
Deep Learning for Real-time Motion Management during MRI-guided Radiotherapy

Elia Lombardo



München 2024

Aus der
Klinik und Poliklinik für Strahlentherapie und Radioonkologie
der
Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. med. Claus Belka

Deep Learning for Real-time Motion Management during MRI-guided Radiotherapy

Elia Lombardo

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität
München

vorgelegt von
Elia Lombardo
aus Palermo

Jahr
2024

Gedruckt mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Erstgutachter: Prof. Dr. rer. nat. Guillaume Landry
Zweitgutachter: Prof. Dr. rer. nat. Michael Ingrisch
Dekan: Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung: 28. Juli 2025

Table of contents

Table of contents	v
Affidavit	vii
Confirmation of congruency	ix
List of abbreviations	xi
List of publications	xiii
Candidate's contribution to the publications	xvii
Zusammenfassung	xix
Abstract	xxi
1 Introduction	1
1.1 Modern radiotherapy	2
1.2 MRI-guided radiotherapy	2
1.2.1 Online adaptation	6
1.2.2 Real-time adaptation	7
1.3 Deep learning	11
1.3.1 Long short-term memory networks	11
1.3.2 Transformer networks	13
1.4 State-of-the-art literature	15
1.5 Evaluation metrics	17
1.5.1 Root Mean Squared Error	17
1.5.2 Maximum Error	17
1.5.3 Dice Similarity Coefficient	18
1.5.4 Hausdorff Distance	18
2 Articles	19
2.1 Paper I: Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy	19
2.2 Paper II: Evaluation of real-time tumor contour prediction using LSTM net- works for MR-guided radiotherapy	36
2.3 Paper III: Experimental comparison of linear regression and LSTM motion prediction models for MLC-tracking on an MRI-linac	55

2.4	Paper IV: Patient-specific deep learning tracking framework for real-time 2D target localization in magnetic resonance imaging-guided radiation therapy . .	76
3	Outlook	103
3.1	Conclusions	104
A	Additional publication	105
A.1	Topical review: Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects	105
	Bibliography	117
	Acknowledgments	123



Affidavit

Lombardo, Elia

Surname, first name

I hereby declare, that the submitted thesis entitled

Deep Learning for Real-time Motion Management during MRI-guided Radiotherapy

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, 14/09/2025

Place, Date

Elia Lombardo

Signature doctoral candidate



Confirmation of congruency between printed and electronic version of the doctoral thesis

Doctoral candidate: Elia Lombardo

I hereby declare that the electronic version of the submitted thesis, entitled

Deep Learning for Real-time Motion Management during MRI-guided Radiotherapy

is congruent with the printed version both in content and format.

Munich, 14/09/2025

Place, Date

Elia Lombardo

Signature doctoral candidate

List of abbreviations

AI artificial intelligence

ANN artificial neural network

CBCT cone beam computed tomography

CNN convolutional neural network

CT computed tomography

DIR deformable image registration

DL deep learning

DSC dice similarity coefficient

EPID electronic portal imaging device

GTV gross tumor volume

HD Hausdorff distance

linac linear accelerator

LSTM long short-term memory

ME maximum error

ML machine learning

MLC multi-leaf collimator

MRI magnetic resonance imaging

MRI-linac magnetic resonance imaging-linear accelerator

MRIGRT magnetic resonance imaging-guided radiotherapy

OAR organ at risk

RMSE root mean squared error

RNN recurrent neural network

US ultra-sound imaging

List of publications

Peer-reviewed articles related to the thesis

Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. (2022) Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. *Physics in Medicine & Biology*, 67(9), 095006.

Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. (2023) Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy. *Radiotherapy and Oncology*, 182, 109555.

Lombardo, E., Liu, P. Z., Waddington, D. E., Grover, J., Whelan, B., Wong, E., ... & Keall, P. J. (2023) Experimental comparison of linear regression and LSTM motion prediction models for MLC-tracking on an MRI-linac. *Medical Physics*, 50(11), 7083-7092.

Lombardo, E., Velezmoro, L., Marschner, S.N., Rabe, M., Tejero, C., Papadopoulou, C.I., ... & Landry, G. (2025) Patient-specific deep learning tracking framework for real-time 2D target localization in magnetic resonance imaging-guided radiation therapy. *International Journal of Radiation Oncology*Biophysics*, 122(4), 827-837.

Peer-reviewed topical review related to the thesis

Lombardo, E., Dhont, J., Page, D., Garibaldi, C., Kunzel, L. A., Hurkmans, C., ... & Placidi, L. (2023) Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects. *Radiotherapy and Oncology*, 109970.

Other articles published during the doctorate

Marschner S., **Lombardo, E.**, ..., Walter, F. (2021) Risk Stratification Using 18F-FDG PET/CT and Artificial Neural Networks in Head and Neck Cancer Patients Undergoing Radiotherapy. *Diagnostics*, 11,9.

Eze, C., **Lombardo, E.**, Nierer, L., Xiong, Y., Niyazi, M., Belka, C., Manapov, F., Corradini, S. (2022) 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. *Radiation Oncology*, 17,1.

Zschaek, S., Weingärtner, J., **Lombardo, E.**, ..., Hofheinz, F. (2022) 18F-Fluorodeoxyglucose Positron Emission Tomography of Head and Neck Cancer: Location and HPV Specific Parameters for Potential Treatment Individualization. *Frontiers in Oncology*, 12.

Wang, Y., **Lombardo, E.**, ..., Landry, G. (2022) Deep learning based time-to-event analysis with PET, CT and joint PET/CT for head and neck cancer prognosis. *Computer Methods and Programs in Biomedicine*, 222, 106948.

Lombardo, E., Hess, J., ..., Landry, G., Unger, K. (2022) DeepClassPathway: Molecular pathway aware classification using explainable deep learning. *European Journal of Cancer*, 176, 41-49.

Nikulin, P., Zschaek, S., Maus, J., Cegla, P., **Lombardo, E.**, Furth, C., ..., van den Hoff, J. (2023) A convolutional neural network with self-attention for fully automated metabolic tumor volume delineation of head and neck cancer in 18-FDG PET/CT. *European Journal of Nuclear Medicine and Molecular Imaging*, 1-16.

Schmitz, H., Thummerer, A., Kawula, M., **Lombardo, E.**, ... Landry, G. (2023) ScatterNet for projection-based 4D cone-beam computed tomography intensity correction of lung cancer patients. *Physics and Imaging in Radiation Oncology*, 27, 100482.

Ballhausen, H., Li, M., **Lombardo, E.**, Landry, G., Belka, C. (2023) Planning CT Identifies Patients at Risk of High Prostate Intrafraction Motion. *Cancers*, 15(16), 4103.

Wang, Y., **Lombardo, E.**, Huang, L., Avanzo, M., Fanetti, G., Franchin, G., ..., Landry, G. (2024) Comparison of deep learning networks for fully automated head and neck tumor delineation on multi-centric PET/CT images. *Radiation Oncology*, 19(1), 3.

Vagni, M., Tran, H. E., Catucci, F., Chiloire, G., D'Aviero, A., Re, A., ..., **Lombardo, E.**, ..., Placidi, L. (2024) Impact of bias field correction on 0.35 T pelvic MR images: evaluation on generative adversarial network-based OARs' auto-segmentation and visual grading assessment. *Frontiers in oncology*, 14, 1294252.

Conference contributions

Lombardo, E., Hess, J., ..., Landry, G. & Unger, K. Linear regression and deep learning models for transcriptome-based HPV-status prediction in head and neck squamous cell carcinoma. Oral presentation at *Head and Neck Symposium* (2022), Essen, Germany.

Lombardo, E., Xiong, Y., Rabe, M., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. LSTM networks for real-time respiratory motion prediction for a 0.35 T MR-linac. Oral presentation at *ESTRO Annual Congress* (2022), Copenhagen, Denmark.

Wang, Y., **Lombardo, E.**, ..., Landry, G. Deep learning based time to event analysis with PET, CT and joint PET/CT for H&N cancer prognosis. Oral presentation at *ESTRO Annual Congress* (2022), Copenhagen, Denmark.

Unger, K., **Lombardo, E.**, ..., Landry, G. Regression and deep learning for transcriptome-based HPV-status prediction in head and neck cancers. Digital poster at *ESTRO Annual Congress* (2022), Copenhagen, Denmark.

Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. Offline und online LSTM Netzwerke für die Echtzeitvorhersage von Atembewegungen an einem 0.35 T MR-Linac. Oral presentation at *Jahrestagung der DGMP* (2022), Aachen, Germany.

Lombardo, E. AI tracking in MRI guided radiotherapy. Invited talk at *ESTRO Physics Workshop* (2022), Lisbon, Portugal.

Lombardo, E., Xiong, Y., Rabe, M., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. Offline and online LSTM networks for respiratory motion prediction at MR-linacs. Oral presentation at *Australian MR in RT* (2022), Noosa, Australia.

Vagni, M., Tran, H.E., Romano, A., Boldrini, L., Chiloiro, G., Landry, G., Kurz, C., Corradini, S., Kawula, M., **Lombardo, E.**, ..., Cusumano, D. A comparison between 2D and 3D GAN as a supporting tool for rectum segmentation on 0.35 T MR images. Oral presentation at *ESTRO Annual Congress* (2023), Vienna, Austria.

Schmitz, H., **Lombardo, E.**, ..., Landry, G. ScatterNet for 4D cone-beam CT intensity correction of lung cancer patients. Oral presentation at *ESTRO Annual Congress* (2023), Vienna, Austria.

Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. Comparing LSTM networks for real-time target segmentation prediction on low-field MR-linacs. Digital poster at *ESTRO Annual Congress* (2023), Vienna, Austria.

Lombardo, E., Liu, P. Z., Waddington, D. E., Grover, J., Wong, E., Reiner, M., ... & Keall, P.J. Experimental investigation of AI motion prediction for MLC-tracking at an MRI-linac. Oral presentation by supervisor Marco Riboldi at *AAPM Annual Meeting & Exhibition* (2023), Houston, USA.

Lombardo, E. KI in der Strahlenphysik - Aktueller Stand und Zukunftsperspektiven. Invited talk at *DEGRO Konferenz* (2023), Kassel, Germany.

Lombardo, E., Liu, P. Z., Waddington, D. E., Grover, J., Wong, E., Reiner, M., ... & Landry, G. Experimenteller Vergleich konventioneller und KI-basierter Algorithmen zur Bewegungsvorhersage für MLC-Tracking an einem MR-Linac. Oral presentation at *Jahrestagung der DGMP* (2023), Magdeburg, Germany.

Lombardo, E., Velezmoro, L., Marschner, S.N., Rabe, M., Tejero, C., Papadopoulou, C.I., ... & Landry, G. Transformers for real-time 2D target tracking in MRI-guided radiotherapy. Oral presentation at *10th MR in RT Symposium* (2024), Rome, Italy.

Lombardo, E., Velezmoro, L., Marschner, S.N., Reiner, M., Corradini, S., Belka, C., ... & Landry, G. Comparing Transformer training strategies for real-time tumor tracking in MRI-guided radiotherapy. Oral presentation at *ESTRO Annual Congress* (2024), Glasgow, UK.

Velezmoro, L., **Lombardo, E.**, Marschner, S.N., Reiner, M., Corradini, S., Belka, C., ... & Landry, G. Patient-specific deep learning for real-time tumor tracking in MR-guided radiotherapy. Oral presentation at *ESTRO Annual Congress* (2024), Glasgow, UK.

Lombardo, E. Unlocking the promise of AI in real-time for a moving target. Teaching lecture at *ESTRO Annual Congress* (2024), Glasgow, UK.

Wang, Y., **Lombardo, E.**, Wang, J., ... & Landry, G. AI assisted tumor tracking in 1.5 T MRgRT based on transfer learning from 0.35 T. Oral presentation at *ECMP* (2024), Munich, Germany.

Lombardo, E. Clinical Work and Academic Research as MPE in Radiotherapy. Meet the Experts Session at *ECMP* (2024), Munich, Germany.

Newsletters

Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. (2022). Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. ESTRO newsletter <https://www.estro.org/About/Newsroom/Newsletter/Physics/Offline-and-online-LSTM-networks-for-respiratory-m>.

Candidate's contribution to the publications

Paper I: Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy

The candidate curated the data, optimized and evaluated the centroid prediction LSTM network, benchmarked against a linear regression model, performed the literature review and drafted the manuscript.

Paper II: Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy

The candidate curated the data, optimized and evaluated the LSTM and the two convolutional LSTMs to work with 2D contours, performed the literature review and drafted the manuscript.

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

During a 2 months research stay in Sydney, Australia, the candidate changed the LSTM code developed in Munich to work with a real-time stream of data from a prototype MRI-linac located in Sydney. The candidate also contributed to the experimental sessions and evaluated the results. Back in Munich he performed the literature review and drafted the manuscript.

Paper IV: Patient-specific deep learning tracking framework for real-time 2D target localization in magnetic resonance imaging-guided radiation therapy

The candidate curated the data, optimized and evaluated the transformer network, benchmarked against a B-spline and the vendor's algorithm, supported the second author in developing the comparison with an auto-segmentation model, performed the literature review and drafted the manuscript.

Zusammenfassung

Ionisierende Strahlung wird seit Ende des 19. Jahrhunderts zur Behandlung von Krebs eingesetzt. Die moderne Strahlentherapie ermöglicht es, den Strahl mit Millimeterpräzision so zu formen, dass die höchstmögliche Dosis auf den Tumor gerichtet wird, während die Dosis auf das umliegende gesunde Gewebe so gering wie möglich bleibt. Tumore in bestimmten anatomischen Bereichen bewegen sich jedoch während der Behandlung. Im Bauchraum und im Brustkorb beispielsweise können sich Tumore, die durch die Atmung des Patienten beeinflusst werden, um Zentimeter verschieben, was die Genauigkeit der Strahlentherapie verringert. Hybride Magnetresonanztomographie (MRI)-Linearbeschleuniger (MRI-Linacs) ermöglichen es die sich bewegende Anatomie während der Bestrahlung sichtbar zu machen und somit den Strahl in Echtzeit (Hunderte von Millisekunden) an die Bewegung anzupassen. In dieser kumulativen Dissertation wurden verschiedene Methoden der künstlichen Intelligenz (KI) entwickelt, um Echtzeit-Bewegungsmanagement mit MRI-Linacs zu unterstützen. Jeder Schritt der Echtzeitanpassung benötigt eine bestimmte Zeit, was zu einer Verzögerung zwischen der physikalischen Bewegung des Tumors und der tatsächlichen Strahlanpassung führt, die als Systemlatenz bezeichnet wird. Die ersten drei Projekte zielten darauf ab, die Latenzzeit der MRI-Linacs durch den Einsatz von KI-basierter Zeitreihenvorhersage zu kompensieren. Im ersten Projekt wurde ein Long-Short-Term Memory-(LSTM-)Netzwerk zur Vorhersage von Atembewegungen auf der Grundlage von Cine-MRI-Daten, die während MRI-Linac-Behandlungen aufgenommen wurden, entwickelt. Das LSTM wurde sowohl anhand von Populationsdaten als auch auf patientenspezifische Weise optimiert und mit einem populationsbasierten und einem patientenspezifischen konventionellen linearen Regressionsmodell verglichen. Bei der Auswertung mit Daten aus zwei verschiedenen Einrichtungen in München und Rom zeigte sich, dass das patientenspezifische LSTM die beste Leistung bei der Vorhersage der zukünftigen Schwerpunktposition des Tumors in superior-inferior Richtung erzielte. Im zweiten Projekt dieser Arbeit wurden dieselben Cine-MRI-Daten verwendet, um die Modelle auf die Vorhersage zukünftiger 2D Tumorkonturen zu erweitern. Das LSTM aus dem vorherigen Projekt wurde mit einer rigiden Verschiebung der letzten verfügbaren Tumorkontur kombiniert und mit Faltungs-LSTMs verglichen, die entweder direkt die zukünftigen Konturen oder zukünftige Deformationsfelder vorhersagen, die zur Verformung der letzten verfügbaren Kontur verwendet werden können. Das patientenspezifische LSTM in Kombination mit der rigiden Verschiebung übertraf die faltungs-basierten LSTMs in beiden Auswertungsdatensätzen. Im dritten Projekt wurde das patientenspezifische LSTM aus dem ersten Projekt mit einem Forschungsprototyp MRI-Linac in Sydney integriert und mit einem Multi-Leaf-Collimator-(MLC-)Tracking Experiment validiert. In einer Phantomstudie wurde das Modell wiederum mit einem populationsbasierten LSTM, einer patientenspezifischen linearen Regression und einem Szenario ohne Bewegungsvorhersage verglichen. Die in-silico-Ergebnisse aus dem vorherigen Projekt bestätigend wurde gezeigt, dass die Genauigkeit, mit

der die Strahlung das Ziel im Phantom traf, am höchsten war, wenn das patientenspezifische LSTM verwendet wurde, und dass der Fehler im Vergleich zu dem Szenario, bei dem die Latenz nicht kompensiert wird, fast halbiert wurde. Im vierten Projekt dieser Arbeit wurde der Schwerpunkt von der Bewegungsvorhersage auf die Echtzeitlokalisierung des Tumors auf dem aktuellen Cine-MRI-Bild verschoben. Konkret wurde unter Verwendung von in München aufgenommenen MRI-Linac-Daten ein Transformer-Netzwerk für schnelle deformierbare Bildregistrierung (DIR) entwickelt und mit einem KI-Autosegmentierungsmodell und einem langsamen konventionellen B-Spline-DIR-Modell verglichen. Der Transformer wurde mit verschiedenen unüberwachten, überwachten und patientenspezifischen Strategien optimiert. Der patientenspezifische Transformer zeigte die beste Leistung und übertraf alle anderen Modelle und Optimierungsstrategien, ohne dass es zu signifikanten zusätzlichen Latenzzeiten kam. Zukünftige Studien könnten basierend auf den hier entwickelten Modellen fortschrittlichere KI-basierte Bewegungsmanagementmodelle in 3D entwickeln, um die Genauigkeit der echtzeitadaptiven MRI-geführten Strahlentherapie weiterhin zu verbessern, wie im Ausblick und im Review Paper im Anhang dieser Arbeit beschrieben.

Abstract

Ionizing radiation has been used to treat cancer since the end of the 19th century. Modern radiotherapy allows to shape the radiation beam with millimeter precision to deliver the highest possible dose to the tumor while keeping the dose to surrounding healthy tissue as low as possible. Tumors in certain anatomical locations, however, move during the treatment. In the abdomen and thorax for instance, tumors affected by the patient's respiration can move by centimeters, thus decreasing the accuracy of radiotherapy. Hybrid magnetic resonance imaging (MRI)-linear accelerators (MRI-linacs) allow to visualize the moving anatomy during irradiation and therefore to adapt the beam to the observed motion in real-time (hundreds of milliseconds). In this cumulative dissertation, different artificial intelligence (AI) methods were developed to support real-time motion management with MRI-linacs. Each step of the real-time adaptation takes a certain time to complete, which leads to a delay between the physical motion of the tumor and the actual beam adaptation called the system latency. The first three projects aimed at compensating for MRI-linac latency by using AI for time series prediction. In the first project, a long short-term memory (LSTM) network for respiratory motion prediction based on cine MRI acquired during MRI-linac treatments was developed. The LSTM was optimized both using population data and in a patient-specific fashion and compared with a population-based and a patient-specific conventional linear regression model. The patient-specific LSTM was found to perform best in predicting the future centroid position of the tumor in super-inferior direction when evaluated on data from two different institutions in Munich and Rome. In the second project of this thesis, the same cine MRI data was used to extend the models to the prediction of future 2D tumor contours. The LSTM from the previous project combined with a rigid shift of the last available tumor contour was compared to convolutional LSTMs predicting either directly the future contours or future deformation fields which can be used to warp the last available contour. The patient-specific LSTM combined with the rigid shift was found to outperform the convolutional LSTMs on both evaluation sets. In the third project, the patient-specific LSTM for tumor centroid prediction was integrated and validated in a multi-leaf collimator (MLC)-tracking experiment on a research prototype MRI-linac in Sydney. In a phantom study the model was again compared to a population-based LSTM, a patient-specific linear regression and a scenario without any motion prediction. Confirming the in-silico results from the previous project, it was shown that the accuracy with which the irradiation hit the target in the phantom was highest when the patient-specific LSTM was used and that the error compared to the scenario where the latency was not compensated for was nearly halved. In the fourth project of this thesis, the focus was changed from motion prediction to the real-time localization of the tumor on the current cine MRI frame. Specifically, using MRI-linac data acquired in Munich a transformer network was developed for fast deformable image registration (DIR) and compared to an AI auto-segmentation model and a slow conventional

B-spline DIR model. The transformer was optimized with different unsupervised, supervised and patient-specific strategies. The patient-specific transformer was found to perform best, outperforming all other models and optimization strategies without introducing significant additional latency. Future studies could start from the models developed herein to build more advanced AI-based motion management models in 3D to further improve the accuracy of real-time adaptive MRI-guided radiotherapy, as described in the outlook and in the review paper included in the appendix of this thesis.

Chapter 1

Introduction

According to the World Health Organization, in 2022 there were about 20 million new cancer cases and 9.7 million deaths, with about 1 out of 5 people developing cancer in their lifetime [1]. Estimates from the Global Cancer Observatory for the same year show that lung cancer was the most common type of cancer worldwide (12.4%), followed by breast (11.6%) and colorectal cancer (9.6%). In terms of mortality, lung cancer was the leading type (18.7%), followed by colorectal (9.3%) and liver cancer (7.8%). Over the past century, several therapeutic options have been developed to treat cancer, including surgery, radiotherapy and systemic therapies such as chemotherapy or immunotherapy. Often, combinations of therapies are used, with approximately 50% of all cancer patients undergoing radiotherapy at some point during their illness [2].

In modern radiotherapy, the radiation beam can be shaped with millimeter accuracy to deliver the highest possible dose to the tumor while sparing surrounding healthy tissue. Current developments to improve efficacy can be broadly categorized in [3]: 1) biology-driven personalized treatment prescription and 2) technology-driven improvement of irradiation conformity. The former aims at using biomarkers, such as expression of a particular gene or for instance hypoxia levels obtained from imaging, to tailor the prescribed irradiation dose in a patient-specific fashion based on the measured values. The latter aims at improving the delivered dose distributions by either leveraging the different physical interaction with matter of some particles with respect to others (proton or heavy ion therapy instead of photon therapy) or by using imaging such as x-rays, magnetic resonance imaging (MRI) or ultra-sound imaging (US) to guide the radiotherapy treatment at different stages.

The research conducted in this thesis falls in the category of image-guided radiotherapy. More specifically, different deep learning (DL) algorithms were developed to manage motion imaged during radiotherapy delivery fractions. So called intra-fractional motion, which mainly derives from patient respiration, has been shown to decrease the accuracy of radiotherapy [4], as tumors for instance in the lung can move by centimeters while radiotherapy treatment plans assume a static anatomy. Early solutions included the usage of additional margins around the gross tumor volume (GTV), which added to the other margins applied to account for uncertainties in delineation, positioning uncertainty and dose delivery, lead to higher dose in healthy tissues [5]. The last decade has seen the clinical introduction of hybrid MRI-linear accelerators (MRI-linacs), which allow to capture internal motion in real-time during the treatment and, by adapting to it, reduce these treatment margins [6]. The workflow of real-time adaptation includes several challenging steps, from imaging and localizing

the tumor, to compensating for latencies and adapting the radiation beam. Being both fast and accurate, DL motion management models can play an increasingly important role in supporting the different steps of real-time adaptive MRI-guided radiotherapy (MRIgRT), as shown in this thesis.

This cumulative dissertation is structured as follows: in chapter 1 an introduction to radiotherapy in general, MRIgRT, adaptive radiotherapy, and to the DL algorithms, state-of-the-art literature and evaluations metrics relevant to the candidate's work is given. In chapter 2 the articles published in the scope of this thesis are shown. Specifically, in section 2.1 a long short-term memory (LSTM) network developed for the real-time prediction of future tumor centroid positions is presented. In section 2.2 the LSTM was extended to predict the future 2D tumor contour while in section 2.3 the model from the first publication was experimentally validated in a phantom study. In section 2.4 the development of a transformer network for real-time localization of the tumor is described. In section A.1 a review article on the role of DL for real-time motion management during MRIgRT published by the candidate as a result of a workshop organized by the European Society for Radiotherapy and Oncology is shown. Finally, chapter 3 summarizes the results and outlooks possible future studies based on this work.

1.1 Modern radiotherapy

External beam radiotherapy uses a linear accelerator (linac) to treat cancer with ionizing radiation by accelerating electrons in a waveguide to typical energies between 6 and 15 MeV. In most treatments nowadays, the electrons are then converted to photons via bremsstrahlung on a high-density material. The generated photon beam is then flattened (i.e., a uniform dose distribution is achieved) using filters and shaped with millimeter precision using jaws (high-density rectangular blocks) and a multi-leaf collimator (MLC) (multiple thin leaves made of a high-density material that can move independently). The linac is mounted on a gantry that can rotate around the patient lying on a couch to irradiate the tumor from multiple angles [7]. Radiotherapy treatments are usually delivered in multiple sessions called fractions as radiobiological studies have shown this to lead to a higher differential cell killing between cancerous and healthy cells [8].

