Phys*. 2009;74:1092-1099.
  37. Feng M, Balter JM, Normolle D, Adusumilli S, Cao Y, Chenevert TL, Ben-Josef E. Characterization of pancreatic tumor motion using cine MRI: surrogates for tumor position should be used with caution. *Int J Radiat Oncol Biol Phys*. 2009;74:884-891.
  38. Klueter S. Technical design and concept of a 0.35 TMR-linac. *Clin Transl Radiat Oncol*. 2019;18:98-101.
  39. Ge Y, O'Brien RT, Shieh CC, Booth JT, Keall PJ. Toward the development of intrafraction tumor deformation tracking using a dynamic multi-leaf collimator. *Med Phys*. 2014;41:061703.
  40. Wachowicz K, Murray B, Fallone BG. On the direct acquisition of beam's-eye-view images in MRI for integration with external beam radiotherapy. *Phys Med Biol*. 2018;63:125002.
  41. Nie X, Rimner A, Li G. Feasibility of MR-guided radiotherapy using beam-eye-view 2D-cine with tumor-volume projection. *Phys Med Biol*. 2021;66:045020.
  42. Stemkens B, Tijssen RH, de Senneville BD, Lagendijk JJ, van den Berg CA. Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy. *Phys Med Biol*. 2016;61:5335.
  43. Paganelli C, Lee D, Kipritidis J, et al. Feasibility study on 3D image reconstruction from 2D orthogonal cine-MRI for MRI-guided radiotherapy. *J Med Imaging Radiat Oncol*. 2018;62:389-400.
  44. Rabe M, Paganelli C, Riboldi M, et al. Porcine lung phantom-based validation of estimated 4D-MRI using orthogonal cine imaging for low-field MR-Linacs. *Phys Med Biol*. 2021;66:055006.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Lombardo E, Liu PZY, Waddington DEJ, et al. Experimental comparison of linear regression and LSTM motion prediction models for MLC-tracking on an MRI-linac. *Med Phys*. 2023;50:7083–7092.  
<https://doi.org/10.1002/mp.16770>