Technological improvements over the last three decades led to the integration of imaging devices into radiotherapy linacs [9]. Image-guided radiotherapy using x-rays is today's standard-of care: most linacs are equipped with a kV x-ray source (mounted on the gantry 90° from the MV treatment beam), which is routinely used prior to treatment to acquire a cone beam computed tomography (CBCT). CBCTs are volumetric images having lower quality compared to computed tomography (CT) images but which are sufficient to account for inter-fractional changes, i.e., changes in the anatomy occurring from one fraction to the next, as detailed in the next section. X-ray guidance to account for intra-fractional changes is emerging but limited by the low soft-tissue contrast.

1.2 MRI-guided radiotherapy

Over the past decade, treatment machines that integrate MRI with external photon irradiation have been introduced into clinical practice. The primary motivations for adopting MRIgRT are the need to account for anatomical changes both before and during a treatment fraction,

Table 1.1: A summary of integrated MRI-linac designs as of 2024 is provided. The third column details the magnetic field strength, radiation beam energy and orientation of the field relative to the beam (perpendicular/inline). (*) Asterisk denotes that the Siemens MRI-linac was never built, only a patent was issued. Adapted from [6, 10].

Company/institution	Commercial?	MRI and beam specification
Viewray	yes	0.35 T split bore, 6 MV (originally Co-60), perpendicular
Elekta	yes	1.5 T closed bore, 7 MV, perpendicular
MagnetTx	yes	0.5 T biplanar, 4 and 6 MV, inline/perpendicular
Australian MRI-linac	no	1.0 T split bore, 4 and 6 MV, inline/perpendicular
Siemens	no*	0.5 T closed bore, 6 MV (inside bore), perpendicular

and the ability to selectively target the most aggressive and radiation-resistant tumor sub-regions.

The advantage of MRI compared to x-ray based guidance lies in its superior soft-tissue contrast, which comes with no additional dose and allows an excellent visualization of the tumor and surrounding organs at risk (OAR). The integration of an MRI scanner with a linac posed a remarkable engineering challenge which took decades to complete and has led to the development of a few different MRI-linac designs, as can be seen in Table 1.1. In the following, detailed descriptions of the two first commercial systems, the 0.35 T Viewray system and the 1.5 T Elekta system, and of the 1.0 T Australian system are provided, the Viewray and the Australian MRI-linac having been used in this thesis.

The Viewray-MRIdian system, originating from early work at the University of Florida and clinically implemented in 2014, represents the first commercial MRI-guided external beam radiotherapy device [11]. Initially featuring three Co-60 radiation sources mounted 120° apart on a ring gantry, it later transitioned into a 6 MV linac-based delivery system leveraging a double stack MLC configuration to shape the beam (leaf travel direction perpendicular to B_0). With a B_0 magnetic field of 0.35 T, the MRIdian system employs for the magnet a split-bore design with a vertical gap, the linac and MRI being electro-magnetically isolated from each other using shielding components (multi-layer ferromagnetic and carbon buckets) mounted around the gantry, as visible in Figure 1.1. The radiation beam coming from the linac rotates around the stationary magnet and remains always perpendicular to the B_0 magnetic field. This minimizes skin dose as contaminant electrons generated in the linac head are swept away. However, these contaminant electrons can land somewhere else on the patient's surface. Compared to inline configurations (B_0 parallel to radiation beam), it also presents dosimetric challenges related to the secondary electrons generated inside the patient which get deviated by the magnetic field [12, 13]. A MRIdian system was installed at the LMU University Hospital and started treatments beginning of 2020. As of 2024, about 60 machines are in clinical use worldwide.

Clinically used for the first time in 2017, the second commercial MRI-linac is the Elekta-Unity system [12]. The system features a 7 MV linac mounted on a rotating ring within a 1.5 T MRI system, as shown in Figure 1.2. As the MRIdian, the Unity system employs passive and active shimming to maintain magnetic field homogeneity and its coil configuration was tuned to create a low magnetic field region in which the most sensitive linac components were placed. The linac rotates within this low field region with the radiation passing through the magnet's cryostat, which leads to attenuation and explains the slightly higher linac energy of

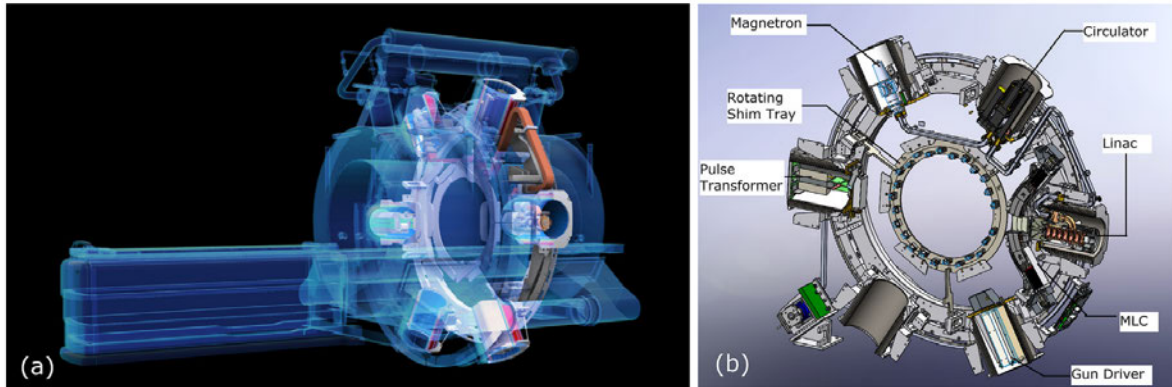


Figure 1.1: (a) The Viewray-MRIdian MRI-linac system consisting of superconducting split-bore magnet, rotating circular radiation gantry and patient couch; (b) detailed drawing of the gantry with linac components and shielding buckets. Reproduced from [13].

7 MV compared to the MRIdian system. The MLC is similar to the one used in conventional Elekta linacs, however, it was rotated by 90° such that the leaf motion is in the superior-inferior direction, facilitating the MLC-tracking in this main direction of motion planned in a future upgrade. The Unity system's 1.5 T MRI provides higher signal compared to the lower field MRI-linacs. To prevent significant image quality degradation, the noise from the linac components is minimized through a specially designed Faraday cage, which isolates the gantry (with all linac components) from the MRI. As of 2024, about 80 machines are in clinical use worldwide.

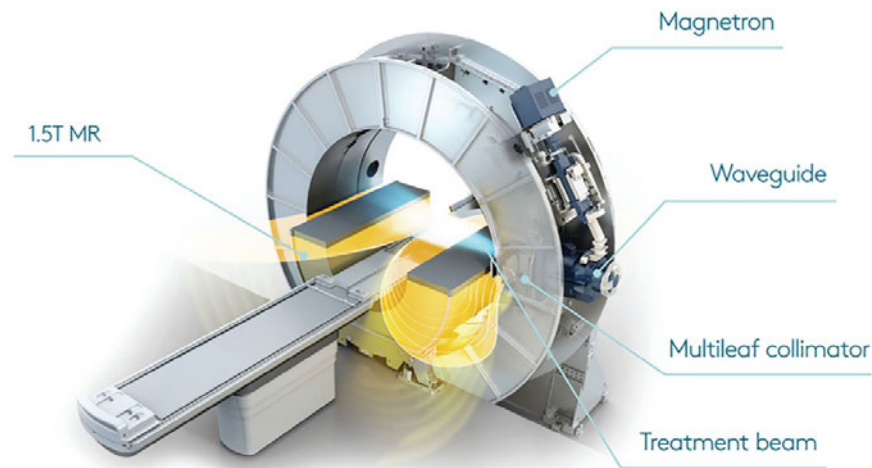


Figure 1.2: Illustration of the Elekta-Unity MRI-linac system with superconducting closed bore magnet, rotating circular radiation gantry and patient couch. B_0 iso-field lines are indicated in yellow. Reproduced from [12].

Depicted in Figure 1.3, the MRI-linac from the Australian MRI-linac Program features a 1.0 T split bore magnet combined with an inline 6 MV industrial linac placed on rails at a

variable distance between 1.8 and 3.2 m [12]. Active shielding was designed to have a low magnetic field at the linac's location. The radiation beam can be shaped with a 120 leaves MLC placed right after the linac in the region of low B-field. As for the Unity and in contrast to the MRIdian, the leaf motion is parallel to the patient's superior-inferior direction. The system presents a 50 cm gap between the two magnet halves which is smaller than that in modern computed tomography (CT) and MRI devices and was chosen as a trade-off between patient comfort and imaging quality. A rotating couch can be inserted in the gap to allow treatments of the patient from multiple angles. As the patient is immobilized and rotated this approach leads to more discomfort but it has the advantage of being more cost effective compared to the gantry approaches of other MRI-linac systems [14]. An electronic portal imaging device (EPID) can be mounted along the beam-line behind the patient and the second magnet half to measure the attenuated radiation beam. A single prototype of the system was installed at the Liverpool Hospital in Sydney and acquired first images in 2016. Being a prototype, it provides easier access for integrating novel developments for pre-clinical research.

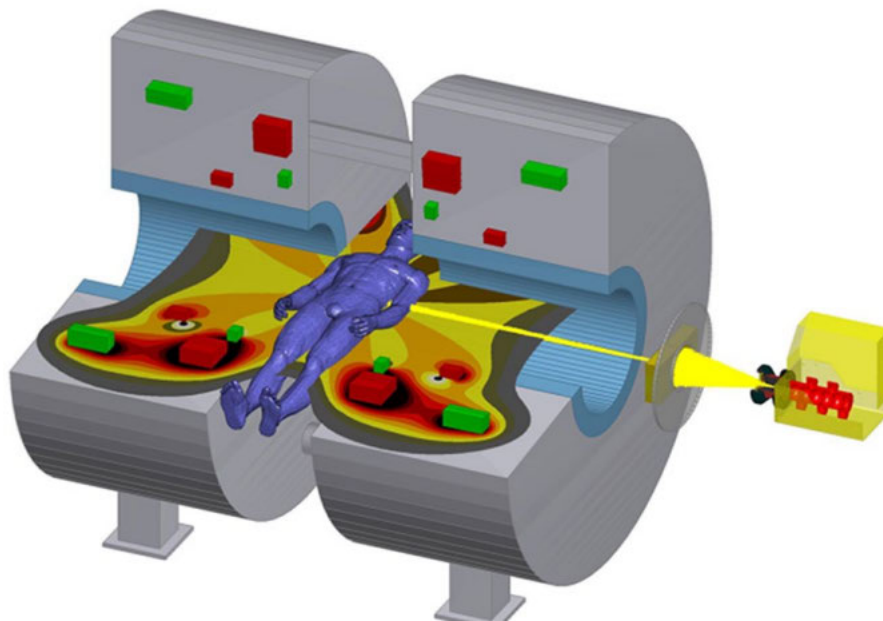


Figure 1.3: Illustration of the Australian MRI-linac system with superconducting double-donut magnet and inline (parallel to B_0) linac without gantry. An MLC is positioned between the linac and magnet (leaf motion in superior-inferior direction) while the patient lies on a rotating couch (not shown) which can be inserted between the two magnet halves. Reproduced from [12].

MRI-linac systems are considered the state of the art when it comes to adaptive radiotherapy. Broadly speaking, the aim of adaptive radiotherapy is to tailor the irradiation to changes in the patient's anatomy, which includes for example plan adaptation to increase the dose to the GTV if the daily OAR position allows it or adaptation of the beam in real-time based on the patient's respiration. While in standard radiotherapy a treatment plan is calculated only once based on a single planning CT and delivered without changes over

several days or weeks, in adaptive radiotherapy the treatment plan can be changed during the irradiation fraction. The evolution of adaptive radiotherapy is closely linked to the one of image-guided radiotherapy as adaptation is based on changes observed with imaging. Based on the timescale of the adaptation, adaptive radiotherapy can be divided in three categories [12]:

- Offline adaptation (days)
- Online adaptation (minutes)
- Real-time adaptation (seconds)

Early examples of adaptive radiotherapy feature the usage of in-room CBCT scanners on standard linacs for offline adaptation of head and neck cancer patients. If on the daily CBCT a substantial weight loss or lesion shrinkage was observed, the original plan would be irradiated on that fraction but the patient would then be sent to acquire a new planning CT for which a new treatment plan is calculated. Offline adaptation therefore takes place after the current treatment fraction and typically follows the same clinical workflow as the initial treatment planning. In contrast, online adaptation takes place prior to the start of the daily treatment fraction, while the patient is on the couch, and is possible only with dedicated machines as rapid 3D imaging, re-planning, plan review and patient-specific quality assurance are needed. Similarly, dedicated machines are needed for real-time adaptation which is performed without any therapist intervention in seconds or even milliseconds during the course of the treatment.

In the following two subsections, a more detailed description of online adaptation and real-time adaptation, with focus on MRI-linac implementations, is given.

1.2.1 Online adaptation

During online adaptation, the patient's treatment plan is adjusted to account for inter-fractional changes in the anatomy such as tumor shrinkage or OAR configurations which differ from the planning CT (e.g., different rectal filling). Currently, MRI-guided online adaptive radiotherapy is clinically applied using either the Viewray-MRIDian or the Elekta-Unity system. Being more relevant for this thesis, a detailed description of the MRIdian workflow is provided in the following together with an illustration in Figure 1.4. The online adaptive workflow on the Unity does not differ substantially.

As for radiotherapy on conventional linacs, an initial planning CT providing the electron density information is required to calculate the treatment dose distribution in the patient. Additionally, a volumetric planning MRI scan is acquired on which the target/OAR contour delineation is performed. The planning CT is then deformably registered to the MRI to generate the so called synthetic CT. The synthetic CT provides both segmentations and electron density information so it can be used to perform the treatment plan optimization.

On the day of the treatment, the process of on-table adaptation involves several steps [6]. First, an in-room volumetric MRI is acquired to visualize the anatomy of the day. The synthetic CT from the planning stage is then used to obtain a segmented in-room synthetic CT by performing a DIR with respect to the in-room MRI. This provides an up-to-date 3D electron density image of the patient in treatment position, and in theory also up-to-date contours. However, in clinical practice a manual adjustment of the contours by the physicians is always performed. As this is time intensive, physicians usually focus only on the relevant region around the irradiation target with DL auto-segmentation models having been proposed

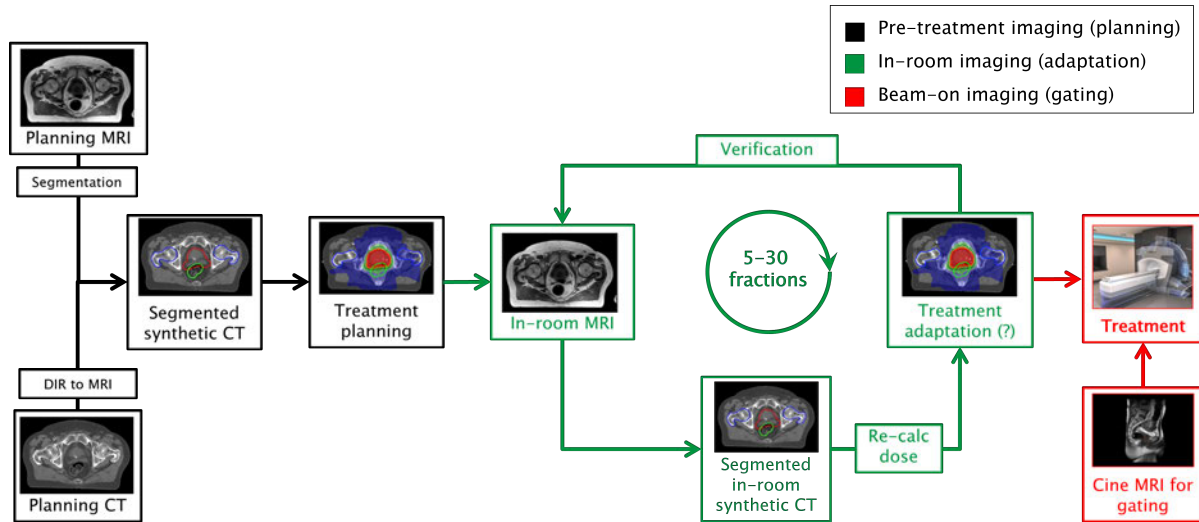


Figure 1.4: Viewray-MRIdian online adaptive MRIGRT workflow. Figure courtesy of PD Dr. Christopher Kurz.

to accelerate this process [15, 16]. After the correction of the contours, a re-calculation of the baseline treatment plan on the daily anatomy is performed to determine whether treatment adaptation is required. A fast dose engine capable of incorporating magnetic fields (Monte Carlo-based) is used for the calculation. Critical parameters such as tumor coverage and dose in the OARs are then compared to the baseline plan and depending on the agreement the physicians might opt for no adaptation or for a full re-optimization of the treatment plan by the physicists in the team. For quality assurance of the newly optimized plan, the MRIdian provides a secondary Monte Carlo dose calculation that can as well incorporate the magnetic field [13]. Shortly before the treatment starts, a so called preview 2D cine MRI is acquired in preparation for the real-time adaptation with gating during the irradiation, as described in the next subsection. The tracking target, the desired gating margin and the percentage of the target allowed outside of the margin before the beam is automatically stopped are also defined at this stage. The online adaptation steps prior to the start of the irradiation have been reported to take up to 30 min on the MRIdian system [17].

1.2.2 Real-time adaptation

During real-time adaptation, the radiation beam is adjusted to compensate for motion occurring during the course of a treatment fraction, for example cardiac-induced or respiratory motion affecting the target and OARs. Clinically, until 2023 the Viewray-MRIdian was the only system providing actual beam adaptation during the treatment based on MRI-guidance [18]. Recently, the first clinical experience of intra-fractional beam adaptation was reported for the first Elekta-Unity system [19], all other 1.5 T clinical systems before this real-time adaptation-enabling upgrade being only able to visualize the motion.

Figure 1.5 also illustrates an optimal real-time adaptive MRIGRT workflow. No single MRI-linac system currently provides the displayed workflow, however, all steps have been separately shown to work either clinically or with prototype machines in a research setting

[20]. As this is the focus of the candidate’s publications, a more comprehensive description of the various techniques used for real-time motion management in MRIGRT is presented below, covering both clinically available methods and those currently under investigation in research settings for each step of the motion management process.

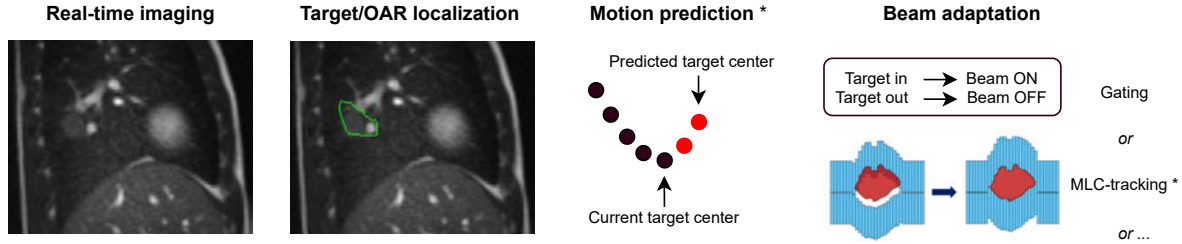


Figure 1.5: Real-time adaptive MRIGRT workflow. (*) Asterisks denote that motion prediction and MLC-tracking have currently been implemented only on prototype MRI-linacs. Reproduced from the candidate’s review article (section A.1) [20].

Real-time imaging

The first step involves the real-time visualization of the patient anatomy using cine MRI with dedicated sequences having short repetition times and which are therefore robust to motion. MRI-linac vendors currently offer cine MRI in either a single 2D sagittal plane or multiple parallel or orthogonal 2D planes. Balanced steady-state free precession sequences with in-plane pixel sizes between 2 and 4 mm, slice thicknesses between 5 and 15 mm and frame rates between 2.5 and 8 Hz are used in clinical practice [20]. The imaging plane is placed in the tumor or a surrogate tracking target in case the tumor is too hard to see to visualize the internal motion and adapt the irradiation to it.

Ideally, continuous 3D+t cine MRI at a frame rate of few Hz and a voxel size of few mm^3 should be used to obtain full anatomical information. However, to achieve this in real-time is a challenge with the current technology and a topic of active research [21]. For instance, 3D+t cine MRI at a frame rate of about 2 Hz but coarse spatial resolution of about 5 mm in all directions has been shown feasible using current commercial acceleration methods [22]. DL solutions to improve image reconstruction of under-sampled data or to convert low-resolution images to high-resolution are also emerging [23, 24].

Target/OAR localization

The second step involves localizing the treatment target or a surrogate thereof and if needed nearby OARs every time a new cine MRI frame is acquired. Various fast algorithms have been proposed for this task.

Clinically, both the Viewray-MRIdian and, more recently, the Elekta-Unity system provide real-time target localization (often referred to as target tracking) based on image registration. The Viewray algorithm first used in 2017 relied on multi-scale deformable image registration (DIR) of a reference frame. The reference frame is automatically selected from the preview cine MRI based on image similarity with a 2D sagittal slice from the 3D daily static MRI [13]. During treatment, the reference frame is in turn deformably registered to the acquired

cine MRI frames (and simultaneously also the other way around to ensure robustness) by minimizing a cross-correlation cost function with gradient magnitude regularization [25]. In a more recent version of the algorithm, multiple trackers with individual cost functions and regularization functioning both locally and globally are implemented and combined [26]. Four different tracker combinations are available for the clinical team to choose from: 1) "Large and Deforming Targets" (three global trackers), 2) "Small Mobile Targets" (three local trackers), 3) "Complex Mobile and Deforming Targets" (three global and three local trackers) and 4) "Default" (one global and two local trackers). Depending on the motion characteristics and the current accuracy, the team is able to freely switch from one algorithm to another during treatment. Since early 2023, a tracking algorithm for the Elekta-Unity has been introduced [27]. Their model requires a 3D daily MRI-linac scan and a pre-treatment cine MRI on which template image pairs are generated in both coronal and sagittal orientation. The template images are then matched to newly acquired frames to rigidly shift a target mask during treatment. While only employed for research, the Australian MRI-linac prototype also features a template matching algorithm to localize the target centroid on newly acquired 2D sagittal cine MRI frames in real-time. Finally, also for target localization, different DL algorithms have been proposed and shown to achieve superior performance to conventional image registration algorithms in-silico [28, 29], as detailed in section 1.4.

Motion prediction

Each step in the workflow requires a certain time to complete, which leads to a time delay between the actual physical motion of the target and the final beam adaptation step. This time delay is called the system latency. Experimental studies have been performed for the different MRI-linac systems and found a latency of 300 ms for the Viewray-MRIdian (clinical system with gating) [30], 350 ms for the Elekta-Unity (prototype with MLC-tracking) [31] and 330 ms for the Australian system (prototype with MLC-tracking) [32]. The Australian researchers have developed a straightforward method to measure the system latency on their MRI-linac: a target in a phantom is moved using a sinusoidal trace while it is irradiated using MLC-tracking. Using the EPID, the sinusoidal of the target in the phantom and the sinusoidal from the MLC aperture can be extracted and the phase shift between the two curves can be calculated to obtain the latency. The main contributor to the system latency is the imaging step with about 200-250 ms (with half of the time lag for this step coming from the low frame-rate of 4 Hz), target localization taking about 20-50 ms and beam adaptation 20-110 ms depending on the adaptation strategy (see next subsection).

To compensate for the latency, respiratory motion prediction algorithms can be used: by predicting for instance the future position of the target, the beam adaptation step can be performed based on the predicted position and therefore in synchrony with the actual motion. Several algorithms have been proposed in-silico for the prediction of the future target center-of-mass or the next 2D frame, as discussed in section 1.4. In general, conventional linear regression is very promising for latencies up to 250 ms but has been shown to be outperformed by DL models such as LSTMs for larger forecasts [33]. One advantage of linear regression is the existence of an analytical solution to calculate its parameters, which demands minimal computation time. This has led to the development of models that can be regularly updated using real-time patient-specific motion data during the treatment [34]. Such a continuously updated linear regression predictor has been used by Uijtewaal et al. to successfully compensate for the latency and improve the dosimetric accuracy in a phantom

study with a prototype Elekta-Unity system using 4 or 8 Hz imaging in combination with MLC-tracking [35]. Clinical systems do not compensate for latencies, however, it can be argued that there is a negligible dosimetric error if gating is used as beam adaptation strategy, as detailed in the next subsection.

Beam adaptation

On MRI-linacs, the two main systems of adaptation to improve beam-target alignment are beam gating and MLC-tracking. These systems are independently controlled and could therefore also be used in parallel.

Beam gating enables treatment of the target in a fixed spatial position, though at the expense of longer treatment duration. The beam is automatically switched on once the tumor or a surrogate volume enters a predetermined area, and switched off when it moves outside. Gating can be applied during free breathing or using breath-hold methods (e.g., for lung treatments). While it has been shown that gating achieves dose conformity similar to static scenarios, it leads to extended treatment times as the target needs to be inside the predetermined area. In literature, duty cycle efficiencies (the ratio of beam-on time to treatment duration) ranging from 20% to 55% have been reported for clinical systems treating with breath-holds [36, 37]. The gating latency can be divided in two parts: the time to switch the beam off when the target moves outside the boundary area and the time to switch the beam on when the target moves inside the area. The former is of more concern as it leads to less accurate irradiation of the target while the latter only reduces the duty cycle efficiency but is nonetheless important for patients with limited breath-hold ability.

MLC-tracking is a more advanced form of beam adaptation in which the leaves shaping the beam are continuously realigned to compensate for target motion. The realignment process can be both rigid, i.e., the entire MLC bank is shifted or deformable, i.e., single leave positions are re-optimized in real-time to account also for target deformation or rotations. Compared to gating which has been used clinically since the first treatments with the Viewray-MRIdian, MLC-tracking has only been used in experimental studies with phantoms. Uijtewaal et al. showed in two studies with a prototype Elekta-Unity system that MLC-tracking combined with different radiation delivery techniques (step-and-shoot intensity-modulated radiation therapy and volumetric-modulated arch therapy) results in highly conformal dose distributions without the cost of additional treatment time [35, 38]. In another phantom study, Liu et al. have shown how simultaneous adaptation to multiple independently moving targets is possible using MLC-tracking on the Australian MRI-linac system. All these studies have underlined how latency is more problematic when using MLC-tracking. Compared to gating, where the beam-off latency leads to incorrect irradiation only in about one frame every time the target exits the boundary area, MLC-tracking with latency leads to incorrect irradiation in most of the treatment, as the leaves constantly lag behind the current target position. Motion prediction as discussed in the previous subsection is therefore more critical with this adaptation strategy. Finally, it can be noted that DL models for the MLC position optimization problem are not needed as accurate analytical solutions, which can be completed in the order of milliseconds, exist [39]. In contrast to MRIGRT, early clinical experiences using MLC-tracking on conventional linacs have been reported [40, 41]. Both studies relied on electromagnetic transponders implanted into the patients' prostate or lung to accurately localize the irradiation target and reported delivered doses similar to the static treatment plan without a prolongation of the irradiation due to beam pauses.

1.3 Deep learning

Many different definitions exist for natural intelligence. If we define natural intelligence as the ability to learn and solve complex problems, then artificial intelligence (AI) can be defined as the simulation of natural intelligence with machines. With first studies dating back to the 1950s, in the last two decades the development of AI models to solve problems including but not limited to perception, recognition, analysis, and decision-making has become increasingly crucial across various fields [42].

To date, machine learning (ML) is considered a sub-category of AI, encompassing numerical algorithms and models designed to analyze data and acquire decision-making capabilities for accomplishing specific tasks [43]. In other words, ML tries to find hidden patterns from data to solve a problem of interest.

DL is in turn considered a sub-category of ML, sharing the same goals of ML but being characterized by more advanced algorithms with more parameters which has lead to breakthroughs in performance in processing images, video, speech and audio [44]. This was enabled by the recent availability of large-scale datasets and technological advances in parallelized computing.

Applications of AI and in particular DL in medicine are manifold, and include diagnosis of diseases, drug discovery, medical robots and outbreak prediction [45]. Also the field of medical physics has seen a steep rise in the utilization of DL to solve various tasks.

Specifically for MRIGRT, DL has found relevant applications in image segmentation, synthetic CT generation, automatic online treatment planning and outcome prediction [46]. Considering the complex treatment workflows, AI is expected to rapidly contribute to MRIGRT, primarily by either safely and efficiently automatising the various manual operations during online adaptation or as a fast and more accurate alternative to conventional algorithms when considering the time constraints of real-time adaptation.

The candidate's publications fall in the category of DL applications for real-time adaptation to motion in MRIGRT. In the two following subsections, the DL algorithms which were found to perform well are described.

1.3.1 Long short-term memory networks

Recurrent neural networks (RNNs) are AI algorithms designed for processing sequential input data, such as time series data. Unlike commonly used regression or classification AI models such as the artificial neural network (ANN) or convolutional neural network (CNN), they include an additional set of weights that connects hidden layers from one time step to the next, which allows them to capture temporal dependencies more efficiently. However, the original RNN architecture was found to be limited due to instability when training over longer input sequences [42]. Consequently, more sophisticated architectures have been proposed, with the most widely used RNN being the LSTM model [47].

The LSTM module was specifically developed to facilitate the learning of longer sequences of data more effectively and is shown in Figure 1.6. The core idea of LSTM is the introduction of the memory cell state c^t , which enables more stable training (back-propagation of errors) and a smooth flow of information. To control the addition or removal of information from the cell state, so called gates are utilized: The forget gate f^t is used for deciding whether to retain or discard past information in c^{t-1} , while the input gate i^t regulates the flow of information into the new memory cell state $\tilde{c}^t$. These gates, along with the previous memory,

are combined to create the final memory cell state $\mathbf{c}^t$. This final memory cell state is then filtered using the output gate $\mathbf{o}^t$ to determine which information for making predictions to retain or discard, ultimately producing the hidden state $\mathbf{h}^t$. In mathematical terms, at a given time step t , the LSTM module can be described as follows:

$$\text{Forget gate: } \mathbf{f}^t = \sigma(\mathbf{W}_f \mathbf{x}^t + \mathbf{U}_f \mathbf{h}^{t-1} + \mathbf{b}_f) \quad (1.1)$$

$$\text{Input gate: } \mathbf{i}^t = \sigma(\mathbf{W}_i \mathbf{x}^t + \mathbf{U}_i \mathbf{h}^{t-1} + \mathbf{b}_i) \quad (1.2)$$

$$\text{New memory cell state: } \tilde{\mathbf{c}}^t = \tanh(\mathbf{W}_c \mathbf{x}^t + \mathbf{U}_c \mathbf{h}^{t-1} + \mathbf{b}_c) \quad (1.3)$$

$$\text{Final memory cell state: } \mathbf{c}^t = \mathbf{f}^t \odot \mathbf{c}^{t-1} + \mathbf{i}^t \odot \tilde{\mathbf{c}}^t \quad (1.4)$$

$$\text{Output gate: } \mathbf{o}^t = \sigma(\mathbf{W}_o \mathbf{x}^t + \mathbf{U}_o \mathbf{h}^{t-1} + \mathbf{b}_o) \quad (1.5)$$

$$\text{Hidden state: } \mathbf{h}^t = \mathbf{o}^t \odot \tanh(\mathbf{c}^t) \quad (1.6)$$

where $\mathbf{b}$, $\mathbf{W}$ and $\mathbf{U}$ denote the biases, input sequence weights and recurrent weights which

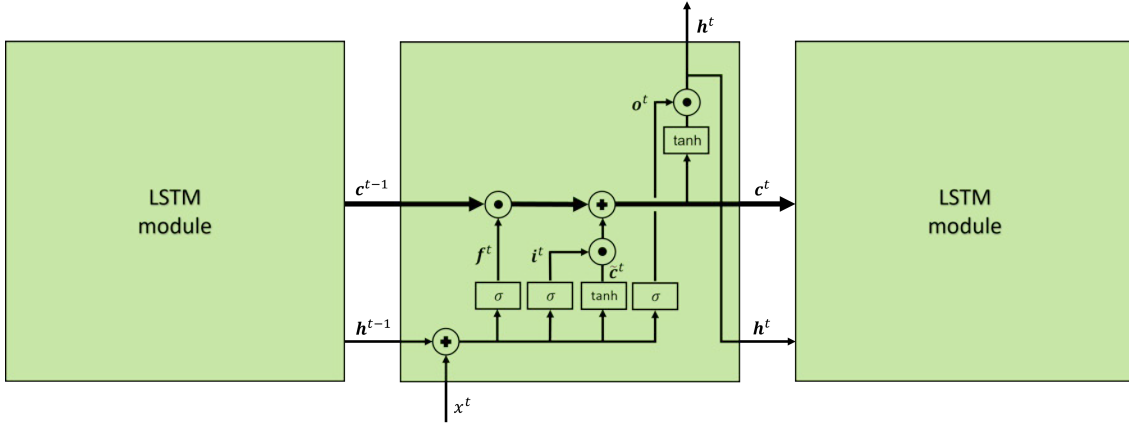


Figure 1.6: Illustration of repeating LSTM module with different gates. Reproduced from the candidate's Paper I (section 2.1) [48].

are learned during the training process. The symbol $\odot$ denotes element-wise multiplication between matrices or vectors. The sigmoid function $\sigma(x)$ is usually employed for the gates, while the hyperbolic tangent function is used for the states. At each time step, the hidden state of one LSTM layer serves as the input for the subsequent LSTM layer. In the final hidden layer of the LSTM, the hidden state from the last time point t_f is fed into a fully connected layer to obtain the predicted output sequence:

$$\text{Predicted output: } \hat{\mathbf{y}}_i = \mathbf{W}_{\text{FC}} \mathbf{h}^{t_f} + \mathbf{b}_{\text{FC}} \quad (1.7)$$

where $\mathbf{W}_{\text{FC}}$ and $\mathbf{b}_{\text{FC}}$ denote the weight matrix and bias vector for the fully connected layer.

While the LSTM module presented above performs well with scalar sequences, its fully connected nature leads to redundancy when dealing with sequences of images. Therefore, Shi et al. [49] proposed the convolutional LSTM: by replacing matrix multiplications between the hidden and cell states with convolutions, they found the model to capture spatiotemporal correlations more efficiently. Mathematically, the convolutional LSTM module is defined as

follows:

$$\text{Forget gate: } f^t = \sigma(W_f * x^t + U_f * h^{t-1} + b_f) \quad (1.8)$$

$$\text{Input gate: } i^t = \sigma(W_i * x^t + U_i * h^{t-1} + b_i) \quad (1.9)$$

$$\text{New memory cell state: } \tilde{c}^t = \tanh(W_c * x^t + U_c * h^{t-1} + b_c) \quad (1.10)$$

$$\text{Final memory cell state: } c^t = f^t \odot c^{t-1} + i^t \odot \tilde{c}^t \quad (1.11)$$

$$\text{Output gate: } o^t = \sigma(W_o * x^t + U_o * h^{t-1} + b_o) \quad (1.12)$$

$$\text{Hidden state: } h^t = o^t \odot \tanh(c^t) \quad (1.13)$$

where b, W and U denote the biases, input sequence convolutional kernels and recurrent convolutional kernels which are learned during model training. The symbol $*$ represents the convolution operation.

1.3.2 Transformer networks

Even though the LSTM was designed to better learn longer sequences, it was realized that the longer the input the harder it is for the model to focus on elements lying further in the past: as all information needs to be encoded in the memory cell state vector, this leads to a bottleneck. One of the attempts to tackle the problem resulted in the introduction of a mechanism at the recurrent neural network's bottleneck that allows the model to pay attention to the most relevant parts of the encoded input [50].

In a seminal paper, Vaswani et al. [51] proposed to get rid of all recurrent neural network components and to implement an architecture which is solely based on the attention mechanism. They called the proposed machine translation model the transformer. As illustrated in Figure 1.7, their model has an encoder-decoder structure. On a high-level, the encoder made up of N identical layers is used to map an input sequence into an abstract representation. Each layer has two sub-layers, the so called multi-head attention layer and a simple feed-forward layer, both sub-layers being combined with layer normalization and residual connections to facilitate model training. The abstract representation from the last encoder layer is then used in the decoder to generate the output sequences one element at the time. The decoder is called auto-regressive, as previously generated elements are used as additional input to the decoder when generating the next element. Also the decoder contains several stacked layers with a slightly different version of multi-head attention and feed-forward sub-layers.

One advantage of attention compared to recurrent neural networks is that all input elements, or tokens, can be processed at once in a highly efficient way. In the following, a mathematical description of the attention mechanism is provided. Given the input sequence of embedded tokens (e.g., words) $\hat{X} \in \mathbb{R}^{N \times D}$ with N the number of tokens and D the embedding dimension (set in the word embedding algorithm which is separate from the transformer), the first step is to retain the positional information of the single input tokens, as these are processed all at once. For that, Vaswani et al. proposed a fixed sinusoidal positional encoding scheme which is simply added to the input embedding

$$X = \hat{X} + E_{\text{pos}} \quad (1.14)$$

with $E_{\text{pos}} \in \mathbb{R}^{N \times D}$. However, the authors note that using a randomly initialized E_{pos} matrix which is learned during model optimization leads to similar results. The embedded and

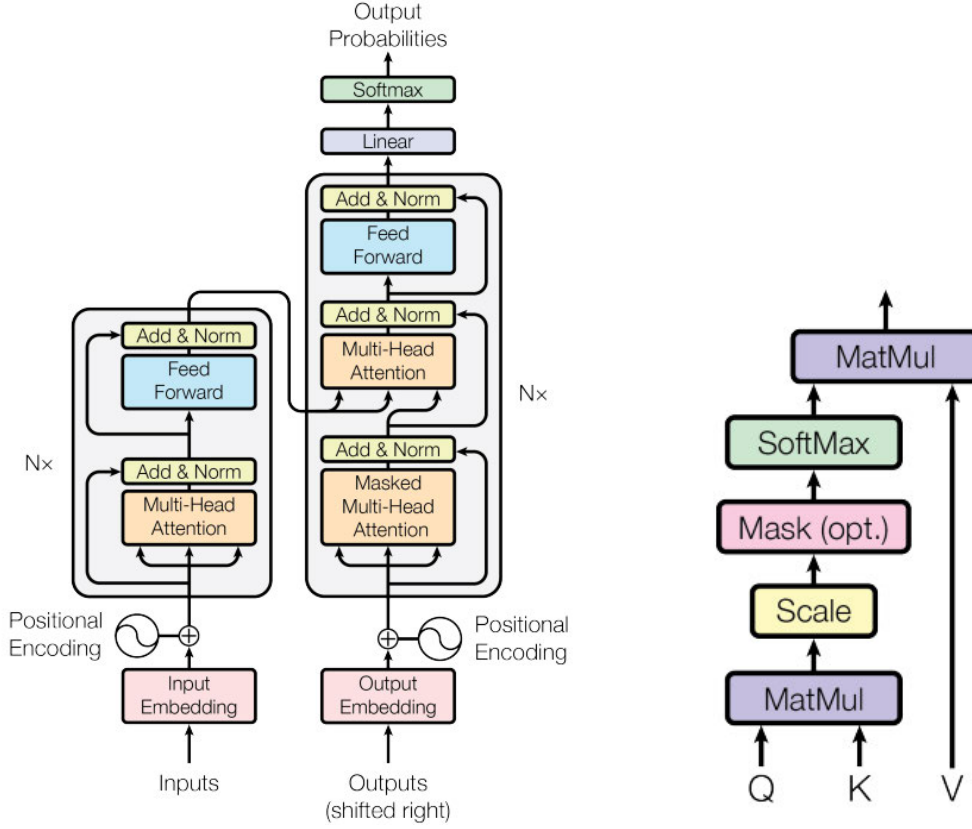


Figure 1.7: Illustration of the transformer architecture (*left*) and the attention mechanism (*right*). Reproduced from [51].

positionally encoded tokens are then linearly transformed to build the queries, keys and value matrices $Q, K, V \in \mathbb{R}^{N \times D_{qkv}}$

$$[Q, K, V] = XW_{q,k,v} \quad (1.15)$$

with the learnable weight matrices $W_{q,k,v} \in \mathbb{R}^{D \times D_{qkv}}$ and D_{qkv} being the new embedding dimension of queries, keys and values. Queries and keys are then multiplied, scaled and normalized to get the attention matrix $A \in \mathbb{R}^{N \times N}$

$$A = \text{softmax} \left(\frac{QK^T}{\sqrt{D_{qkv}}} \right) \quad (1.16)$$

The attention matrix is in turn multiplied with the values to get the self-attention matrix $SA \in \mathbb{R}^{N \times D_{qkv}}$

$$SA = AV \quad (1.17)$$

The intuition behind the usage of the queries, keys and value matrices comes from retrieval systems. The queries pinpoint what we are interested in (e.g., a specific topic we are looking for in a book), the keys provide context and relevance (e.g., the summary of a book) and the values give us the actual information we need (e.g., the detailed information in that book).

When computing self-attention, information from multiple tokens (e.g., books) is obtained to understand which tokens are more relevant than others [52]. In practice, multiple attention heads are computed in parallel with each of the h heads being randomly initialized with a separate weight matrix. The set of weight matrices $\mathbf{W}_O \in \mathbb{R}^{hD_{qv} \times D}$ is then multiplied with the self-attention matrix obtained from each head to obtain a single multi-head self-attention matrix $\mathbf{MSA} \in \mathbb{R}^{N \times D}$

$$\mathbf{MSA} = \text{concat}(\mathbf{SA}_1, \dots, \mathbf{SA}_h) \mathbf{W}_O \quad (1.18)$$

This way, the transformer gets increased representational power (more learnable weights) in an efficient way (the head computations can be parallelized). The two (multi-head) attention mechanisms in the decoder only present small differences with respect to the self-attention computation described above. When computing masked self-attention (also called causal self-attention), all entries in the attention matrix $\mathbf{A}$ above the diagonal are set to zero to avoid that the decoder has access to future words. When computing the attention that connects the encoder to the decoder, which is often called cross-attention, the keys and values matrices come from the encoder while the queries come from the decoder in contrast to encoder self-attention where all matrices come from the input tokens.

Since their publication in 2017, transformers have been replacing other DL algorithms in a wide range of tasks, from natural language processing where they were first proposed to computer vision, audio recognition or healthcare [53]. A key idea allowing the spread of transformers in many domains was presented by Dosovitskiy et al. when applying transformers for image classification [54]: by splitting each input image into smaller patches and treating each patch as an input token, they could leverage the attention mechanism without the computational overhead which would be obtained if every single pixel was considered as a token. While this works well for classification tasks, a problem with this approach arises when the model's prediction is on a dense pixel level, such as for segmentation tasks. To solve this issue, Liu et al. proposed the shifted-window (swin) attention mechanism, in which the patches are organized into windows where attention is computed only locally, which is less computationally expensive [55].

1.4 State-of-the-art literature

The candidate's work focused on the application of DL algorithms to motion prediction in the first three papers and to target localization in the last paper. In the following, an overview of state-of-the-art methods related to these two topics in the medical imaging and radiotherapy field is provided.

In a comparative study from 2019, Joehl et al. optimized a total of 18 respiratory motion prediction algorithms based on 93 1D motion traces obtained from LED markers externally attached to patients' chests [57]. Prediction horizons of 160 ms and 480 ms were investigated for traces resampled to a frequency of 25 Hz. The implemented algorithms included conventional algorithms such as linear regression, support vector regression and Kalman filters as well as an ANN. For all models a patient-specific optimization was performed, i.e., the data values used for training were updated at every time step by adding the current and discarding the oldest time sample of the current patient. They found that for both prediction horizons linear algorithms such as linear regression were sufficient for accurate respiratory motion prediction. For this reason, a linear regression was implemented as the baseline for comparisons in the motion prediction publications in this thesis which used 1D traces as

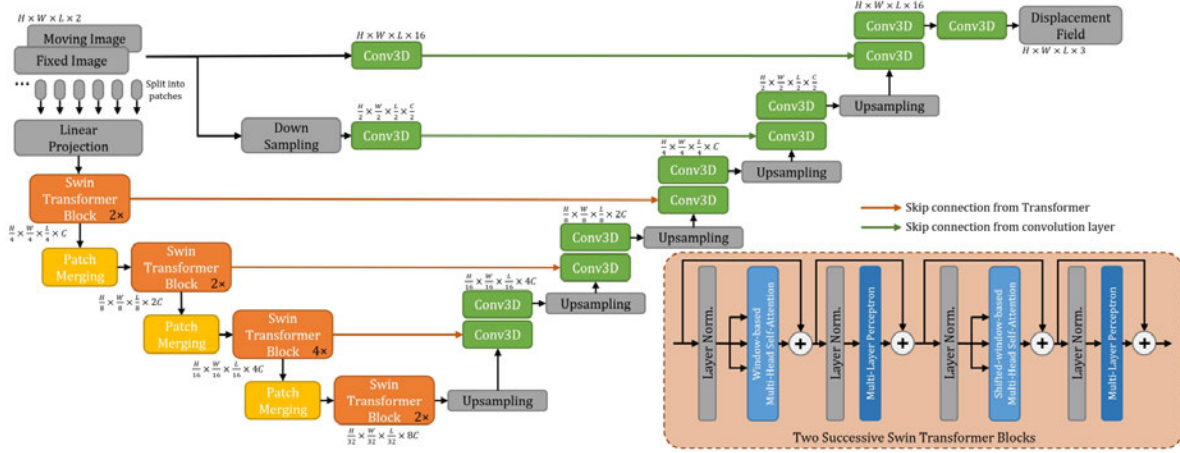


Fig. 1: The architecture of the proposed TransMorph registration network.

Figure 1.8: Illustration of the TransMorph architecture for 3D deformable medical image registration. Reproduced from [56].

input (section 2.1 and section 2.3). Also in 2019, a study showing the superiority of LSTM networks in predicting respiratory motion compared to ANNs was published by Lin et al. [58]. The authors used a total of 1703 1D motion traces obtained from markers positioned on the patients' chest or abdomen and localized with an infrared camera and could show that tuning of the LSTM hyper-parameters could improve the performance by up to 20%. However, they did not optimize the models in a patient-specific fashion. The theoretical advantage of LSTMs when dealing with sequential data and the promising performance in a radiotherapy setting found by Lin et al. lead to the usage of an LSTM network in the first paper in this thesis (section 2.1). Of inspiration for the candidate's second paper (section 2.2) was the model developed by Romaguera et al., who instead of predicting future 1D target positions implemented an encoder-decoder convolutional LSTM for the prediction of future 2D deformation fields [59]. Specifically, they could show that the proposed model outperformed both a conventional principal component-based and a DL-based spatio-temporal prediction algorithm, achieving vessel position accuracy in the next temporal image (320-400 ms) below 0.5 mm for MRI, CT and US datasets affected by respiratory motion. However, the cine MRI data used in this study was from healthy subjects imaged on a 3 T diagnostic scanner. Of relevance for the third publication in this thesis (section 2.3) was the experimental work published in 2020 by the research group in Sydney on MLC-tracking with the Australian MRI-linac [32]. The authors showed in a phantom study that it is possible to simultaneously track multiple independent targets moved with 1D motion traces from lung and prostate cancer patients. They showed the accuracy of the irradiation to be significantly improved compared to a scenario without MLC-tracking and identified in the system latency the major remaining source of error for MLC-tracking with lung motion traces.

Looking at target localization for MRI-linacs, the work by Friedrich et al. [28] from 2021 and the work by Hunt et al. [29] from 2023 should be mentioned. Friedrich et al. implemented an auto-segmentation CNN and found it to outperform a conventional B-spline DIR model when applied to under-sampled 2D cine MRI data from a 0.35 T Viewray-MRIdian

machine. The CNN was trained in a patient-specific fashion using the first 10 frames from each patient's cine MRI, however, only three manually contoured liver cancer patients were investigated in total. Hunt et al. also implemented a CNN but for DIR (VoxelMorph) of 2D cine MRIs from a Viewray-MRIdian. They found their model to outperform affine, B-spline and Demons image registration methods on a large set of about 629,000 frames. However, they evaluated their models only using image metrics in contrast to Friedrich et al. which used more relevant target contour metrics (see next section). Of relevance for the candidate's fourth publication (section 2.4) was also the model published by Chen et al. for fast DIR of 3D medical images [56]. As can be seen in Figure 1.8, the authors leveraged swin transformer blocks in the encoder to efficiently learn spatial correspondences in the input fixed and moving images. Convolutions were instead used in the decoder to recover the dense deformation field which was applied to the moving image to obtain the output image. They could show that TransMorph outperformed a variety of other registration algorithms, including both conventional and DL ones, when applied to register 3D brain MRI and 3D CT to phantom images.

1.5 Evaluation metrics

In the following section, the evaluation metrics used in the candidate's papers are defined. Even though the task changed from motion prediction to target localization, the metrics used are the same and measure the alignment between model outputs and ground truth scalars/vectors or contours. Additionally, in all papers the model inference (forward-pass) time was measured to ensure that a real-time application would be possible.

1.5.1 Root Mean Squared Error

The root mean squared error (RMSE) is a measure of the differences between predicted values and ground truth values (e.g., target centroid positions). Mathematically, it is defined as the square root of the average of the squared differences between the predicted and true values [48]:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (1.19)$$

where y_i are the true values, $\hat{y}_i$ are the predicted values, and N is the number of observations. Smaller values indicate better model performance.

1.5.2 Maximum Error

The maximum error (ME) is the largest absolute difference between the predicted values and the true values. It is defined as [48]:

$$\text{ME} = \max_i |y_i - \hat{y}_i| \quad (1.20)$$

This metric indicates the worst-case performance of a model with smaller values indicating better model performance.

1.5.3 Dice Similarity Coefficient

The dice similarity coefficient (DSC) is a measure of overlap between two contours, in this thesis between the predicted contour and the manually annotated ground truth contour. It is defined as [60]:

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B|} \quad (1.21)$$

where A and B are the two sets representing the predicted and ground truth contours, respectively. The value of DSC ranges from 0 to 1, with 1 indicating perfect overlap and 0 indicating no overlap.

1.5.4 Hausdorff Distance

The Hausdorff distance (HD) is a measure of the maximum distance between boundary points in two sets. It is used in this thesis to indicate the greatest of all the distances from a point in the predicted contour to the closest point in the ground truth contour. For two sets of points A and B , the HD is defined as [61]:

$$\text{HD} = \max \left\{ \max_{a \in A} \min_{b \in B} \|a - b\|, \max_{b \in B} \min_{a \in A} \|b - a\| \right\} \quad (1.22)$$

where $\|a - b\|$ denotes the Euclidean distance between points a and b . To account for outliers, a modified version of the HD using distance percentiles was used in this thesis. The α -percentile HD is defined as:

$$\text{HD}_\alpha = \max \left\{ P_\alpha \left(\left\{ \min_{b \in B} \|a - b\| \mid a \in A \right\} \right), P_\alpha \left(\left\{ \min_{a \in A} \|b - a\| \mid b \in B \right\} \right) \right\} \quad (1.23)$$

where P_α denotes the α -percentile of the set of distances, i.e., the distances are ranked by magnitude and a value is selected according to the desired percentile instead of taking the maximum distance. Smaller HD values indicate a better alignment of the contours.

Chapter 2

Articles

2.1 Paper I: Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy

In this study different LSTM networks for the prediction of future 1D target centroid positions in superior-inferior direction were implemented. The models were trained on respiratory motion traces extracted from 2D cine MRI acquired at the Viewray-MRIdian at the LMU University Hospital in Munich and tested on LMU and independent data obtained from a Viewray-MRIdian at the Gemelli Hospital in Rome. It was shown that the LSTM which is continuously re-optimized based on recent motion from a specific patient outperformed a population-based LSTM, a population-based linear regression and a continuously updated linear regression, which is considered the current state of the art. The investigated prediction horizons of 250 ms, 500 ms and 750 ms imply that LSTM models could be used to successfully compensate for the system latency in MRigRT.

Reprinted with permission from "Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. (2022) Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. *Physics in Medicine & Biology*, 67(9), 095006".

Physics in Medicine & Biology



PAPER

OPEN ACCESS

RECEIVED
13 December 2021

REVISED
1 March 2022

ACCEPTED FOR PUBLICATION
24 March 2022

PUBLISHED
19 April 2022

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy

Elia Lombardo¹ , Moritz Rabe¹ , Yuqing Xiong¹, Lukas Nierer¹, Davide Cusumano² , Lorenzo Placidi², Luca Boldrini², Stefanie Corradini¹, Maximilian Niyazi¹, Claus Belka^{1,3}, Marco Riboldi⁴ , Christopher Kurz¹ and Guillaume Landry¹

¹ Department of Radiation Oncology, University Hospital, LMU Munich, Munich, D 81377, Germany

² Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, I 00168, Italy

³ German Cancer Consortium (DKTK), Partner Site Munich, Munich, D 81377, Germany

⁴ Department of Medical Physics, Faculty of Physics, Ludwig Maximilians Universität München, Garching, D 85748, Germany

E-mail: guillaume.landry@med.uni-muenchen.de

Keywords: AI, long short term memory networks, time series prediction, respiratory motion, MR linac, MRI guidance

Supplementary material for this article is available [online](#)

Abstract

Objective. Gated beam delivery is the current clinical practice for respiratory motion compensation in MR-guided radiotherapy, and further research is ongoing to implement tracking. To manage intra-fractional motion using multileaf collimator tracking the total system latency needs to be accounted for in real-time. In this study, long short-term memory (LSTM) networks were optimized for the prediction of superior–inferior tumor centroid positions extracted from clinically acquired 2D cine MRIs. **Approach.** We used 88 patients treated at the University Hospital of the LMU Munich for training and validation (70 patients, 13.1 h), and for testing (18 patients, 3.0 h). Three patients treated at Fondazione Policlinico Universitario Agostino Gemelli were used as a second testing set (1.5 h). The performance of the LSTMs in terms of root mean square error (RMSE) was compared to baseline linear regression (LR) models for forecasted time spans of 250 ms, 500 ms and 750 ms. Both the LSTM and the LR were trained with offline (*offline* LSTM and *offline* LR) and online schemes (*offline+online* LSTM and *online* LR), the latter to allow for continuous adaptation to recent respiratory patterns. **Main results.** We found the *offline+online* LSTM to perform best for all investigated forecasts. Specifically, when predicting 500 ms ahead it achieved a mean RMSE of 1.20 mm and 1.00 mm, while the best performing LR model achieved a mean RMSE of 1.42 mm and 1.22 mm for the LMU and Gemelli testing set, respectively. **Significance.** This indicates that LSTM networks have potential as respiratory motion predictors and that continuous online re-optimization can enhance their performance.

1. Introduction

Magnetic resonance imaging guided radiotherapy (MR-guided RT) provides radiation-free and high soft tissue contrast imaging, allowing for inter-fractional/intra-fractional motion management and treatment adaptation (Paganelli *et al* 2018b, Kurz *et al* 2020). For tumors affected by respiratory motion such as lung, pancreatic or liver and more generally for malignancies affected by inter-fractional anatomical changes, MR-guided RT offers advantages, such as individualized planning and treatment thanks to its adaptation capabilities (Corradini *et al* 2019, Placidi *et al* 2020).

Both commercially available MR-linacs (Liney *et al* 2018), the MRIdian (ViewRay Inc., Oakwood Village, Ohio, USA) and the Unity (Elekta AB, Stockholm, Sweden) provide the possibility to monitor intra-fractional respiratory motion via 2D+*t* cine MR imaging (Green *et al* 2018, Jackson *et al* 2019, Menten *et al* 2020). For the Unity, recent studies have investigated the usage of multileaf collimator (MLC)-tracking for intra-fractional

motion management, showing its potential to increase the dose delivery accuracy (Menten *et al* 2016) and the technical feasibility of its implementation (Glitzner *et al* 2019, Uijtewaal *et al* 2021). On the other hand, automatic gated beam delivery by use of cine MRI has been clinically used on the MRIdian for years (Green *et al* 2018). To reduce treatment time, gating on the 0.35 T MR-linac is mostly performed in combination with breath-holds. However, not all patients can perform breath-holds or can perform them in a reproducible way (Persson *et al* 2019). MLC-tracking could address this limitation and achieve similar treatment accuracy as gating in a more-time efficient way (Keall *et al* 2021).

To be able to fully exploit the potential of MLC-tracking, the system latency needs to be accounted for in real-time (Poulsen *et al* 2010). The total latency for MLC-tracking is defined as the time lag between the physical target motion and the execution of the MLC motion instructions (Keall *et al* 2021). Recently, Glitzner *et al* have experimentally quantified the total latency for 4 Hz MRI-guidance with MLC-tracking to be about 350 ms for the Elekta Unity (Glitzner *et al* 2019). For the MRIdian, beam-off latencies for gating have been quantified to about 400 ms for the Cobalt-60 version (Green *et al* 2018) and to about 250 ms for the linac version with 4 Hz cine MRI (Kim *et al* 2020). To overcome RT system latencies, several motion prediction algorithms have been proposed in the past (Sharp *et al* 2004, Krauss *et al* 2011, Yun *et al* 2012); in a recent review study by Joehl *et al* a continuously re-optimized (i.e. online) linear regression (LR) model was found to perform best on average compared to other motion predictors such as artificial neural networks or Kalman filters (Joehl *et al* 2020).

In the past few years, several different artificial intelligence (AI) algorithms have found relevant applications in the field of MR-guided RT, e.g. in image segmentation, synthetic CT reconstruction or automatic online planning (Cusumano *et al* 2021). Long short-term memory (LSTM) networks (Hochreiter and Schmidhuber 1997) are a class of AI models which were designed to efficiently capture temporal dependencies in the input data and are therefore ideally suited for motion prediction. In fact, studies have shown the potential of LSTMs for motion prediction in RT based on infrared real-time position management data from a standard linear accelerator (Lin *et al* 2019) and on optical fiducial marker data from a robotic radiosurgery system (Wang *et al* 2018).

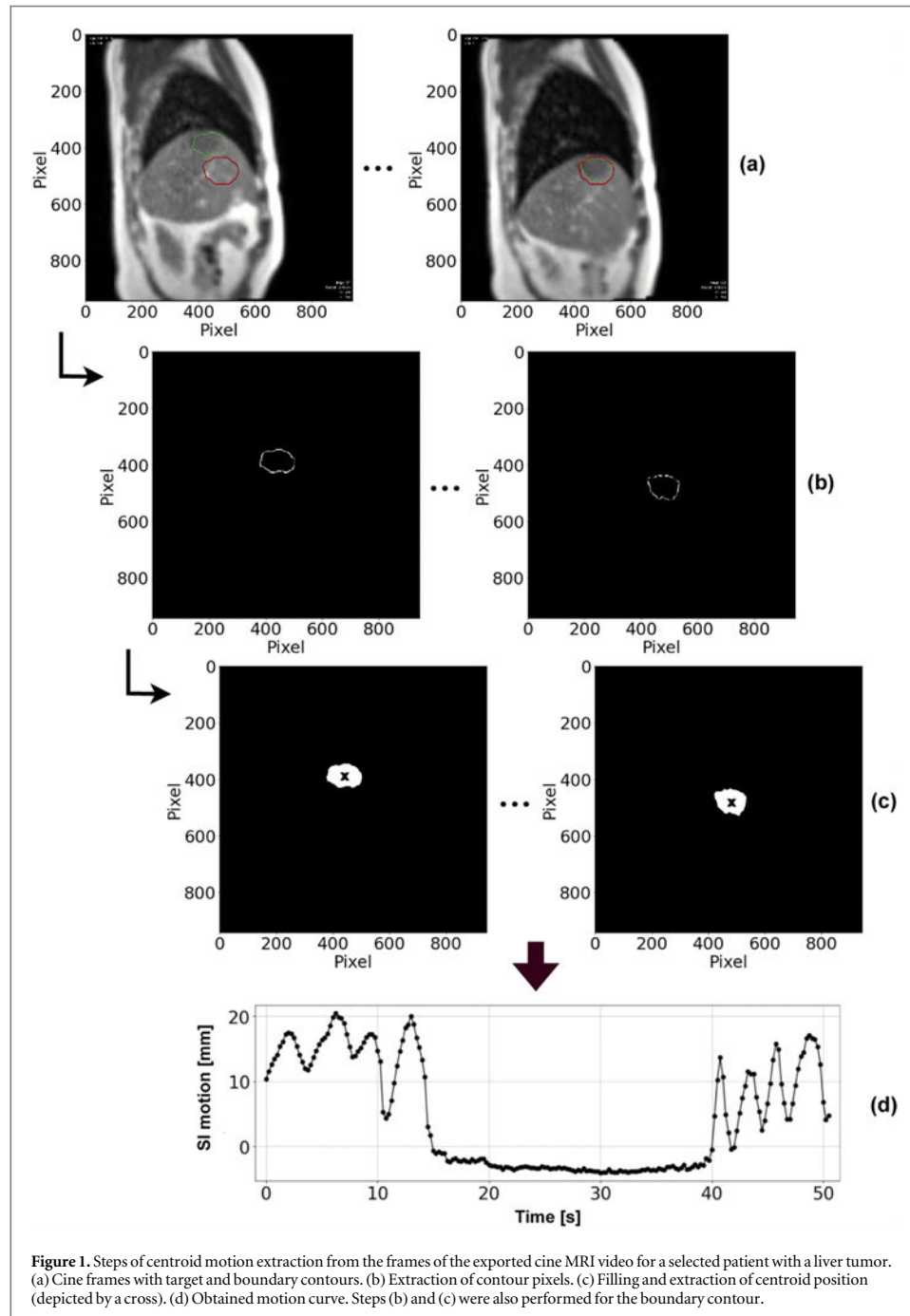
In this work, we developed LSTM networks and benchmarked their performance with LR models for the prediction of tumor centroid positions based on 4 Hz cine MRI data acquired at two different institutes with a MRIdian MR-Linac and a MRIdian MR-Cobalt-60. Specifically, motion curves from patients treated at the University Hospital of the LMU Munich were used for training, validation and testing of the models. Additionally, patients treated at the Fondazione Policlinico Universitario Agostino Gemelli in Rome were used as independent testing set. Both the LSTM and the LR were implemented with offline and online training schemes, taking into account feasibility in a 4 Hz intra-fractional motion management clinical scenario. To the best of our knowledge, this is the first study in which LSTMs were applied to MR-guided RT data and in which the usage of continuously re-optimized LSTMs was investigated for motion prediction in RT.

2. Material and methods

2.1. Respiratory motion data

We retrospectively collected respiratory motion data from 2D+*t* cine MRIs across two institutions. Specifically, cine videos from 88 patients were collected at the Department of Radiation Oncology of the University Hospital of the LMU Munich. As the RT treatment received by every patient is usually split into several fractions and for each fraction a cine MRI sequence is acquired, we obtained 556 videos from the 88 LMU patients. All patients were treated with the MRIdian MR-linac using breath-hold techniques and comprised tumors in the lung (37 cases), pancreas (22), heart (6), liver (20) and mediastinum (3). At the Fondazione Policlinico Universitario Agostino Gemelli in Rome, three patients with in total 15 cine videos were collected. For this cohort, we only selected patients treated in free-breathing using a MRIdian MR-Cobalt-60 machine. Tumor sites comprised lung (2) and pancreas (1).

For all cohorts, the 2D+*t* cine MRs were acquired at 4 Hz in a sagittal plane with a balanced steady-state free precession sequence (TRUF1; in-plane resolution $3.5 \times 3.5 \text{ mm}^2$; field-of-view $270 \times 270 \text{ mm}^2$ or $350 \times 350 \text{ mm}^2$; slice thickness of 5, 7 or 10 mm). The information on the field-of-view was used to convert the motion amplitudes from video pixels into mm. The cine MRs were exported with target and boundary contours in the OGV video format, as supported by the vendor. This resampled and interpolated video file was then used for analysis. The contours are present in every exported cine MR frame as they are used for the gated beam delivery: prior to treatment, a user defines a target structure (tumor) in a sagittal slice of the volumetric MRI, as well as a boundary structure which defines the gating area where the beam is turned on. During treatment, the target contour is continuously propagated to the current cine MR frame using fast deformable image registration by the vendor's software (Green *et al* 2018, Klueter 2019).



2.2. Data pre-processing

2.2.1. Centroid position extraction

To obtain motion trajectories from the cine MRIs we used an in-house developed software. Figure 1 summarizes the workflow of the motion extraction from the cine MR frames containing target (green) and boundary (red) contours (a). Briefly, both target and boundary contours were extracted from the videos using thresholds in RGB-space (b). The contours were then filled using the watershed algorithm (Roerdink and Meijster 2000). From the filled contours we subsequently computed the superior–inferior (SI) tumor centroid position relative

to the fixed boundary SI centroid position (c). Once the tumor SI motion curves (d) were obtained for all patients, further pre-processing was done, as detailed in the following sections.

2.2.2. Outlier replacement and filtering

First, we replaced outliers arising from incorrect filling of contours using sliding windows of size three. In detail, we computed the median centroid position within the current sliding window and if the absolute difference between the central data point in the window and the median value was a larger than an optimized threshold, we replaced that point with the median value of the window. For this step, the curves were temporarily normalized such that a single threshold independent from the absolute motion amplitudes could be used. The normalization was then reversed as the motion in mm is needed to exclude cine videos with small motion (see 2.2.4). After that, the curves were smoothed with a moving average filter applied on sliding windows of size three.

2.2.3. Breath-holds and image pauses exclusion

For the LMU cohort only, we then analyzed the motion trajectories to detect the breath-holds. Using sliding windows of size 20, we considered the current window a breath-hold if the median deviation from the median centroid position of the window was smaller than an optimized threshold. This information was then used to exclude all breath-hold data points from the motion trajectories and keep only the free-breathing sub-trajectories in between. Additionally, for this cohort we separated the data according to detected imaging pauses. Cine imaging pauses are inherently part of the MRIdian MR-linac treatment: when the gantry rotates from one irradiation angle to the next, its moving electronics interferes with the MRI causing the image quality to degrade. These degraded cine MRI frames are automatically excluded from the exported videos by the vendor, but their start is indicated by displaying the statement ‘imaging paused’ on the top right of the video. We automatically detected the frames where this statement was displayed and used this information to separate the motion trajectories into two sub-trajectories. This avoids jumps in the curves arising from an imaging pause between data points.

2.2.4. Small motion exclusion and data normalization

For both the LMU and the Gemelli cohorts we excluded all data of cine videos for which the interquartile range (IQR) of SI free-breathing motion was below 3.5 mm (in-plane resolution of MRIdian cine MRIs), as this motion is more substantially affected by imaging noise. This led to the exclusion of 73 LMU and 5 Gemelli videos. Finally, we again normalized all motion curves to the range -1 to $+1$ using the minimum and maximum tumor centroid position of each cine MRI. These min/max values were saved to disk and used during evaluation to undo the normalization of the predicted curves.

After pre-processing, we obtained 16.1 h of motion data without breath-holds (105.8 h if the breath-holds were not excluded) for the LMU cohort and 1.5 h of free-breathing motion for the Gemelli cohort.

2.3. Motion prediction models

2.3.1. Mathematical formulation of prediction problem

Following the terminology used by Remy *et al* (2021), motion prediction is about obtaining future target positions (at time $t + \Delta t$) from the current target motion (at time t). In general, for a given time step i , every prediction task can be simply formulated as

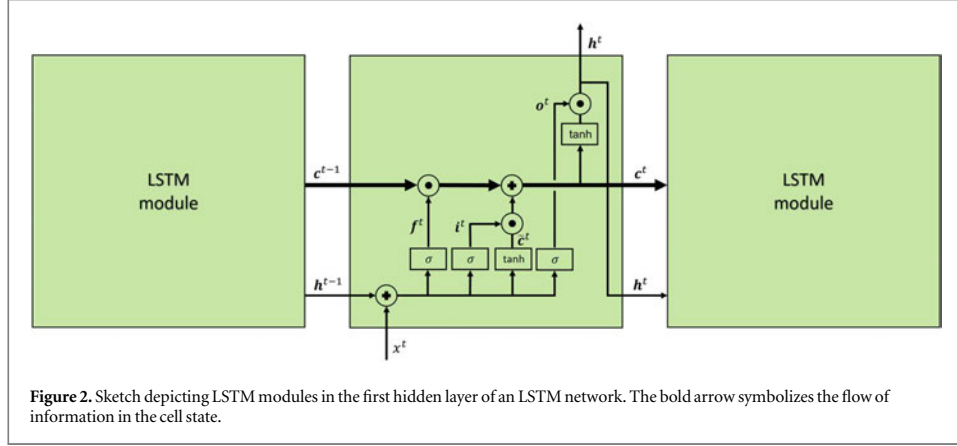
$$\hat{y}_i = f(\mathbf{x}_i), \quad (1)$$

where $f()$ is a motion prediction algorithm, $\mathbf{x}_i$ is the i th vector containing the input data window and $\hat{y}_i$ is the i th vector with the predicted output data window. The corresponding vector $\mathbf{y}_i$ contains the ground truth output data window used to optimize the algorithm. In our case, $\mathbf{x}$ and $\mathbf{y}$ contain input and output SI target centroid positions. The length of $\mathbf{x}$ was treated as a hyper-parameter (see section 2.4.2). On the other hand, the length of $\hat{\mathbf{y}}$ (and $\mathbf{y}$) is automatically related to the forecasted time span. In this study, we investigated forecasts of 250 ms, 500 ms and 750 ms, corresponding to $\hat{\mathbf{y}}$ having length of 1, 2 or 3, respectively, for 4 Hz imaging.

2.3.2. Linear ridge regression

Over the last decade, several motion prediction algorithms have been proposed to account for latencies in image guided RT. Joehl *et al* used motion traces from a robotic radiosurgery system to compare 18 different predictors for 160 ms and 480 ms forecasts (Joehl *et al* 2020). On average, they found an (online) LR model to perform best, so we decided to leverage it as baseline model. Mathematically, the regression function is defined as (Krauss *et al* 2011)

$$f(\mathbf{x}) = \beta^T \mathbf{x} + \beta_0, \quad (2)$$



where the vector β contains the parameters of the regression model. The loss function to be minimized to solve the regression is given by

$$L(\beta) = \sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2 + \lambda \|\beta\|^2, \quad (3)$$

where N is the number of input/output training windows and λ is an L2-regularization parameter. If $\lambda \neq 0$, the term ridge regression is usually used.

If we define a matrix $\mathbf{X}$ such that its rows equal the input window vectors $\mathbf{x}_i$ and $\mathbf{Y}$ a matrix with the true output window vectors $\mathbf{y}_i$, the loss function $L(\beta)$ is analytically solved by the optimal parameters

$$\beta^* = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (4)$$

where $\mathbf{I}$ is the identity matrix. Therefore, the LR model has a closed form solution and does not need iterative optimization.

2.3.3. LSTM

Recurrent neural networks (RNN) are a class of machine learning algorithms ideally suited for sequential input data (e.g. time series data). Compared to artificial neural networks or convolutional neural networks, recurrent neural networks have an extra set of weights which connects hidden layers from one time point to the next. However, the original RNN module comprising a simple fully connected layer was found to be limited due to unstable gradient issues when back-propagating over longer input sequences (Shen *et al* 2020). Therefore, more advanced RNN architectures have been proposed and the most widely adopted is the LSTM model (Hochreiter and Schmidhuber 1997), which was specifically designed to more easily learn longer sequences of data.

The repeating module of LSTMs is shown in figure 2. LSTMs introduce the memory cell state $\mathbf{c}^t$, which allows a stable back-propagation of errors and straightforward flow of information. To remove or add information to the cell state, structures called gates are used. Intuitively, the forget gate $\mathbf{f}^t$ is used to keep/discard past information in $\mathbf{c}^{t-1}$ while the input gate $\mathbf{i}^t$ allows to filter information in the new memory cell state $\tilde{\mathbf{c}}^t$. The previous gates and memory are then combined to build the final memory cell state $\mathbf{c}^t$. The final memory cell is filtered with the output gate $\mathbf{o}^t$ to keep/discard some information and build the hidden state $\mathbf{h}^t$. Mathematically, at a specific time step t , the LSTM module is described as follows:

$$\text{Forget gate: } \mathbf{f}^t = \sigma(\mathbf{W}_f \mathbf{x}^t + \mathbf{U}_f \mathbf{h}^{t-1} + \mathbf{b}_f) \quad (5)$$

$$\text{Input gate: } \mathbf{i}^t = \sigma(\mathbf{W}_i \mathbf{x}^t + \mathbf{U}_i \mathbf{h}^{t-1} + \mathbf{b}_i) \quad (6)$$

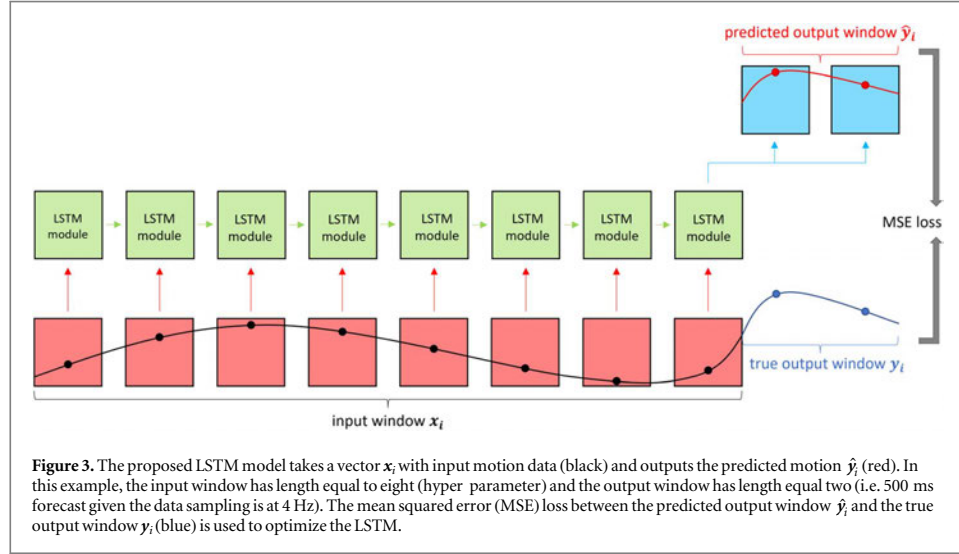
$$\text{New memory cell state: } \tilde{\mathbf{c}}^t = \tanh(\mathbf{W}_c \mathbf{x}^t + \mathbf{U}_c \mathbf{h}^{t-1} + \mathbf{b}_c) \quad (7)$$

$$\text{Final memory cell state: } \mathbf{c}^t = \mathbf{f}^t \odot \mathbf{c}^{t-1} + \mathbf{i}^t \odot \tilde{\mathbf{c}}^t \quad (8)$$

$$\text{Output gate: } \mathbf{o}^t = \sigma(\mathbf{W}_o \mathbf{x}^t + \mathbf{U}_o \mathbf{h}^{t-1} + \mathbf{b}_o) \quad (9)$$

$$\text{Hidden state: } \mathbf{h}^t = \mathbf{o}^t \odot \tanh(\mathbf{c}^t), \quad (10)$$

where $\mathbf{b}$, $\mathbf{W}$ and $\mathbf{U}$ denote the biases, input window weights and recurrent weights which are learned during the optimization process. The symbol $\odot$ represents element-wise multiplication between matrices/vectors. The sigmoid function $\sigma(x)$ was used for the gates and the hyperbolic tangent function was used to generate the states. For each time step, the hidden state of one LSTM layer is used as input for the next LSTM layer. For the last



hidden layer of the LSTM, the hidden state of the last time point t_f is input to a fully connected layer to build the predicted output window

$$\text{Predicted output: } \hat{y}_i = W_{FC} h^{t_f} + b_{FC}, \quad (11)$$

where W_{FC} and b_{FC} denote the weight matrix and bias vector for the fully connected layer.

In this study, a stateless LSTM was implemented, which means that during optimization the hidden state and the cell state were cleared after every batch of data. The LSTM architecture was inspired by the one used by Lin *et al* (2019). Specifically, we performed our hyper-parameter optimization based on the range of values used in their hyper-parameter search. More details can be found in section 2.4. Figure 3 schematically shows the working principle of the proposed LSTM. At every time point, a single SI tumor centroid position is given as input for as many points as the length of the input data window. The LSTM modules in the hidden layer (green boxes) process the time-dependent information as shown in figure 2 until the last LSTM module is reached. The hidden vector which is output by the last LSTM module is mapped via a fully connected layer to the predicted output window following equation (11). Note that in figure 3 only one hidden layer is shown whereas this number was treated as a hyper-parameter in our optimizations (see section 2.4.2).

2.4. Model optimization

2.4.1. Data subdivision

To optimize and evaluate the models we split the LMU data into training, validation and testing sets. Specifically, we assigned the motion trajectories belonging to 60% of the patients to the training set (52 patients), 20% to the validation set (18 patients) and the remaining 20% to the testing set (18 patients) and did this procedure only once at the beginning. This splitting roughly also led to 60% of the motion trajectories being in training (9.1 h), 20% in validation (4.0 h) and 20% in testing (3.0 h). As the Gemelli cohort was smaller (1.5 h) but at the same time in free-breathing, we decided to use this dataset as an independent additional testing set. Finally, we also applied the best models trained/validated on the LMU data without breath-holds to the LMU testing set without excluding the breath-holds during pre-processing.

2.4.2. Hyper-parameter search

To find the optimal set of hyper-parameters for both the LSTM and the LR we repeatedly performed training and validation while varying the parameters for all three analyzed forecasts and for all four training strategies (see section 2.4.3) separately. For the LSTM, the following hyper-parameters were varied, based on the hyper-parameter search performed by Lin *et al*:

- *Number of layers*: the number of hidden layers of the LSTM was chosen among the following values {1, 3, 5, 10}.
- *Dropout*: the dropout rate on the outputs of each hidden layer (but the last one) was sampled from the set {0, 0.1, 0.2}.

- *Learning rate*: the learning rate of the optimizer was sampled from the set $\{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}, 1 \times 10^{-2}\}$.
- *Batch size*: the batch size, i.e. the number of input windows fed to the network simultaneously, can be varied only for the *offline* trained LSTM (see section 2.4.3) and was set to either 64 or 128.
- *L2-regularization*: the L2-regularization parameter λ (also called weight decay) was sampled from the set $\{0, 1 \times 10^{-6}, 1 \times 10^{-5}, 1 \times 10^{-4}\}$.

All optimizations for the LSTMs were carried out using the Adam optimizer (Kingma and Ba 2015) with a normalized mean squared error (MSE) loss function and learning rates from the set shown above. We set the number of features in the hidden layer vector $\mathbf{h}^t$ to 15 like Lin *et al*. No batch normalization was used. For the LR, the following hyper-parameter was varied in logarithmic steps over a large range of values:

- *L2-regularization*: the L2-regularization parameter λ was sampled from the set $\{1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}, 1, 10\}$.

Both for the LSTM and for the LR we varied the length of the input data window $\mathbf{x}$ between 8, 16, 24 and 32 data points, corresponding to 2, 4, 6 and 8 seconds of past motion.

2.4.3. Training strategies

Retraining models on recent motion data has been shown to improve the predictive performance (Krauss *et al* 2011, Sun *et al* 2020). Thus, two different training strategies were investigated for the LSTM and for the LR model.

- *Offline LSTM*: the *offline* LSTM optimization was carried out following the typical machine learning training/validation/testing subdivision. The model was iteratively optimized on the training set for 600 epochs while monitoring training and validation losses. If the validation loss improved, the model weights were saved to disk. If the validation loss did not improve for 100 epochs, the optimization was stopped, a technique known as early stopping. For final inference, we loaded the weights and hyper-parameters of the best performing model on the validation set and applied it unchanged to the testing set.
- *Offline+online LSTM*: to allow adaptation to recent motion patterns, we continuously retrained the LSTM on current data. Specifically, we first loaded the weights of a previously optimized *offline* LSTM. The LSTM was then re-optimized on the last 20 s of validation data using a sliding set of validation input/output windows updated with a first-in-first-out approach, as shown in figure 4. The online optimization of the LSTM was done for 10 epochs, taking about 150 ms. This would allow an implementation in a 4 Hz image acquisition clinical scenario. To prevent the iterative optimization to introduce an additional latency, within the 250 ms between one cine MRI frame and the next, we performed the prediction *before* optimizing the LSTM. To calculate the validation loss we used the ground truth data point lying 250 ms, 500 ms or 750 ms (depending on the forecast) in the future with respect to the last centroid position in the currently used 20 s of optimization data. For final inference, we loaded the *offline* LSTM, set the hyper-parameters leading to the best result on the validation set and continuously retrained and evaluated the model on the testing set.
- *Offline LR*: the *offline* LR training is analogous to the *offline* LSTM training but for the fact the LR is solved analytically while the LSTM is iteratively optimized. Specifically, the LR was solved on the training set and then applied unchanged to validation set to perform the hyper-parameter search. For final inference, as for the *offline* LSTM, we loaded the weights and set the hyper-parameters of the best performing model on the validation set and applied it unchanged to the testing set.
- *Online LR*: on the other hand, the *online* LR is different from *offline+online* LSTM. As no iterative fine-tuning is needed for the LR, no weights from a pre-trained *offline* LR were loaded. The *online* LR was continuously solved ‘from scratch’ based on the last 20 s of validation data using a sliding set of validation input/output windows updated with a first-in-first-out approach (figure 4). As solving the LR is simply a matrix multiplication (see equation (2)), it takes less than 1 ms. As this additional latency is not significant, for the *online* LR we performed the prediction *after* the optimization, as illustrated in figure 4. This is advantageous as the model’s prediction can take into account the most recently acquired data point.

As mentioned in section 2.3.1, our data was subdivided in input windows $\mathbf{x}$ where the number of entries $\text{len}(\mathbf{x})$ is a hyper-parameter. To obtain a set of windows with a total duration of 20 s to be used for online training, we need several input windows. Given that every input window is shifted by one and that the cine imaging is

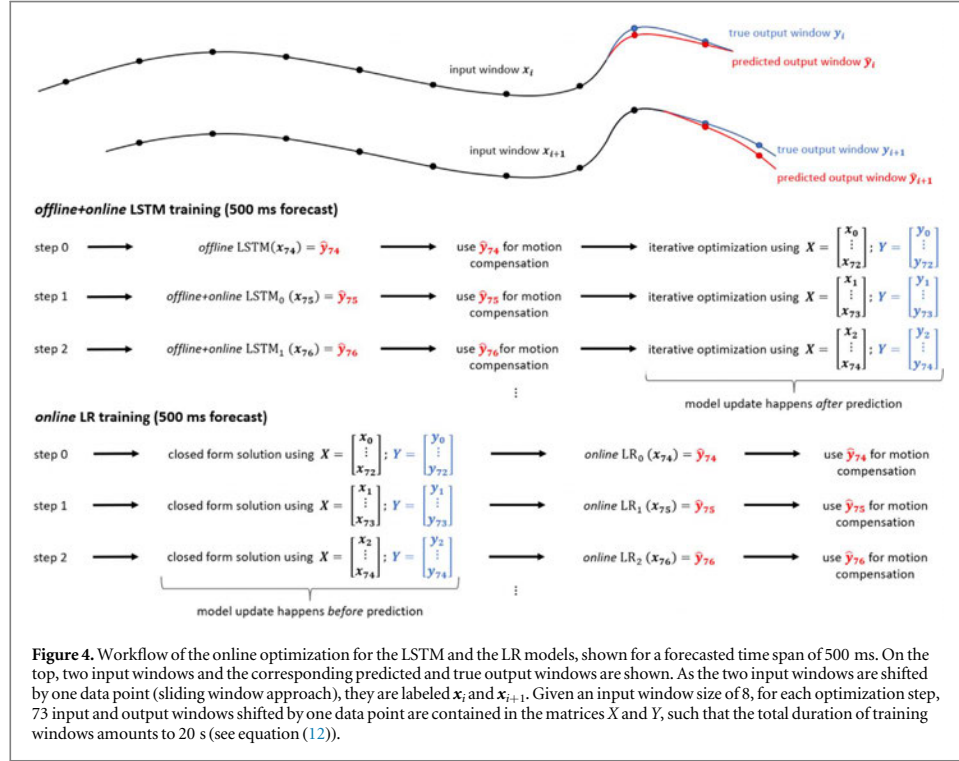


Figure 4. Workflow of the online optimization for the LSTM and the LR models, shown for a forecasted time span of 500 ms. On the top, two input windows and the corresponding predicted and true output windows are shown. As the two input windows are shifted by one data point (sliding window approach), they are labeled x_i and x_{i+1} . Given an input window size of 8, for each optimization step, 73 input and output windows shifted by one data point are contained in the matrices X and Y , such that the total duration of training windows amounts to 20 s (see equation (12)).

performed at a frame rate of 4 Hz, the number of input windows needed is given by

$$\text{Nr. input windows} = 20 \cdot 4 - \text{len}(x) - 1. \quad (12)$$

Therefore, 73 input windows are needed if we choose $\text{len}(x) = 8$ as in figure 4. This number corresponds to the batch size for the *offline+online* LSTM, which is thus not freely selectable if we fix the duration of the online optimization data to 20 s. The number of input windows between the last window used for optimization and the window used for prediction is given by the current forecasted time span. For example for the 500 ms forecast scenario shown in figure 4, this difference is equal to two (e.g. step 0: x_{72} is the last window in the optimization matrix and x_{74} is the window used for prediction). For the 250 ms forecast this difference is one and for the 750 ms forecast this difference is equal to three.

2.5. Loss and evaluation metrics

The loss function used to optimize the LSTM was the MSE, which is defined as

$$\text{MSE} = \frac{1}{B} \sum_{i=1}^B (y_i - \hat{y}_i)^2, \quad (13)$$

where B is the batch size, y_i is the vector with the true output window and $\hat{y}_i$ is the vector with the predicted output window of centroid positions. Note that the MSE was computed using normalized output windows.

The root mean squared error (RMSE) and maximum error (ME) error were used to evaluate the LSTM and LR predictive performance on the validation and testing sets. Prior to the computation of the evaluation metrics, the normalization of the ground truth and predicted curves was reversed, such that the metrics are in mm. The RMSE and ME were calculated on a treatment fraction basis (one RMSE/ME per cine MRI video) and are defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (14)$$

$$\text{ME} = \max\{|y_i - \hat{y}_i|, i = 1, 2, \dots, N\}, \quad (15)$$

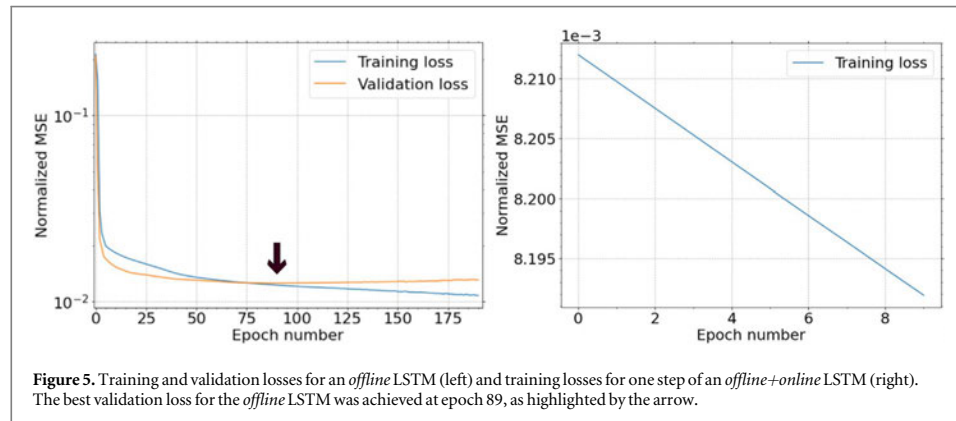


Figure 5. Training and validation losses for an *offline* LSTM (left) and training losses for one step of an *offline+online* LSTM (right). The best validation loss for the *offline* LSTM was achieved at epoch 89, as highlighted by the arrow.

where N is the number of data points in the motion curves belonging to a single cine MRI, y_i is the true future centroid position and $\hat{y}_i$ is the predicted centroid position. The analysis was done using only the last element of each output window, which is the hardest to predict (Sharp et al 2004).

Finally, we averaged over the RMSEs/MEs of different fractions to build the mean RMSE/ME with corresponding standard deviation.

2.6. Statistical tests

To analyze if there is a statistically significant difference between the RMSE values obtained with the different models on the different testing sets, non-parametric Friedman tests were performed (Friedman 1937). A p -value < 0.05 was considered significant. If the Friedman test revealed a significant difference, we consecutively performed a post-hoc Nemenyi test (Nemenyi 1963) to infer which model obtained significantly better RMSEs in a pair-wise fashion.

2.7. Implementation details

All code used for this study was written in Python 3.8.5 and is freely available: https://github.com/LMUK-RADONC-PHYS-RES/lstm_centroid_prediction. To build and optimize the LSTMs, the PyTorch library (Paszke et al 2017) version 1.8.0 was used. Training for both the *offline* and the *offline+online* LSTM was carried out on an NVIDIA Quadro RTX 8000 GPU with 48 GB of memory. The LR was built and solved using the scikit-learn library (Pedregosa et al 2011) version 0.24.1. The LR was trained on an Intel Xeon Gold 6254 (Cascade Lake-EP) 18-Core CPU.

3. Results

In terms of prediction speed, a forward pass with an LSTM takes about 5 ms while for the LR models less than 1 ms.

3.1. Validation

Figure 5 shows the normalized MSE losses for an optimization of an *offline* LSTM and an *offline+online* LSTM. For the shown *offline* LSTM, the best validation loss was obtained at epoch 89 which led to early stopping of the optimization at epoch 189. On the other hand, no validation loss was monitored for the *offline+online* LSTM. As shown in figure 4, within one training step we first performed the prediction and then iteratively re-optimized the LSTM for 10 epochs (see section 2.4.3), as this is the maximum number of epochs which would still allow a re-optimization in a 4 Hz clinical scenario.

Table 1 shows the best RMSEs obtained with the four different models on the validation set. The corresponding set of best hyper-parameters for each model is shown in the appendix (tables A1, A2 and A3). For all three forecasted time spans, the *offline+online* LSTM achieved the best performance, reaching basically the same as the *offline* LSTM for the 250 ms forecast, and slightly better performance for the 500 ms and 750 ms forecasts. The best performing LR was the *offline* one, however its performance was worse than both LSTM training schemes.

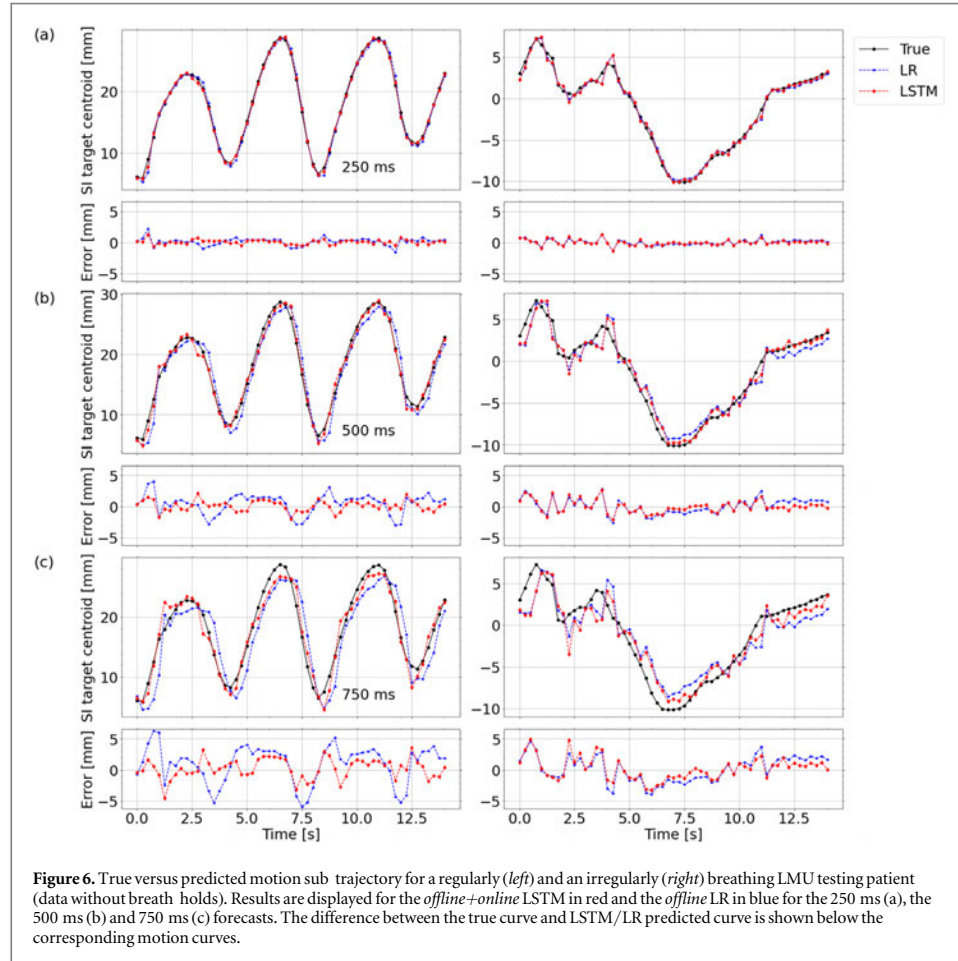


Figure 6. True versus predicted motion sub trajectory for a regularly (left) and an irregularly (right) breathing LMU testing patient (data without breath holds). Results are displayed for the *offline+online* LSTM in red and the *offline* LR in blue for the 250 ms (a), the 500 ms (b) and 750 ms (c) forecasts. The difference between the true curve and LSTM/LR predicted curve is shown below the corresponding motion curves.

Table 1. Mean and standard deviation of RMSEs for the validation set. The RMSE of the best performing model is shown in bold for each forecasted time span.

Model	250 ms forecast	500 ms forecast RMSE [mm]	750 ms forecast
Offline LSTM	0.55 ± 0.44	1.40 ± 1.00	2.58 ± 1.71
Offline+online LSTM	0.54 ± 0.43	1.36 ± 0.94	2.54 ± 1.63
Offline LR	0.63 ± 0.49	1.68 ± 1.11	3.09 ± 1.91
Online LR	0.74 ± 0.53	1.76 ± 1.13	3.15 ± 1.87

3.2. Testing

Figure 6 shows for two selected patients of the LMU testing set (data without breath-holds) ground truth versus predicted respiratory motion trajectories for the best LSTM model (*offline+online*) and the best LR model (*offline* LR) for all the forecasted time spans. Qualitatively, no noticeable difference is seen when comparing the best LR and the best LSTM for the 250 ms forecast. On the other hand, both for the 500 ms forecast and for the 750 ms one can see how the LSTM outperforms the LR especially when it comes to predicting steep inhalations/exhalations. Similar observations can be made when looking at figure 7 displaying true vs predicted curves for Gemelli testing patients. Although we noticed that the LSTM overshoots more often than the LR, the former is able to more quickly adapt to changes in the motion trajectories (from steeper/shallower inhalations/exhalations to irregularities) which leads to an overall smaller error, as can be seen in the error plots in figures 6 and 7. Table 2 shows the RMSEs obtained with the four best validation models on the LMU testing set (data without breath-holds). The *offline+online* LSTM was confirmed the best model for all three forecasts. These

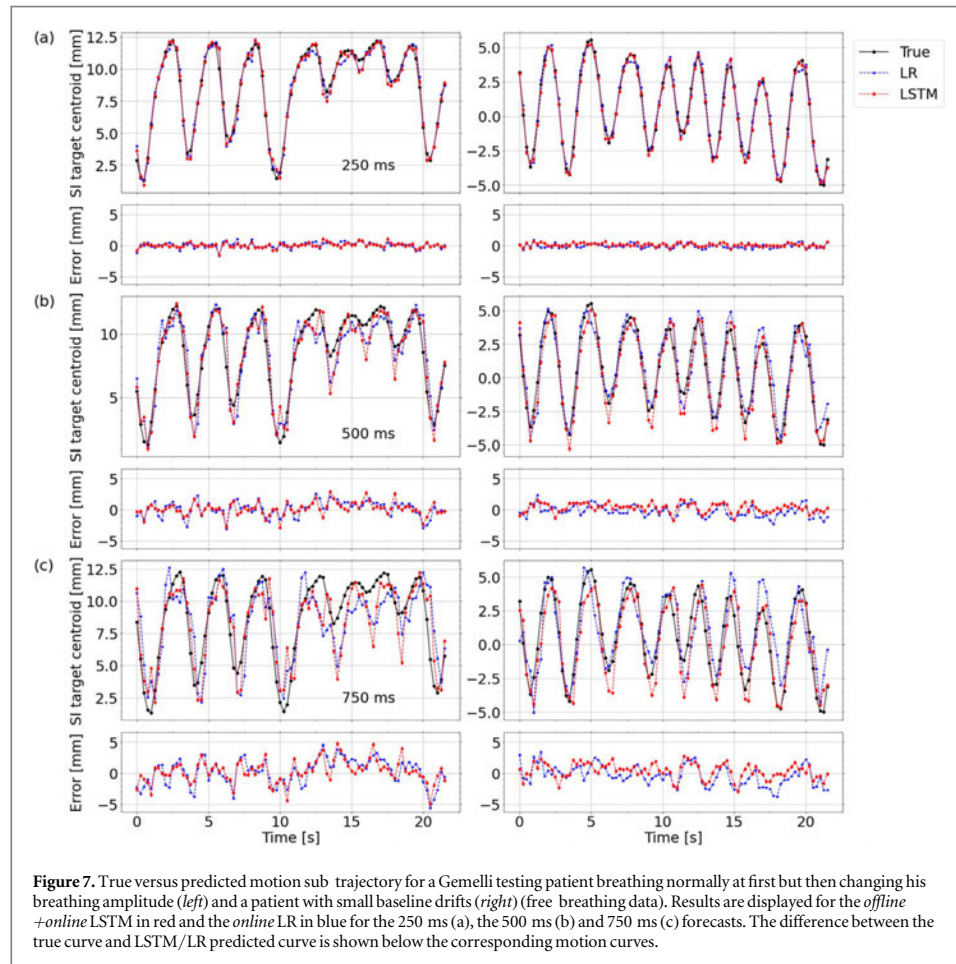


Figure 7. True versus predicted motion sub trajectory for a Gemelli testing patient breathing normally at first but then changing his breathing amplitude (*left*) and a patient with small baseline drifts (*right*) (free breathing data). Results are displayed for the *offline+online* LSTM in red and the *online* LR in blue for the 250 ms (a), the 500 ms (b) and 750 ms (c) forecasts. The difference between the true curve and LSTM/LR predicted curve is shown below the corresponding motion curves.

Table 2. Mean and standard deviation of RMSEs for the LMU testing set without breath holds. The RMSE of the best performing model is shown in bold for each forecasted time span.

Model	250 ms forecast	500 ms forecast	750 ms forecast
	RMSE [mm]		
Offline LSTM	0.49 ± 0.29	1.24 ± 0.70	2.34 ± 1.25
Offline+online LSTM	0.48 ± 0.28	1.20 ± 0.65	2.20 ± 1.12
Offline LR	0.54 ± 0.30	1.42 ± 0.78	2.61 ± 1.38
Online LR	0.64 ± 0.38	1.54 ± 0.79	2.73 ± 1.41

results were confirmed also when looking at MEs, as shown in table A4 in the appendix. In general, all models performed slightly better than during validation both in terms of mean and standard deviation of the RMSE.

When applying the Friedman test, we found a significant difference among the models for all forecasts and testing sets. For the LMU testing set without breath-holds, the post-hoc Nemenyi test yielded the *p*-values shown in table 3. The best model in terms of RMSEs, i.e. the *offline+online* LSTM was found to perform significantly better than both LR models while there was no significant difference between the *offline* LSTM and the *offline+online* LSTM for all investigated forecasts.

As shown in table 4, the results obtained on the LMU validation and testing set (data without-breath-holds) were confirmed with the Gemelli testing set (free-breathing data). The *offline+online* LSTM was found to perform best for all three forecasted time spans followed by the *offline* LSTM. This time, the *online* LR performed better than the *offline* LR and reached the same RMSE as the *offline* LSTM for the 750 ms forecast. Table 5 shows

2.1. Paper I: Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy

31

Table 3. *P* values obtained from the post hoc Nemenyi test for the LMU testing set without breath holds for all possible pairwise model comparisons. Significant *p* values ($< 5e-2$) are denoted with an asterisk.

Comparison		250 ms forecast	500 ms forecast	750 ms forecast
Model 1	Model 2	<i>p</i> value		
Offline LSTM	Offline LR	3e-1	3e-2*	2e-1
Offline LSTM	Online LR	6e-1	1e-1	3e-1
Offline+online LSTM	Offline LR	1e-3*	1e-3*	1e-3*
Offline+online LSTM	Online LR	2e-3*	1e-3*	1e-3*
Offline LSTM	Offline+online LSTM	9e-2	6e-2	6e-2
Offline LR	Online LR	9e-1	9e-1	9e-1

Table 4. Mean and standard deviation of RMSEs for the Gemelli free breathing testing set. The RMSE of the best performing model is shown in bold for each forecasted time span.

Model	250 ms forecast	500 ms forecast	750 ms forecast
	RMSE [mm]		
Offline LSTM	0.47 ± 0.12	1.14 ± 0.29	2.02 ± 0.49
Offline+online LSTM	0.42 ± 0.13	1.00 ± 0.30	1.77 ± 0.54
Offline LR	0.57 ± 0.14	1.52 ± 0.34	2.76 ± 0.71
Online LR	0.53 ± 0.17	1.22 ± 0.30	2.02 ± 0.49

Table 5. *P* values obtained from the post hoc Nemenyi test for the Gemelli free breathing testing set for all possible pairwise model comparisons. Significant *p* values ($< 5e-2$) are denoted with an asterisk.

Comparison		250 ms forecast	500 ms forecast	750 ms forecast
Model 1	Model 2	<i>p</i> value		
Offline LSTM	Offline LR	3e-2*	3e-2*	1e-2
Offline LSTM	Online LR	2e-1	8e-1	9e-1
Offline+online LSTM	Offline LR	1e-3*	1e-3*	1e-3*
Offline+online LSTM	Online LR	3e-3*	2e-2*	8e-2
Offline LSTM	Offline+online LSTM	4e-1	2e-1	8e-2
Offline LR	Online LR	4e-1	6e-2	1e-2*

Table 6. Mean and standard deviation of RMSEs for the LMU testing set with breath holds. The RMSE of the best performing model is shown in bold for each forecasted time span.

Model	250 ms forecast	500 ms forecast	750 ms forecast
	RMSE [mm]		
Offline LSTM	0.34 ± 0.17	0.83 ± 0.45	1.59 ± 0.95
Offline+online LSTM	0.30 ± 0.17	0.74 ± 0.39	1.34 ± 0.74
Offline LR	0.36 ± 0.19	0.96 ± 0.51	1.83 ± 1.03
Online LR	0.63 ± 0.65	1.39 ± 0.93	2.81 ± 2.12

the *p*-values obtained with the Nemenyi test on the Gemelli testing results. This time, the *offline+online* LSTM was found to be significantly better than the *offline* LR for all forecasts and significantly better than the *online* LR for the 250 ms and the 500 ms forecast.

Finally, we also applied the best models obtained on the LMU data without breath-holds to the LMU testing set with breath-holds. As shown in table 6, the *offline+online* LSTM again outperformed all other models for all three forecasts followed by the *offline* LSTM and the *offline* LR. While these three models substantially improved their performance compared to the LMU testing set without breath-holds (table 2), the *online* LR even worsened its performance for the 500 ms and 750 ms forecasts. When applying the Nemenyi test, we found significant differences between all pairwise model combinations excluding the *offline* LSTM versus the *offline* LR, as shown in table 7.

Table 7. *P* values obtained from the post hoc Nemenyi–Friedman test for the LMU testing set with breath-holds for all possible pairwise model comparisons. Significant *p* values ($< 5e-2$) are denoted with an asterisk.

Comparison Model	Model 2	250 ms forecast	500 ms forecast <i>p</i> value	750 ms forecast
Offline LSTM	Offline LR	3e-1	2e-3*	9e-3*
Offline LSTM	Online LR	1e-3*	1e-3*	1e-3*
Offline+online LSTM	Offline LR	1e-3*	1e-3*	1e-3*
Offline+online LSTM	Online LR	1e-3*	1e-3*	1e-3*
Offline LSTM	Offline+Online LSTM	1e-3*	1e-3*	1e-3*
Offline LR	Online LR	1e-3*	1e-3*	1e-3*

Animated figures with sliding input, true output and predicted output windows for a testing patient from the LMU set without breath-holds, the Gemelli set and the LMU set with breath-holds respectively are shown in the online material (stacks.iop.org/PMB/67/095006/mmedia).

4. Discussion

LSTM networks have been successfully applied for time series prediction in many fields, making them one of the most popular versions of RNNs (Shen *et al* 2020). In this study, we applied LSTMs to forecast tumor centroid positions based on respiratory motion trajectories obtained from 0.35 T MR-linacs. The fact that the proposed *offline+online* LSTM outperformed all the other models for all testing cohorts and for all forecasted time spans confirms our hypothesis that LSTMs are well suited for motion prediction in MR-guided RT. The *offline+online* LSTM was found to perform significantly better than the best performing LR in 8/9 testing scenarios. The only scenario where the better RMSE of the *offline+online* LSTM was not significant compared to the best LR was for the 750 ms forecast with the Gemelli testing data. However, the Gemelli testing set presents less data compared to the other two testing sets, as there are less videos over which the RMSE is calculated.

As expected from literature (Murphy and Pokhrel 2009, Sun *et al* 2020), the *offline+online* LSTM achieved better performance than the *offline* LSTM for all testing cohorts when looking at RMSE. Additionally, for the LMU testing set with breath-holds this difference was statistically significant. As this testing cohort differs more substantially from the training and validation sets than the other two, we expected this improvement. In general, we conclude that iterative fine-tuning using the latest respiratory patterns is beneficial also for LSTMs. The *offline+online* LSTM was implemented such that online optimization took about 150 ms and could therefore be used in a 4 Hz cine MRI guided RT treatment.

All models achieved better mean RMSE on the LMU testing set with breath-holds included. This is expected since this data contains long time intervals of flat motion trajectories, which are easy to predict.

For the LMU testing set without breath-holds, the *offline* LR regression was found to perform better than the *online* LR (see table 2), a finding in disagreement with literature (Krauss *et al* 2011, Uijtewaal *et al* 2021). However, the difference was not significant, as shown in table 3. When comparing the mean RMSE obtained for the 500 ms forecast with our *offline* LR to the mean RMSE obtained with the *online* LR by Uijtewaal *et al* for the 500 ms forecast, we can see that both models achieved a value of about 1.5 mm. Furthermore, the *online* LR was found to perform better than the *offline* on the Gemelli testing set, which likely differs from the LMU training set. Additionally, the free-breathing Gemelli data might be easier to predict than the LMU data without breath-holds as the latter consists of sub-trajectories of free-breathing motion in-between breath-holds and can thus contain irregular breathing or steep inhalations and exhalations. This could also explain why the *online* LR performed better on the Gemelli testing set.

To compare the performance obtained in this study with the one obtained by the LSTM implemented by Lin *et al*, we report here normalized RMSEs obtained with our *offline+online* LSTM for the 500 ms forecast. The normalization was done using the min-max amplitudes saved to disk during pre-processing. We found a mean normalized RMSE of 0.086 for the LMU testing set without breath-holds and 0.107 for the Gemelli testing set. These results are in agreement with the mean testing RMSE of 0.139 found by Lin *et al* (2019). Furthermore, we can approximately compare the RMSE obtained with our *offline+online* LSTM for the 500ms forecast with the RMSE obtained by Wang *et al* using a Bi-LSTM for a 400 ms forecast (Wang *et al* 2018). Since they found a mean validation normalized RMSE of 0.081 (no testing set was used, unlike in this study), we conclude that our *offline+online* LSTM is comparable.

In general, we noticed large standard deviations of the RMSEs. This suggests that substantial performance differences might be observed among different patients. The standard deviations for the Gemelli testing set and the LMU testing set with breath-holds were smaller than for the LMU testing set without breath-holds. As the Gemelli data consists of regular free-breathing and the LMU data with breath-holds largely consists of flat

motion regions, we hypothesize that this decreased variability in the data leads to smaller standard deviations. The fact that the mean and standard deviation of the validation RMSE were larger than the LMU testing results (compare tables 1 and 2), might be explained by the fact that by chance, when splitting the LMU cohort in training, validation and testing, more irregular motion curves were assigned to the validation compared to the testing set.

As observed in several studies (Sharp *et al* 2004, Seregini *et al* 2016, Wang *et al* 2018, Uijtewaal *et al* 2021), the predictive performance decreased with increasing forecasted time span. However, sub-resolution accuracy (<3.5 mm) was still reached for all three forecasts. The good RMSE of about 1 mm of the *offline+online* LSTM for the 500 ms forecast shows that this model could be used to successfully account for the system latencies found by Glitzner *et al* (2019) when performing MLC tracking on an Elekta Unity MR-linac.

The current study has a few limitations. The first is that all models were optimized and applied on motion curves which were normalized based on the global minimum and maximum SI centroid position of each cine video, following (Lin *et al* 2019, Yu *et al* 2020). In clinical practice of course, the global minimum and maximum for the entire fraction cannot be known before the treatment ends. However, with the 0.35 T MR-linac, right before the treatment starts a preview cine MRI is acquired (Klueter 2019) for automatic selection of a tracking key frame, to inspect if the gating window needs to be adjusted and similar aspects. This cine MRI could also be used to get the min-max amplitudes to be used for the normalization of the motion curves acquired for the treatment. A small window size (equal to three) was taken for both the outlier replacement and the moving average filter to make an implementation in a real-time clinical scenario possible. The second limitation consists in the fact that our models only predict the future centroid position in SI direction. While this could already be used for centroid MLC tracking in parallel direction, where the MLC shape is shifted to the predicted SI position, latencies for deviations in anterior-posterior direction would not be accounted for. To achieve this, a second model predicting the other direction could be run in parallel. Alternatively, the anterior-posterior motion could be included as input, a possible extension to the models presented in this work. However, only predicting centroid positions would not allow for more advanced forms of dynamic MLC tracking (Ge *et al* 2014), where the MLC shape is adapted the predicted tumor location and shape, possibly taking into account in-plane rotations and deformations (Keall *et al* 2021). In a future study, we plan to extend the proposed LSTM to directly predict future 2D cine MR frames, thus allowing for dynamic MLC tracking. Finally, our model cannot predict out-of-plane motion as the cine MRIs are acquired in a single sagittal plane. However, several methods have been proposed to obtain time-resolved volumetric MRI (Fayad *et al* 2012, Stemkens *et al* 2016, Paganelli *et al* 2018a, Rabe *et al* 2021), which might be combined with our motion prediction model in future studies.

5. Conclusions

In this study, we developed LSTMs for SI tumor centroid position prediction based on cine MRIs acquired with 0.35 T MRIdian machines from two different institutions and showed that they outperformed state-of-the-art LR algorithms for all investigated forecasts (250 ms, 500 ms and 750 ms). The proposed models generalized their predictive performance to different testing sets with different breathing patterns, ranging from free-breathing to treatments with prolonged breath-holds. The continuously re-optimized *offline+online* LSTM network achieved superior performance in all tasks compared to offline optimized models. In conclusion, LSTMs were shown to have great potential as respiratory motion predictors to account for the system latencies present in MR-guided RT with MLC tracking.

Acknowledgments

Henning Schmitz is thanked for support related to the manuscript preparation, Andreas Haslauer for initial work on the cine videos and Seyed-Ahmad Ahmadi for fruitful discussions. This work was supported by the German Research Foundation (DFG) within the Research Training Group GRK 2274.

Conflict of interest

The Department of Radiation Oncology of the University Hospital of the LMU Munich has research agreements with Brainlab, Elekta and ViewRay.

Appendix

Table A1. Best hyper parameters found for each model by repeatedly performing training and validation for the 250 ms forecast.

Model	Nr. of layers	Dropout rate	Learning rate	Batch size	L2 weight	Input window length
Offline LSTM	3	0	5×10^{-4}	64	0	32
Offline+online LSTM	3	0	1×10^{-6}	Fixed	1×10^{-6}	32
Offline LR					1×10^{-2}	32
Online LR					1×10^{-5}	16

Table A2. Best hyper parameters found for each model by repeatedly performing training and validation for the 500 ms forecast.

Model	Nr. of layers	Dropout rate	Learning rate	Batch size	L2 weight	Input window length
Offline LSTM	5	0	5×10^{-4}	128	0	32
Offline+online LSTM	5	0	1×10^{-6}	Fixed	1×10^{-6}	32
Offline LR					1×10^{-2}	32
Online LR					1×10^{-4}	8

Table A3. Best hyper parameters found for each model by repeatedly performing training and validation for the 750 ms forecast.

Model	Nr. of layers	Dropout rate	Learning rate	Batch size	L2 weight	Input window length
Offline LSTM	5	0	5×10^{-4}	64	0	32
Offline+online LSTM	5	0	1×10^{-6}	Fixed	1×10^{-6}	32
Offline LR					1×10^{-2}	32
Online LR					1×10^{-5}	8

Table A4. Mean and standard deviation of MEs for the LMU testing set without breath holds. The ME of the best performing model is shown in bold for each forecasted time span.

Model	250 ms forecast	500 ms forecast ME [mm]	750 ms forecast
Offline LSTM	2.30 ± 2.04	5.12 ± 3.99	8.19 ± 5.57
Offline+online LSTM	2.18 ± 1.95	4.81 ± 3.73	7.81 ± 5.21
Offline LR	2.43 ± 2.09	5.31 ± 4.15	8.33 ± 5.87
Online LR	2.83 ± 2.05	6.02 ± 3.78	9.88 ± 6.81

ORCID iDs

Elia Lombardo  <https://orcid.org/0000-0003-4522-8298>

Moritz Rabe  <https://orcid.org/0000-0002-7085-4066>

Davide Cusumano  <https://orcid.org/0000-0001-8626-0003>

Marco Riboldi  <https://orcid.org/0000-0002-2431-4966>

Guillaume Landry  <https://orcid.org/0000-0003-1707-4068>

References

- Corradini S *et al* 2019 MR guidance in clinical reality: current treatment challenges and future perspectives *Radiat. Oncol.* **14** 92
- Cusumano D *et al* 2021 Artificial intelligence in magnetic resonance guided radiotherapy: Medical and physical considerations on state of art and future perspectives *Phys. Med.* **85** 175–91
- Fayad H J *et al* 2012 A generic respiratory motion model based on 4D MRI imaging and 2D image navigators *IEEE Nuclear Science Symp. and Medical Imaging Conf.*

2.1. Paper I: Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy

35

- Friedman M 1937 The use of ranks to avoid the assumption of normality implicit in the analysis of variance *J. Am. Stat. Assoc.* **32** 675–701
- Ge Y et al 2014 Toward the development of intrafraction tumor deformation tracking using a dynamic multi-leaf collimator *Med. Phys.* **41** 061703
- Glitzner M et al 2019 Technical note: MLC tracking performance on the Elekta Unity MRI linac *Phys. Med. Biol.* **64** 15NT02
- Green O et al 2018 First clinical implementation of real-time, real anatomy tracking and radiation beam control *Med. Phys.* **45** 3728–40
- Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
- Jackson S et al 2019 MRI B₀ homogeneity and geometric distortion with continuous linac gantry rotation on an Elekta Unity MR linac *Phys. Med. Biol.* **64** 12NT01
- Joehl A et al 2020 Performance comparison of prediction filters for respiratory motion tracking in radiotherapy *Med. Phys.* **47** 643–50
- Josipovic M, Anzar M C, Thomsen J B, Scherman J, Damkjaer S M, Nygard L, Specht L, Pohl M and Persson G F 2019 Deep inspiration breath hold in locally advanced lung cancer radiotherapy: validation of intrafractional geometric uncertainties in the INHALE trial *Br J Radiol.* **92** 1104
- Keall P et al 2021 AAPM task group 264: the safe clinical implementation of MLC tracking in radiotherapy *Med. Phys.* **48** e44–64
- Kim T et al 2020 Clinical experience of MRI4D QUASAR motion phantom for latency measurements in 0.35 T MR Linac *J. Appl. Clin. Med. Phys.* **22** 128–36
- Kingma D P and Ba J 2015 Adam: A method for stochastic optimization *3rd Int. Conf. for Learning Representations*
- Klueter S 2019 Technical design and concept of a 0.35 T MR linac *Clin. Transl. Radiat. Oncol.* **18** 98–101
- Krauss A, Nill S and Oelfke U 2011 The comparative performance of four respiratory motion predictors for real-time tumour tracking *Phys. Med. Biol.* **56** 5303–17
- Kurz C et al 2020 Medical physics challenges in clinical MR-guided radiotherapy *Radiat. Oncol.* **15** 93
- Lin H et al 2019 Towards real-time respiratory motion prediction based on long short-term memory neural networks *Phys. Med. Biol.* **64** 085010
- Liney G P et al 2018 MRI linear accelerator radiotherapy systems *Clin. Oncol.* **30** 686–91
- Menten M J et al 2016 Lung stereotactic body radiotherapy with an MR linac—quantifying the impact of the magnetic field and real-time tumor tracking *Radiother. Oncol.* **119** 461–6
- Menten M J et al 2020 Automatic reconstruction of the delivered dose of the day using MR linac treatment log files and online mr imaging *Radiother. Oncol.* **145** 88–94
- Murphy M J and Pokhrel D 2009 Optimization of an adaptive neural network to predict breathing *Med. Phys.* **36** 40–7
- Nemenyi P B 1963 Distribution-free multiple comparisons *PhD Thesis* Princeton University
- Paganelli C et al 2018a Feasibility study on 3D image reconstruction from 2D orthogonal cine-MRI for MRI-guided radiotherapy *J. Med. Imaging Radiat. Oncol.* **62** 389–400
- Paganelli C et al 2018b MRI guidance for motion management in external beam radiotherapy: current status and future challenges *Phys. Med. Biol.* **63** 22TR03
- Paszke A et al 2017 Automatic differentiation in PyTorch *NIPS W*
- Pedregosa F et al 2011 Scikit-learn: Machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
- Placidi L et al 2020 On-line adaptive MR-guided radiotherapy for locally advanced pancreatic cancer: Clinical and dosimetric considerations *Tech. Innov. Patient Support Radiat. Oncol.* **15** 15–21
- Poulsen P R et al 2010 Detailed analysis of latencies in image-based dynamic MLC tracking *Med. Phys.* **37** 4998–5005
- Rabe M et al 2021 Porcine lung phantom-based validation of estimated 4D-MRI using orthogonal cine imaging for low-field MR-Linacs *Phys. Med. Biol.* **66** 055006
- Remy C et al 2021 Potential of a probabilistic framework for target prediction from surrogate respiratory motion during lung radiotherapy *Phys. Med. Biol.* **66** 105002
- Roerdink J B T M and Meijster A 2000 The watershed transform: definitions, algorithms and parallelization strategies *Fundam. Inform.* **41** 187–228
- Seregni M et al 2016 Motion prediction in MRI-guided radiotherapy based on interleaved orthogonal cine-MRI *Phys. Med. Biol.* **61** 872–87
- Sharp G et al 2004 Prediction of respiratory tumour motion for real-time image-guided radiotherapy *Phys. Med. Biol.* **49** 425
- Shen C et al 2020 An introduction to deep learning in medical physics: advantages, potential, and challenges *Phys. Med. Biol.* **65** 05TR01
- Stemkens B et al 2016 Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy *Phys. Med. Biol.* **61** 5335
- Sun W et al 2020 Adaptive respiratory signal prediction using dual multi-layer perceptron neural networks *Phys. Med. Biol.* **65** 185005
- Uijtewaal P et al 2021 Dosimetric evaluation of MRI-guided multi-leaf collimator tracking and trailing for lung stereotactic body radiation therapy *Med. Phys.* **48** 1520–32
- Wang R et al 2018 A feasibility of respiration prediction based on deep bi-LSTM for real-time tumor tracking *IEEE Access* **6** 51262–8
- Yu S et al 2020 Rapid prediction of respiratory motion based on bidirectional gated recurrent unit network *IEEE Access* **8** 49424–35
- Yun J et al 2012 An artificial neural network (ANN)-based lung tumor motion predictor for intrafractional MR tumor tracking *Med. Phys.* **39** 4423–33

2.2 Paper II: Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy

In this follow-up study to the first publication, using the same datasets the models were extended to predict the future 2D target contour instead of the 1D centroid position. Specifically, three LSTMs were implemented: 1) the LSTM-shift model is an extension of the model from the previous study to predict both the centroid in superior-inferior and in anterior-posterior direction. The difference between the predicted centroid position and the last available centroid position is used to rigidly shift the 2D contour obtained from the Viewray target localization algorithm. 2) an encoder-decoder convolutional LSTM predicting directly the future 2D contour and 3) an encoder-decoder convolutional LSTM predicting future 2D deformations which are used to warp the available tumor contour, similarly to Romaguera et al. [59]. As the convolutional LSTMs have more parameters, only population-based training was performed, while the LSTM-shift was continuously re-optimized as in the previous study. It was shown that the two LSTM-shift models, and in particular the patient-specific one, performed the best. This suggests that no large changes in the tumor contour happen from the last available frame to the predicted frame 500 ms in the future.

Reprinted with permission from "Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. (2023) Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy. *Radiotherapy and Oncology*, 182, 109555".



Original Article

Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy



Elia Lombardo^a, Moritz Rabe^a, Yuqing Xiong^a, Lukas Nierer^a, Davide Cusumano^b, Lorenzo Placidi^b, Luca Boldrini^b, Stefanie Corradini^a, Maximilian Niyazi^a, Michael Reiner^a, Claus Belka^{a,c}, Christopher Kurz^a, Marco Riboldi^d, Guillaume Landry^{a,*}

^a Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany; ^b Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome 00168, Italy; ^c German Cancer Consortium (DKTK), Munich 81377; and ^d Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Garching b. München 85748, Germany

ARTICLE INFO

Article history:

Received 26 September 2022

Received in revised form 24 January 2023

Accepted 5 February 2023

Available online 21 February 2023

Keywords:

Artificial intelligence

Long short-term memory networks

Time series prediction

Respiratory motion

MR-linac

MLC-tracking

ABSTRACT

Background and purpose: Magnetic resonance imaging guided radiotherapy (MRgRT) with deformable multileaf collimator (MLC) tracking would allow to tackle both rigid displacement and tumor deformation without prolonging treatment. However, the system latency must be accounted for by predicting future tumor contours in real-time. We compared the performance of three artificial intelligence (AI) algorithms based on long short-term memory (LSTM) modules for the prediction of 2D-contours 500 ms into the future.

Materials and methods: Models were trained (52 patients, 3.1 h of motion), validated (18 patients, 0.6 h) and tested (18 patients, 1.1 h) with cine MRs from patients treated at one institution. Additionally, we used three patients (2.9 h) treated at another institution as second testing set. We implemented 1) a classical LSTM network (LSTM-shift) predicting tumor centroid positions in superior-inferior and anterior-posterior direction which are used to shift the last observed tumor contour. The LSTM-shift model was optimized both in an *offline* and *online* fashion. We also implemented 2) a convolutional LSTM model (ConvLSTM) to directly predict future tumor contours and 3) a convolutional LSTM combined with spatial transformer layers (ConvLSTM-STL) to predict displacement fields used to warp the last tumor contour.

Results: The *online* LSTM-shift model was found to perform slightly better than the *offline* LSTM-shift and significantly better than the ConvLSTM and ConvLSTM-STL. It achieved a 50% Hausdorff distance of 1.2 mm and 1.0 mm for the two testing sets, respectively. Larger motion ranges were found to lead to more substantial performance differences across the models.

Conclusion: LSTM networks predicting future centroids and shifting the last tumor contour are the most suitable for tumor contour prediction. The obtained accuracy would allow to reduce residual tracking errors during MRgRT with deformable MLC-tracking.

© 2023 Elsevier B.V. All rights reserved. Radiotherapy and Oncology 182 (2023) 109555

Magnetic resonance imaging (MRI) guided radiotherapy (MRgRT) offers excellent treatment adaptation capabilities thanks to the high soft tissue contrast of MRI coming with no additional dose [1,2]. Intra-fractional respiratory motion, which has been shown to decrease the accuracy of radiotherapy [3,4], can be continuously monitored with real-time cine MRI, allowing the implementation of motion mitigation strategies to optimally treat the tumor while sparing the surrounding healthy tissue. Multileaf collimator (MLC) tracking is a promising motion management strategy which consists of shifting or reshaping the beam aperture to follow the target motion in real-time [5,6]. Therefore, it allows

for smaller tumor margins during the treatment planning stage compared to e.g. mid-ventilation planning, leading to decreased organs-at-risk dose without extending the treatment duration [7,8]. This is an advantage compared to beam gating, for which duty cycle efficiencies of 20% to 50% have been reported for MR-linac treatments [9,10], depending on patient compliance. MLC-tracking can be implemented by rigidly shifting the MLC by the tumor centroid displacement or by dynamically moving the leaves to the positions that best match the target, the latter allowing to additionally account for target rotations and deformations observed during treatment [11,12]. The technical feasibility of rigid MLC-tracking has been demonstrated on the 1.5T Unity MR-linac (Elekta AB, Stockholm, Sweden) for stereotactic lung treatments with intensity modulated radiation therapy (IMRT) plans [13]

* Corresponding author at: Marchionistr. 15, 81377 Munich, Germany.

E-mail address: guillaume.landry@med.uni-muenchen.de (G. Landry).

while a proof-of-principle study showed the feasibility of adaption of the MLC aperture to account for tumor deformation [14].

To fully leverage MLC-tracking, the system latency has to be accounted for in real-time [15]. The latency is defined as the time difference between the motion of the tumor and, in the case of MLC-tracking, the execution of the MLC motion [7]. Experimental studies have found an MLC-tracking end-to-end latency of 350ms for the Unity MR-linac [16], 330ms for the Australian MR-linac [17] and 340ms for the linac-MR (University of Alberta) [18]. In the last years, several motion prediction algorithms have been proposed to compensate for system latencies in RT [19–26]. The presented algorithms take a series of past target positions as input and predict the future target positions in 1D, 2D or 3D, therefore providing latency compensation via rigid shifts of the MLC. On the other hand, only few studies have been conducted to predict future tumor contours (or equivalently binary masks) [27–29]. Romaguera et al. introduced an artificial intelligence (AI) model to predict deformation vector fields which they used to obtain future cine MRI frames (at a diagnostic 3T scanner), however, they did not specifically investigate contour prediction [30].

Recently, AI models have been successfully applied on a variety of tasks in medical physics [31–35], including MRgRT [36]. For motion prediction, a class of AI models called Long Short-Term Memory (LSTM) networks has been shown to perform best in a series of studies [23,24,26], as they are designed to take into account long-term temporal dependencies in the input data [37]. In a previous study [26], we showed that an LSTM model outperformed a state-of-the-art linear regression for the prediction of superior-inferior (SI) tumor center of mass positions. Shi et al. implemented the convolutional LSTM, which by construction can handle both temporal and spatial correlations in the input data and showed it to outperform the LSTM on a precipitation forecasting problem [38].

In this work, we compared three different AI-based contour prediction models using LSTM which could be used for deformable MLC-tracking during MRgRT. Compared to previous studies [27–29], this is the first time that AI algorithms are implemented and compared for this task and their performance is evaluated on clinically acquired contours. Compared to the previous study [26], we extended a classical LSTM to predict future tumor positions in two directions of motion and to shift an existing contour in real-time and compared it with two variants of convolutional LSTM networks [38,30] directly providing the future tumor contour. The models were evaluated with multiple overlap and distance metrics to investigate the accuracy for deformable MLC-tracking using data acquired during patient treatments with 0.35 T MR-linacs at two institutions.

Materials and methods

Datasets

We retrospectively collected 2D + t cine MRIs from 88 patients treated at the Department of Radiation Oncology of LMU Munich. Tumor sites comprised lung (37), pancreas (22), heart (6), liver (20) and mediastinum (3) and all patients were treated using breath-hold and gating with the MRIdian MR-linac. We also collected imaging data of 3 patients irradiated at Gemelli Rome in free-breathing with the MRIdian MR-Cobalt-60 machine. For these patients, tumor sites comprised lung (2) and pancreas (1). All cine MRIs were acquired at 4Hz in a sagittal plane with a balanced steady-state free precession sequence (TrueFISP; in-plane resolution $3.5 \times 3.5 \text{ mm}^2$; field-of-view $270 \times 270 \text{ mm}^2$ or $350 \times 350 \text{ mm}^2$; slice thickness of 5, 7 or 10mm) and exported in the OGV video format as supported by the vendor. Each video contained the up-sampled cine MRI with tracking target and gating

boundary contours (color-coded). The tracking target contours were obtained during treatment by propagating the contour from a reference frame to the current frame with a deformable image registration algorithm by the vendor [39,40]. These contours were used both as input and as ground truth output (at later time points) for model optimization. Informed consent was obtained from all patients and all methods were carried out in accordance with relevant ethics guidelines and regulations.

Data pre-processing

To extract the target tracking contours from the OGV video data, we used the same in-house software developed previously [26]. In contrast to the former study, where we extracted only the SI center of mass motion trajectories, we also extracted the anterior-posterior (AP) motion trajectories and additionally kept the binary masks obtained from the extracted contours. More details can be found in the [supplementary material](#).

In summary, after pre-processing we obtained 4.8h of binary mask and SI/AP center of mass motion for the LMU cohort without breath-holds (40.3 h if the breath-holds were included) and 2.9 h of motion for the free-breathing Gemelli cohort.

Motion prediction models

Motion prediction can be formulated as the task of obtaining future data points from past data points at a given time step t :

$$\hat{\mathbf{y}}_t = f(\mathbf{x}_t) \quad (1)$$

where $f()$ is a predictor algorithm, $\mathbf{x}_t$ is the t -th fixed window of input data and $\hat{\mathbf{y}}_t$ is the t -th fixed window of predicted output data. In this study, an element of the input/output window can be a vector with the centroid coordinates in SI and AP or a 2D array representing the binary tumor contour. The length of the input window is a hyper-parameter while the length of the predicted window determines the forecasted horizon. This parameter was set equal to two, which corresponds to a 500ms prediction when using an imaging frequency of 4Hz. We implemented three different algorithms based on LSTM modules for the prediction of contours and compared them with a baseline no-predictor:

- LSTM-shift: we extended an LSTM model developed in a previous study [26] to predict the future centroid coordinates in SI and AP and used the difference between the predicted centroids and the centroids of the last available tumor contour to shift that contour. The LSTM-shift was trained both *offline* and *online*, the latter to enable real-time patient-specific adaption to recent respiration.
- ConvLSTM: we used the convolutional LSTM introduced by Shi et al. [38] to directly predict future tumor contours from a sequence of observed contours.
- ConvLSTM-STL: we implemented the convolutional LSTM combined with spatial transformer layers proposed by Romaguera et al. [30] to predict future deformation vector fields which were used to warp the last observed tumor contour.
- No-predictor: we considered the last observed contour without any changes as baseline prediction.

A more detailed description of the models can be found in the [supplementary materials](#).

Model optimization

Data partitioning

All three models were trained using 60% of the LMU patients (52 in total), validated on 20% of the LMU patients (18) and tested with

the remaining 20% of the LMU patients (18). Additionally, we used all three Gemelli free-breathing patients as an independent testing set. For the LMU validation and testing sets, we always excluded the breath-holds while for the training set we investigated whether inclusion or exclusion would lead to better performance during inference, that is, we treated the presence of breath-holds as a hyper-parameter (see 2.4.2). If the breath-holds were included, the total duration of the training set amounted to 21.6h (2013 frames on average per fraction). Without the breath-holds the duration of the training set amounted to 3.1h (618 frames on average), the duration of the validation set to 0.6h (389 frames on average) and the duration of the testing set to 1.1h (516 frames on average). The duration of the free-breathing Gemelli testing set was 2.9h (2753 frames on average). The original cine MRI time resolution of 4Hz was not changed.

Hyper-parameters

The optimal set of hyper-parameters for each model was found by repeatedly performing training and validation. The weights of each model leading to the best validation performance were saved to disk and loaded to perform testing without any further changes of the model settings. A detailed description of the investigated parameters can be found in the [supplementary material](#). To assess model stability, once we found the best parameters, we repeated training and validation a second time to see whether the validation accuracy changes between two optimizations with the same hyper-parameters.

Performance and statistics

To quantify the performance of the model, we calculated the dice similarity coefficient [41] (DSC), the Hausdorff distance [42] (HD) 50% and 95% (defined as 50th and 95th percentile of the distances between boundary points, respectively) between the predicted masks and the true output masks. Prior to the computation of these metrics, we thresholded the predicted masks at 0.5 to make them binary. Furthermore, we computed the center of mass coordinates of the predicted masks and evaluated the alignment with the true binary mask by computing the root mean squared error (RMSE) between predicted and true center of mass positions in SI and AP separately. For each of these metrics, we computed the mean and standard deviation of the metric over all cine MRIs in a testing cohort.

To see whether there is a statistically significant difference between the metrics obtained with the different models on the testing sets, we merged all metrics into a single vector and used non-parametric Friedman tests [43], considering a p-value < 0.05 significant. If the Friedman test was significant, we performed a post hoc Nemenyi test [44] to infer which model obtained significantly better metrics in a pair-wise fashion.

Results

Table 1 shows for the two variants of the LSTM-shift, the ConvLSTM, the ConvLSTM-STL and the baseline no-predictor the best geometric accuracy performance achieved on the validation set after the hyper-parameter search. Overall, we found the *online*

LSTM-shift model to perform best for all investigated metrics. Improvements of mean metrics were small (only visible without rounding) compared to the *offline* LSTM-shift and more substantial compared to the convolutional LSTMs. As shown in [supplementary Table S2](#), the models achieved the same or a very similar performance when repeating the optimization.

Tables 2 and 3 show the performance achieved with the best validation models on the LMU and Gemelli testing sets, respectively. Again, we found the *online* LSTM-shift model to perform best in terms of both the mean and the standard deviation of all metrics. The *offline* LSTM-shift performed slightly worse than the *online* LSTM-shift, followed by the ConvLSTM-STL, the ConvLSTM and the no-predictor.

The Friedman test revealed that statistically significant differences were present among the models both for the LMU and for the Gemelli testing set. Table S3 in the [supplementary material](#) shows the p-values obtained with the Nemenyi test for the pair-wise comparisons of all computed metrics. We found the *online* LSTM-shift to perform significantly better than the *offline* for the Gemelli set but not for the LMU set. Both LSTM-shift models were significantly better than the ConvLSTM. The ConvLSTM-STL was significantly better than the ConvLSTM for the Gemelli set but not for the LMU set. All LSTMs were significantly better than the no-predictor on both testing sets.

Fig. 1 shows for three different testing patients the comparison between the ground truth output and the prediction by the different models in the sagittal plane. Qualitatively, it can be noted that the LSTM-shift prediction most closely matched the actual ground truth binary mask, especially due to correct alignment rather than agreement in shape.

Fig. 2 shows a comparison between the SI centroid positions extracted from the true future binary mask, the prediction by the *online* LSTM-shift and the centroids computed from the masks predicted by the ConvLSTM-STL and the no-predictor. It can be noticed that the motion curve predicted by the LSTM-shift achieved the best alignment with the ground truth curve, especially in the middle of inhalation and exhalation, where differences to previous points were larger. The centroids obtained from the ConvLSTM-STL match the true curve better than the no-predictor model during inhalations and exhalations, however the ConvLSTM-STL overshoots slightly more. For comparability among figures, the first six centroid positions in Fig. 2 correspond to the six contours shown in the top row of Fig. 1.

Line plots showing the RMSE_{SI} and the inter-quartile range (IQR) of motion as a function of patient number for the LMU and the Gemelli testing sets are shown in Fig. 3. Substantial differences in the performance of the models can be observed among the patients. These differences appear to correlate with the range of motion, with larger motion leading to larger differences between models.

In Fig. 4 the original cine MRI frames with boundary contour obtained from the OGV videos is overlaid with the extracted tracking target contour used as ground truth for optimization and the prediction of our best model, i.e., the *online* LSTM-shift. It can be seen that, for these frames, the prediction of the LSTM-shift follows the ground truth closely.

Table 1
Mean and standard deviation of different metrics for the 500ms forecasted masks for the validation set.

Model	HD _{50%} [mm]	HD _{95%} [mm]	DSC	RMSE _{SI} [mm]	RMSE _{AP} [mm]
<i>offline</i> LSTM-shift	1.1 ± 0.4	2.6 ± 1.0	0.91 ± 0.05	1.2 ± 0.7	0.7 ± 0.3
<i>online</i> LSTM-shift	1.1 ± 0.4	2.6 ± 1.0	0.91 ± 0.04	1.2 ± 0.7	0.7 ± 0.3
ConvLSTM	1.3 ± 0.6	3.1 ± 1.5	0.90 ± 0.06	1.6 ± 1.0	0.8 ± 0.4
ConvLSTM-STL	1.3 ± 0.6	3.1 ± 1.4	0.90 ± 0.06	1.6 ± 1.0	0.8 ± 0.4
No-predictor	1.6 ± 0.8	3.6 ± 1.7	0.87 ± 0.06	2.2 ± 1.3	1.0 ± 0.5

Contour prediction with LSTMs for MRgRT

Table 2

Mean and standard deviation of different metrics for the 500ms forecasted masks for the LMU testing set.

Model	HD _{50%} [mm]	HD _{95%} [mm]	DSC	RMSE _{SI} [mm]	RMSE _{AP} [mm]
offline LSTM-shift	1.3 ± 0.4	3.6 ± 1.1	0.92 ± 0.04	1.4 ± 0.5	0.8 ± 0.3
online LSTM-shift	1.2 ± 0.4	3.5 ± 1.2	0.92 ± 0.04	1.3 ± 0.6	0.8 ± 0.3
ConvLSTM	1.5 ± 0.6	4.1 ± 1.6	0.91 ± 0.05	1.9 ± 1.0	1.0 ± 0.4
ConvLSTM-STL	1.5 ± 0.6	4.0 ± 1.5	0.91 ± 0.05	1.9 ± 1.1	0.9 ± 0.4
No-predictor	1.8 ± 0.9	4.7 ± 1.9	0.89 ± 0.05	2.8 ± 1.6	1.3 ± 0.6

Table 3

Mean and standard deviation of different metrics for the 500ms forecasted masks for the Gemelli testing set.

Model	HD _{50%} [mm]	HD _{95%} [mm]	DSC	RMSE _{SI} [mm]	RMSE _{AP} [mm]
offline LSTM-shift	1.1 ± 0.3	2.8 ± 0.8	0.92 ± 0.02	1.3 ± 0.8	0.6 ± 0.1
online LSTM-shift	1.0 ± 0.3	2.7 ± 0.7	0.92 ± 0.02	1.1 ± 0.6	0.5 ± 0.1
ConvLSTM	1.3 ± 0.4	3.3 ± 1.0	0.91 ± 0.03	1.7 ± 0.9	0.7 ± 0.2
ConvLSTM-STL	1.2 ± 0.4	3.1 ± 0.9	0.91 ± 0.02	1.6 ± 0.9	0.6 ± 0.1
No-predictor	1.5 ± 0.6	3.6 ± 1.2	0.89 ± 0.03	2.4 ± 1.4	0.8 ± 0.2

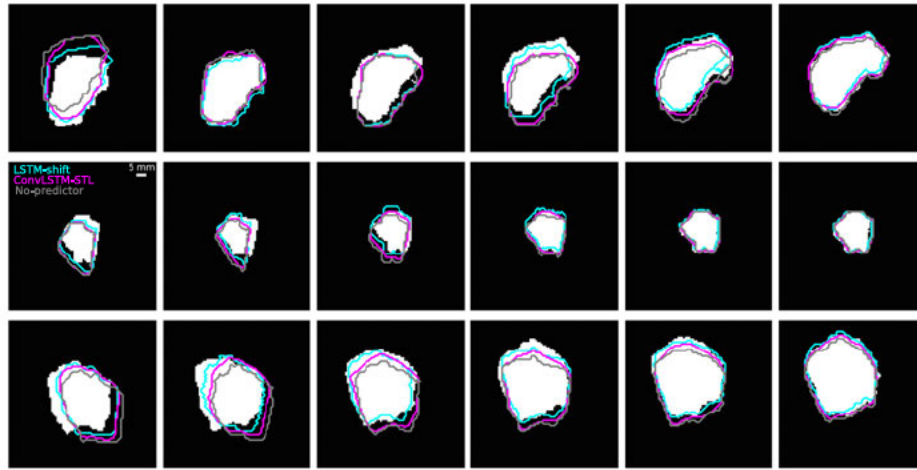


Fig. 1. Comparison between ground truth (binary mask in white) and predictions (colored contours) for the *online* LSTM-shift (cyan), the ConvLSTM-STL (purple) and the baseline no-predictor (grey) model for the investigated forecast (500ms). Sequence of contours for a Gemelli testing patient with a pancreas tumor (top row), an LMU testing patient with a lung tumor (middle row) and an LMU patient with a pancreas tumor (bottom row). For visualization purposes, the predicted masks are presented as contours.

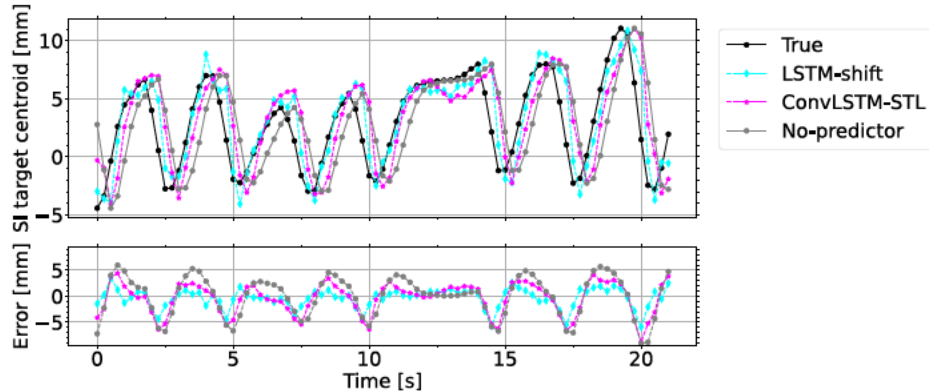


Fig. 2. Comparison between ground truth (black) and predicted SI centroid positions for the *online* LSTM-shift (cyan), the ConvLSTM-STL (purple) and the baseline no-predictor (grey) model for a Gemelli testing patient with a pancreas tumor (same as top row of Fig. 1). The bottom row shows the difference between the true centroid position and the predictions by the different models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Paper II: Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy

41

E. Lombardo, M. Rabe, Y. Xiong et al.

Radiotherapy and Oncology 182 (2023) 109555

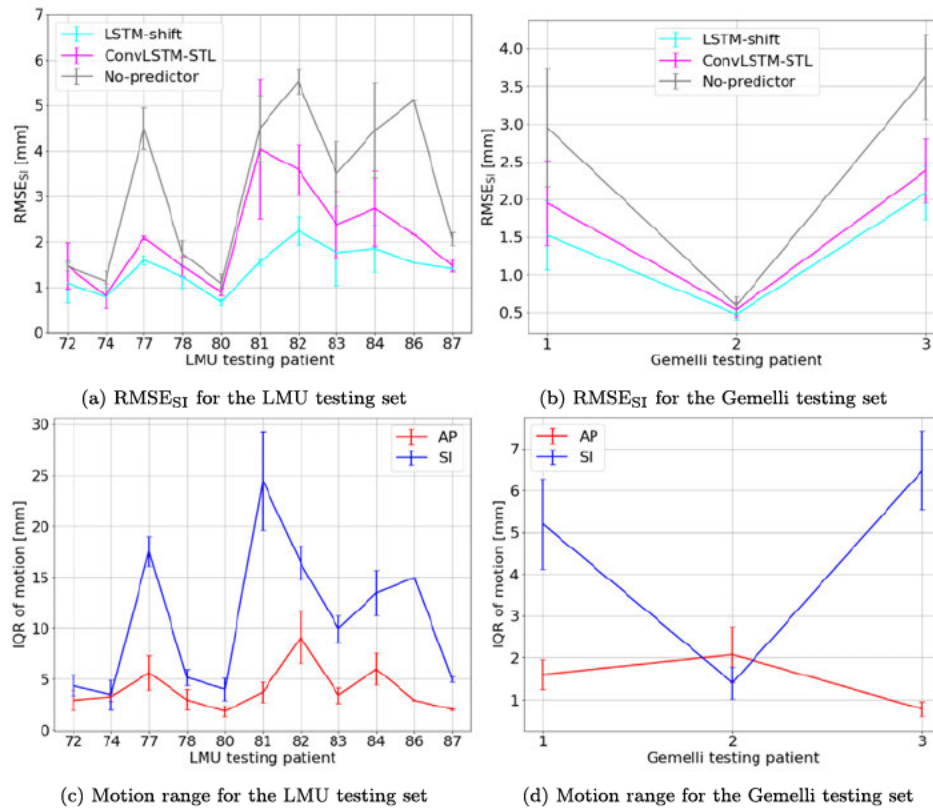


Fig. 3. Line plots showing the RMSE_{SI} for the different models and the IQR of motion in both directions as a function of patient number for the two testing cohorts. Average and standard deviation are computed over treatment fractions of patients.

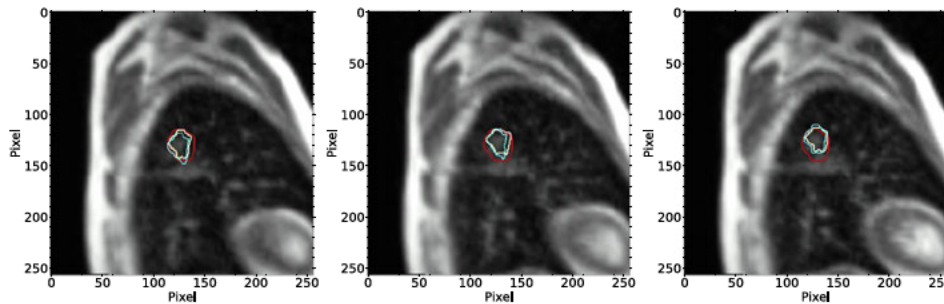


Fig. 4. Cine MRI frames with gating boundary contour (red) for the LMU lung tumor testing patient shown in the middle row of Fig. 1. The ground truth output contour extracted with in-house software is displayed in white while the prediction by the online LSTM-shift is shown in cyan. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In total, a prediction with the LSTM-shift model takes (17 ± 3) ms, where about 1 ms comes from the model's forward pass, 1 ms from the normalization of the input window and 15 ms from shifting the last input contour. A prediction with the ConvLSTM takes (14 ± 1) ms (forward pass) and a prediction with the ConvLSTM-STL takes (45 ± 1) ms (forward pass).

Discussion

In this study, we have implemented three LSTM-based models and evaluated their performance for the prediction of tumor con-

tours on testing cine MRI data from 0.35 T MR-linacs from two different institutions.

The continuously re-optimized LSTM from a previous study[26] was extended to predict both SI and AP centroid positions and used to shift the tumor contour rigidly on basis of the last input frame by the difference between prediction and last input (*online* LSTM-shift). We found this model to perform significantly better than all other models except from the *offline* LSTM-shift on the LMU testing set, where the performance was slightly but not significantly better. This could be explained by the fact that online re-optimization based on recent motion enhances the performance particularly if the respiration is more regular (albeit patient-specific) or if baseline drifts are present, such as for the free-

breathing Gemelli patients. The ConvLSTM-STL by Romaguera et al. [30] was found to achieve better performance than the ConvLSTM by Shi et al. [38], however both models performed less well compared to the *offline* and *online* LSTM-shift. The reason for that might be that, as hypothesized by Bourque et al. [27], the shape of the tumor on the last acquired image is a good proxy of the shape of the tumor 250ms and, in this study, 500ms in the future. From this, we conclude that it might be more relevant to train models which focus on correct alignment of the centroids (i.e., LSTM-shift) rather than models that additionally try to predict changes in shape within the 500ms prediction. Another explanation might be that the amount of training data is not sufficient for the more complex convolutional LSTMs, which have substantially more weights than the LSTM-shift models (about 2×10^6 parameters compared with 1×10^4). The fact that training with data augmentations improved the performance of both convolutional LSTMs seems to support this hypothesis.

The *online* LSTM-shift significantly outperformed both convolutional LSTMs and the no-predictor while showing smaller standard deviations, meaning that it could provide robust predictions over a large span of respiratory patterns. While the standard deviation might still seem large compared to the average improvements, it should be noted that its magnitude does not derive from noisiness in the models, as shown by the stable validation accuracy obtained when re-training (see Table S2). Instead it is caused by the high variability in respiratory patterns translating in substantially different prediction accuracies from one patient to the other (see Fig. 3). Even though the differences between the models were statistically significant, the gain in accuracy achieved by the *online* LSTM-shift will not be clinically relevant for certain patients (e.g. LMU patient 74 and Gemelli patient 2 in Fig. 3). For other patients the improvement can exceed 2mm over the ConvLSTM-STL and 3mm over the no-predictor (e.g. LMU patient 81 and 82). Looking at the corresponding line plots for the motion range, we noticed a correlation between large motion and substantial performance differences between the models, and small motion and negligible differences between the models. Therefore, the larger the patient's motion, the more likely it is to obtain a benefit from using the LSTM-shift compared to the other models. The models were trained to predict both 250ms and 500ms into the future, thus the prediction horizon could easily be adjusted to machine specific latencies (which are typically 330ms to 350ms) [16–18], by using interpolation between the two predictions.

Compared to previous contour prediction studies [27,29], this is the first time different AI models, which have been shown to outperform traditional algorithms, are compared. A direct comparison between the studies is not possible as the testing data is different. However, we can say that the testing RMSE_{SI} of 1.3 mm (LMU) and 1.1 mm (Gemelli) for the 500ms forecast are comparable with the RMSE of 1.3 mm obtained by Bourque et al. for the 250ms forecast and the average median centroid distance of 0.63 mm reported by Ginn et al. for 250–330ms forecasts. In general, the achieved predictive performance would be sufficient to improve MLC-tracking accuracy, while further performance improvements might be limited by the in-plane resolution of $3.5 \times 3.5 \text{ mm}^2$ of the cine MRIs.

Even though it is meaningful to use clinically acquired contours, it represents one of the main limitations of this study. The predictive performance of our algorithms is inherently limited by the performance of the clinical tracking algorithm, which was observed to present some noisiness and fluctuations in target shape, especially when tracking smaller targets. Additionally, the in-house contour extraction software was not always able to perfectly extract the vendor's contour from the videos. Even though we excluded videos for which this was happening more often (see supplementary material), it is very likely that a few incorrectly

extracted contours are still present in our datasets. However, as both the convolutional LSTMs and the LSTM-shift are either based directly on the extracted binary masks or use centroids computed from them, incorrect contours and the 3.5mm pixel size will affect all models in a similar way. In future studies, we would like to build an AI deformation model which could be trained to perform the combined contour tracking and prediction. It should be noted that contour prediction using one of the proposed models would currently be possible with the MRIdian MR-linac [40] or with the Alberta linac-MR [18], which provide the current tumor shape using either deformable image registration or artificial neural networks, while other machines currently localize the tumor using template matching [13,17] and therefore provide the tumor center of mass. Another limitation of the study is that the models cannot predict out-of-plane motion, as the underlying cine MRI data is acquired on a single sagittal 2D slice. However, several methods have been discussed in the literature to obtain time-resolved volumetric MRIs [45–48], which could be combined with motion prediction models especially considering the increase in latency expected when generating 3D real-time data. In particular for MLC-tracking, also beam's eye view cine MRI with tumor-volume projection might represent an acquisition scheme ensuring beam conformity [49,50].

Conclusions

We have implemented and compared AI models based on LSTM for the prediction of future 2D tumor contours using data from treatments with 0.35T MR-linacs from two institutions. The experiments showed that the LSTM-shift models, i.e., LSTMs predicting future center of mass positions, which are used to shift the last available tumor mask, significantly outperformed the convolutional LSTM counterparts and obtained substantial improvements especially when the motion range was larger. They are therefore very promising candidates to overcome the system latency present in deformable MLC-tracking during MRgRT. An advantage of the LSTM model relative to the convolutional LSTMs is that, given its smaller number of parameters, it can be continuously re-optimized on current motion data to further improve its performance.

Conflict of Interest Statement

The Department of Radiation Oncology of the University Hospital of LMU Munich has research agreements with Brainlab, Elekta and ViewRay.

Acknowledgments

Seyed-Ahmad Ahmadi is thanked for fruitful discussions. This work was supported by the German Research Foundation (DFG) within the Research Training Group GRK 2274.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.radonc.2023.109555>.

References

- [1] Kurz C et al. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol* 2020;15:93. <https://doi.org/10.1186/s13014-020-01524-4>.
- [2] Alongi F et al. 1.5 T MR-guided and daily adapted SBRT for prostate cancer: feasibility, preliminary clinical tolerability, quality of life and patient-reported

2.2. Paper II: Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy

43

E. Lombardo, M. Rabe, Y. Xiong et al.

Radiotherapy and Oncology 182 (2023) 109555

- outcomes during treatment. *Radiat Oncol* 2020;15:1 9. <https://doi.org/10.1186/s13014-020-01510-w>.
- [3] Langen KM, Jones DT. Organ motion and its management. *Int J Radiat Oncol Biol Phys* 2001;50:265–78. [https://doi.org/10.1016/s0360-3016\(01\)01453-5](https://doi.org/10.1016/s0360-3016(01)01453-5).
 - [4] Shirato H, Seppenwoolde Y, Kitamura K, Onimura R, Shimizu S. Intrafractional tumor motion: lung and liver. *Seminars in radiation oncology*, vol. 14. Elsevier; 2004. p. 10–8. <https://doi.org/10.1053/j.semradonc.2003.10.008>.
 - [5] Keall P, Kini V, Vedam S, Mohan R. Motion adaptive x-ray therapy: a feasibility study. *Phys Med Biol* 2001;46:1. <https://doi.org/10.1088/0031-9155/46/1/301>.
 - [6] Sawant A, Venkat R, Srivastava V, Carlson D, Povzner S, Cattell H, et al. Management of three-dimensional intrafraction motion through real-time DMLC tracking. *Med Phys* 2008;35:2050–61. <https://doi.org/10.1118/1.2905355>.
 - [7] Keall PJ, Sawant A, Berbeco RI, Booth JT, Cho B, Cerviño LI, et al. AAPM task group 264: The safe clinical implementation of MLC tracking in radiotherapy. *Med Phys* 2021;48:e44–64. <https://doi.org/10.1002/mp.14625>.
 - [8] Menten MJ, Fast MF, Nill S, Kamerling CP, McDonald F, Oelfke U. Lung stereotactic body radiotherapy with an mr-linac quantifying the impact of the magnetic field and real-time tumor tracking. *Radiother Oncol* 2016;119:461–6. <https://doi.org/10.1016/j.radonc.2016.04.019>.
 - [9] Eze C 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. URL <https://ro-journal.biomedcentral.com/articles/10.1186/s13014-022-02011-8>.
 - [10] Finazzi T, de Koste JRVs, Palacios MA, Spoelstra FO, Slotman BJ, Haasbeek CJ, et al. Delivery of magnetic resonance-guided single-fraction stereotactic lung radiotherapy. *Phys Imaging Radiat Oncol* 2020;14:17–23. <https://doi.org/10.1016/j.phro.2020.05.002>.
 - [11] 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;74:1092–9. <https://doi.org/10.1016/j.ijrobp.2008.09.031>.
 - [12] Feng M, Balter JM, Normolle D, Adusumilli S, Cao Y, Chenevert TL, et al. Characterization of pancreatic tumor motion using cine MRI: surrogates for tumor position should be used with caution. *Int J Radiat Oncol Biol Phys* 2009;74:884–91. <https://doi.org/10.1016/j.ijrobp.2009.02.003>.
 - [13] 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;48:1520–32. <https://doi.org/10.1002/mp.14772>.
 - [14] Ge Y, O'Brien RT, Shieh C-C, Booth JT, Keall PJ. Toward the development of intrafraction tumor deformation tracking using a dynamic multi-leaf collimator. *Med Phys* 2014;41:061703. <https://doi.org/10.1118/1.4873682>.
 - [15] Poulsen PR et al. Detailed analysis of latencies in image-based dynamic MLC tracking. *Med Phys* 2010;37:4998–5005. <https://doi.org/10.1118/1.3480504>.
 - [16] Glitzner M et al. Technical note: MLC-tracking performance on the Elekta Unity MRI-linac. *Phys Med Biol* 2019;64:15NT02. <https://doi.org/10.1088/1361-6560/ab2667>.
 - [17] Liu PZ 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;47:6440–9. <https://doi.org/10.1002/mp.14536>.
 - [18] 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. <https://doi.org/10.1118/1.4802735>.
 - [19] Sharp G et al. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys Med Biol* 2004;49:425. <https://doi.org/10.1088/0031-9155/49/3/006>.
 - [20] Krauss A, Nill S, Oelfke U. The comparative performance of four respiratory motion predictors for real-time tumour tracking. *Phys Med Biol* 2011;56:5303–17. <https://doi.org/10.1088/0031-9155/56/16/015>.
 - [21] Yun J et al. An artificial neural network (ANN)-based lung-tumor motion predictor for intrafractional MR tumor tracking. *Med Phys* 2012;39:4423–33. <https://doi.org/10.1118/1.4730294>.
 - [22] Seregni M et al. Motion prediction in MRI-guided radiotherapy based on interleaved orthogonal cine-MRI. *Phys Med Biol* 2016;61:872–87. <https://doi.org/10.1088/0031-9155/61/2/872>.
 - [23] Wang R et al. A feasibility of respiration prediction based on deep bi-LSTM for real-time tumor tracking. *IEEE Access* 2018;6. <https://doi.org/10.1109/ACCESS.2018.2869780>.
 - [24] Lin H et al. Towards real-time respiratory motion prediction based on long short-term memory neural networks. *Phys Med Biol* 2019;64:085010. <https://doi.org/10.1088/1361-6560/ab13fa>.
 - [25] Joehl A et al. Performance comparison of prediction filters for respiratory motion tracking in radiotherapy. *Med Phys* 2020;47:643–50. <https://doi.org/10.1002/mp.13929>.
 - [26] Lombardo E et al. Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. *Phys Med Biol* 2022;67:095006. <https://doi.org/10.1088/1361-6560/ac60b7>.
 - [27] Bourque AE, Carrier J-F, 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 Exp* 2017;3:035001. <https://doi.org/10.1088/2057-1976/aa6b5b>.
 - [28] Bourque A et al. Particle filter based target tracking algorithm for magnetic resonance guided respiratory compensation: robustness and accuracy assessment. *Int J Radiat Oncol Biol Phys* 2018;100:325–34. <https://doi.org/10.1016/j.ijrobp.2017.10.004>.
 - [29] 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;47:404–13. <https://doi.org/10.1002/mp.13948>.
 - [30] Romaguera LV, Plantefève R, Romero FP, Hébert F, Carrier J-F, Kadoury S. Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks. *Medical Image Anal* 2020;64:101754. <https://doi.org/10.1016/j.media.2020.101754>.
 - [31] Zanca F, Avanzo M, Colgan N, Crijns W, Guidi G, Hernandez-Giron I, et al. Focus issue: Artificial intelligence in medical physics. *Phys Med* 2021;83:287–91. <https://doi.org/10.1016/j.ejmp.2021.05.008>.
 - [32] Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Phys Med Biol* 2020;65:20TR01. <https://doi.org/10.1088/1361-6560/ab843c>.
 - [33] Kawula M et al. Dosimetric impact of deep learning-based CT auto-segmentation on radiation therapy treatment planning for prostate cancer. *Radiat Oncol* 2022;17:1–12. <https://doi.org/10.1186/s13014-022-01985-9>.
 - [34] Avanzo M et al. Machine and deep learning methods for radiomics. *Med Phys* 2020;47:e185–202. <https://doi.org/10.1002/mp.13678>.
 - [35] Lombardo E, Kurz C, Marschner S, Avanzo M, Gagliardi V, Fanetti G, et al. Distant metastasis time to event analysis with CNNs in independent head and neck cancer cohorts. *Sci Rep* 2021;11:1–12. <https://doi.org/10.1038/s41598-021-85671-y>.
 - [36] Cusumano D et al. Artificial intelligence in magnetic resonance guided radiotherapy: Medical and physical considerations on state of art and future perspectives. *Phys Med* 2021;85:175–91. <https://doi.org/10.1016/j.ejmp.2021.05.010>.
 - [37] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9. <https://doi.org/10.1162/neco.1997.9.8.1735>.
 - [38] Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv Neural Inform Process Syst* 2015;28. <https://doi.org/10.48550/arXiv.1506.04214>.
 - [39] Green O et al. First clinical implementation of real-time, real anatomy tracking and radiation beam control. *Med Phys* 2018;45. <https://doi.org/10.1002/mp.13002>.
 - [40] Kluter S. Technical design and concept of a 0.35 T MR-linac. *Clin Transl Radiat Oncol* 2019;18:98–101. <https://doi.org/10.1016/j.ctro.2019.04.007>.
 - [41] Zou K et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol* 2004;11:178–89. [https://doi.org/10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8).
 - [42] Huttenlocher D, Klanderman G, Rucklidge W. Comparing images using the hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 1993;15:850–63. <https://doi.org/10.1109/34.232073>.
 - [43] 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. <https://doi.org/10.1080/01621459.1937.10503522>.
 - [44] Nemenyi PB. Distribution-free multiple comparisons. Princeton University; 1963. PhD thesis.
 - [45] Fayad HJ et al. A generic respiratory motion model based on 4D MRI imaging and 2D image navigators. *IEEE Nuclear Science Symp. and Medical Imaging Conf*, 2012.
 - [46] Stemkens B et al. Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy. *Phys Med Biol* 2016;61:5335. <https://doi.org/10.1088/0031-9155/61/14/5335>.
 - [47] Paganelli C 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. <https://doi.org/10.1111/1754-9485.12713>.
 - [48] Rabe 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. <https://doi.org/10.1088/1361-6560/abc937>.
 - [49] 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. <https://doi.org/10.1088/1361-6560/aac5b9>.
 - [50] Nie X, Rimmer A, Li G. Feasibility of MR-guided radiotherapy using beam-eye-view 2D-cine with tumor-volume projection. *Phys Med Biol* 2021;66:045020. <https://doi.org/10.1088/1361-6560/abd66a>.

Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy

Elia Lombardo^a, Moritz Rabe^a, Yuqing Xiong^a, Lukas Nierer^a, Davide Cusumano^b, Lorenzo Placidi^b, Luca Boldrini^b, Stefanie Corradini^a, Maximilian Niyazi^a, Michael Reiner^a, Claus Belka^{a,c}, Christopher Kurz^a, Marco Riboldi^d, Guillaume Landry^a

^a*Department of Radiation Oncology, University Hospital, LMU Munich, Munich, 81377, Germany*

^b*Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, 00168, Italy*

^c*German Cancer Consortium (DKTK), Munich, 81377, Germany*

^d*Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Garching b. München, 85748, Germany*

Supplementary material

Data pre-processing

Using in-house software [1, 2], we extracted SI/AP center of mass motion curves (for the LSTM-shift) and binary masks (for the convolutional LSTMs) from the cine MRI videos. We then performed outlier replacement and applied a moving average filter (as described previously [1]) for both the SI and the AP motion curves to correct for wrongly extracted centroid positions. We did not exclude any cine video based on small motion amplitude and we changed the normalization scheme of the centroid motion trajectories: instead of normalizing based on the minimum and maximum amplitude of the entire individual cine video, we normalized each input window between -1 and +1 separately, based on the minimum and maximum amplitude of the window. This normalization is fast (≈ 1 ms), so it can be applied in real-time. We resampled each mask to a 1×1 mm² grid and cropped each image to 256×256 based on the center of the gating boundary contour.

To avoid data corrupted by unsuccessful extraction of binary masks, we computed the dice similarity coefficient (DSC) between successive masks and if the DSC was smaller than 0.4 for more than eight frames in total, we excluded that whole cine video. For the LMU data, we detected the

breath-holds based on the median deviation from the median centroid position within 20 data points and saved this information to disk to be able to exclude or include the breath-holds during model optimization. As in our previous study, we also detected the start of cine MRI pauses (originating from gantry rotations) and used this information to split the centroid and contour data into subsets. This avoids jumps in the data arising from the temporal discontinuities in the cine videos.

LSTM-shift model

The first model is an extension of the LSTM we implemented in a previous study [1]. Specifically, the input was changed from being a window with SI centroid positions to a window of both SI and AP centroid positions. As before, the input window is fed sequentially to LSTM modules which encode the motion over a variable number of hidden layers. The output of the last LSTM module in the final hidden layer is input to a fully connected layer, for which we doubled the number of neurons, such that the predicted output window contained both the SI and the AP centroid coordinates. In this fashion, however, the network would only allow for rigid MLC-tracking in SI and AP direction. Therefore, following an idea by Bourque et al. [3], we combined the network output with the binary mask corresponding to the last input frame (thus the last contour provided by the tracking algorithm) and then shifted this mask rigidly by the difference between the predicted centroid coordinates at 500 ms and the last centroid coordinates of the input window. Shifting is performed using nearest neighbour interpolation. This way, deformable MLC-tracking taking into account deformations/rotations up to the last available input contour is possible.

In this study, we extended the *offline+online* LSTM, a network which can be continuously re-optimized on current patient data, because we had found it to perform best in Lombardo et al. [1] and labelled it online LSTM-shift. However, we also extended an *offline* LSTM[1] to investigate which impact the continuous re-optimization has in terms of final model performance and labelled it *offline* LSTM-shift.

ConvLSTM model

The convolutional LSTM module was introduced by Shi et al. [4] to be able to capture both temporal and spatial correlations in the input. Compared to the classical LSTM module, matrix multiplications between the

hidden and cell states are replaced by convolutions. At a specific time step t , the convolutional LSTM module is described as follows:

$$\text{Forget gate: } \mathbf{f}^t = \sigma(\mathbf{W}_f * \mathbf{x}^t + \mathbf{U}_f * \mathbf{h}^{t-1} + \mathbf{b}_f) \quad (\text{S1})$$

$$\text{Input gate: } \mathbf{i}^t = \sigma(\mathbf{W}_i * \mathbf{x}^t + \mathbf{U}_i * \mathbf{h}^{t-1} + \mathbf{b}_i) \quad (\text{S2})$$

$$\text{New memory cell state: } \tilde{\mathbf{c}}^t = \tanh(\mathbf{W}_c * \mathbf{x}^t + \mathbf{U}_c * \mathbf{h}^{t-1} + \mathbf{b}_c) \quad (\text{S3})$$

$$\text{Final memory cell state: } \mathbf{c}^t = \mathbf{f}^t \odot \mathbf{c}^{t-1} + \mathbf{i}^t \odot \tilde{\mathbf{c}}^t \quad (\text{S4})$$

$$\text{Output gate: } \mathbf{o}^t = \sigma(\mathbf{W}_o * \mathbf{x}^t + \mathbf{U}_o * \mathbf{h}^{t-1} + \mathbf{b}_o) \quad (\text{S5})$$

$$\text{Hidden state: } \mathbf{h}^t = \mathbf{o}^t \odot \tanh(\mathbf{c}^t) \quad (\text{S6})$$

where $\mathbf{b}$, $\mathbf{W}$ and $\mathbf{U}$ denote the biases, input window convolutional kernels and recurrent convolutional kernels which are learned during the optimization process. The symbol $*$ represents convolution while the symbol $\odot$ represents element-wise multiplication between matrices. As for the LSTM module, the sigmoid function $\sigma(x)$ was used for the gates and the hyperbolic tangent function was used to generate the states.

The second model leverages convolutional LSTM modules to directly generate future 2D masks following the recurrent encoder-decoder architecture [4, 5] illustrated in Figure S1. The purpose of the implemented spatial encoder is to extract high-level features from the input masks. These are then fed to the convolutional LSTM encoder modules which learn the spatio-temporal characteristics of the motion and extrapolate them in time via the decoder modules. A spatial decoder is then used to reconstruct this information to the original contour dimensionality, thus outputting the predicted binary masks. The spatial encoder branch comprises three convolutional blocks, each formed by a convolutional layer, a ReLU activation function and a batch normalization layer [6]. The convolutional layer of the first block has a kernel size of five, stride of two and 32 output channels. The convolutional layer of the second and third blocks both have a kernel size of three, 64 output channels and a stride of one and two, respectively. The convolutional LSTM encoder modules have a kernel size of five, 64 output channels and use a hidden and cell state initialization with zeros. The convolutional LSTM decoder modules have kernel size of five and 64 output channels. The initial hidden and cell states are taken over from the last encoder module and represent the encoded motion. While we did not use the output of the LSTM encoder modules, for the LSTM decoders we fed it to the subsequent spatial decoder branch and to the next convolutional LSTM module. It has been shown

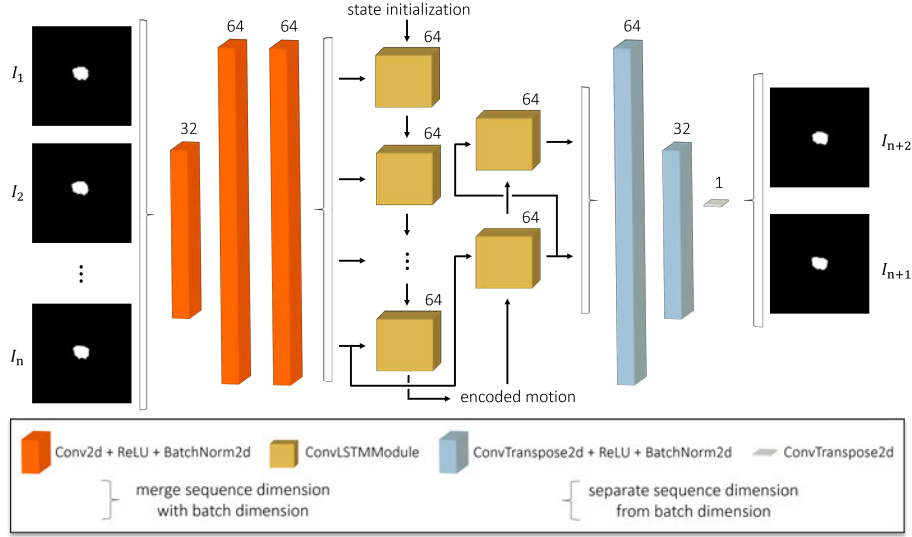


Figure S1: At a specific time step t , the ConvLSTM model takes a window of fixed length x_t containing binary mask I_1 to I_n as input and outputs the predicted mask at 250 ms and 500 ms. Each input sequence is processed by convolutional blocks (orange) to extract high-level features. The convolutional LSTM modules (yellow) are used to learn the spatio-temporal correlations and extrapolate the motion in time using a sequence to sequence arrangement of the modules. A spatial decoder is used to reconstruct the predicted mask from the encoded information (blue). The number of channels is displayed for each feature block. The curly brackets indicate that the sequence dimension is merged with or separated from the batch dimension using tensor reshaping.

that conditioning the modules on previously generated outputs improves the predictive performance [7]. The spatial decoder branch comprises two blocks with a transposed convolution (filter size of three, stride of two and one and 64 and 32 output channels) followed by ReLU and batch normalization and a block with only a transposed convolution with one output channel (filter size of three and stride of two) to obtain the final 2D masks.

ConvLSTM-STL model

The third model is an implementation of the recurrent encoder-decoder introduced by Romaguera et al [8]. As for ConvLSTM, it takes a fixed window of 2D binary masks as input and comprises encoding and decoding

convolutional and convolutional LSTM blocks. However, instead of learning to extrapolate masks in time, it learns to predict a sequence of 2D deformation vector fields which are used to warp the last input mask to obtain the future 2D masks. To be able to train the model in an end-to-end fashion, the warping operation is performed using the differentiable grid sampler present in spatial transformer networks [9]. In contrast to original spatial transformer networks, which are convolutional neural networks on their own, capable of learning transformations, Romaguera et al. inferred the transformations from the last convolutional decoder and used the differentiable grid sampling (which has no learnable weights) to warp the images. For this reason, we prefer to use the term spatial transformer layer (STL), instead of network. In this work, we implemented a ConvLSTM-STL without multi-scale residual block and skip connections, similar to the model developed in a follow-up publication by Romaguera et al. [10]. This model has less parameters and its lack of skip-connections enables a sequence-to-sequence implementation of the convolutional LSTM, which allows for a number of outputs independent of the number of inputs.

No-predictor model

To investigate the performance achieved if no motion predictor was used, we took for each fixed input window the last available contour and compared it with the true contour after two frames, i.e., after 500 ms. The same pre-processed testing datasets were used as for the three previous models to enable a fair comparison.

Hyper-parameters

In the following, we describe for each model the parameters used or if a parameter search was performed, the range of the search.

- LSTM-shift: we mainly used the parameters found in our previous study[1], i.e., number of hidden layers equal 5, no dropout, batch size equal 128 and $L2$ regularization weight of 0 for the offline training of the model (*offline* LSTM-shift). The learning rate of the offline training was varied around 5×10^{-4} , i.e., the best value found in our previous study. Specifically, we sampled from the set $\{1 \times 10^{-3}, 5 \times 10^{-4}, 3 \times 10^{-4}, 1 \times 10^{-4}\}$. The model was trained for 500 epochs with an early stopping patience of 200 (i.e., if the model did not improve after 200 epochs we stopped training and used the best performing model

for evaluation) when using the training set with breath-holds and for 2500 epochs with a patience of 1000 when training without the breath-holds, to make sure that about the same number of model updates were performed for the two scenarios presenting different amounts of data. For online re-training (*online* LSTM-shift), we used a learning rate of 1×10^{-6} and re-optimized the model at each time step for about 150 ms, as in our previous study. No data augmentations were used. The model was trained using the mean squared error (MSE) loss between the predicted and the true output centroid coordinates for both the 250 ms and the 500 ms future time point.

- ConvLSTM: due to memory limitations we trained using a batch size of 32 for a total of 50 epochs with an early stopping patience of 20 epochs for the breath-hold inclusion scenario and for a total of 250 epochs with a patience of 100 when training without breath-holds. The learning rate was chosen among the following values $\{1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}\}$. We used no dropout and an $L2$ regularization weight of 0 but, considering the higher number of parameters of the ConvLSTM compared to the LSTM-shift, we compared scenarios with and without the use of data augmentations. Specifically we used random affine transformations with a probability of 50% which included rotations up to 60° , shearing up to 0.5 and scaling up to 0.2. The augmentations were applied in the same way on both all input and output contours. The model was trained using either the dice loss [11] between the predicted and true output contours (250 ms and 500 ms future time points) or the focal loss [12]. Additionally, we trained the model using both the binary masks and the corresponding cine MRI frames as input. Online re-training in the same fashion as for the LSTM-shift is currently not possible with convolutional LSTMs as one online epoch was found to take more than 250 ms, which is the time at disposal before the next cine MRI frame is available.
- ConvLSTM-STL: due to stronger memory limitations we trained using a batch size of 3 for a total of 5 epochs with an early stopping patience of 2 epochs for the breath-hold inclusion scenario and for a total of 25 epochs with a patience of 10 when training without breath-holds. It should be noted that even though the number of epochs is considerably smaller, the total number of model updates was comparable with the

other models as the batch size was smaller. The learning rate, $L2$ regularization weight, dropout and augmentations were chosen as for the ConvLSTM. The model was trained using either dice loss alone or combined with a smoothing loss based on the gradient of the predicted deformation vector fields as in Romaguera et al.[8]. Additionally, we trained the model using the dice loss of the contour combined with the MSE loss between centroids computed from the ground truth and from the predicted contours. Again, we took both the 250 ms and the 500 ms forecast in consideration for the loss computation. As for the ConvLSTM, we additionally trained this model using both the binary masks and the corresponding cine MRI frames as input.

All models were optimized with the Adam algorithm [13] and we used an input window length of 32 (8 s of data) for the LSTM-shift, as we had found this length to lead to the best performance [1]. For the ConvLSTM and ConvLSTM-STL we trained with an input window length of 8, 16 and 32 and found the latter to lead to the best validation results, so only 32 was used for the rest of the study. When performing validation, we found training on data without breath-holds to slightly increase the performance for all but the ConvLSTM model, for which it slightly decreased. Specifically, we found an average improvement over all metrics of 1.2 % for the LSTM-shift, of 0.3 % for the ConvLSTM-STL and an average decline of 9.1 % for the ConvLSTM. For the *offline* LSTM-shift, we found a learning rate of 5×10^{-4} to perform best, as in our previous study. Both for the ConvLSTM and for ConvLSTM-STL we found a learning rate of 1×10^{-4} , the usage of augmentations and training with dice loss to achieve the best validation performance. Using the cine MRI frames as additional input to the binary masks led to a small decrease in performance for both convolutional LSTMs. For the ConvLSTM-STL we tried to put more focus on the alignment of the centroids by using a combined dice and MSE centroid loss, however, the performance did not improve.

Implementation details

The code for this study was written in Python 3.8.13 and is publicly available: https://github.com/LMUK-RADONC-PHYS-RES/lstm_contour_prediction. All models were optimized and evaluated using the PyTorch based framework MONAI 0.9.0 [14]. Both the optimization and the evaluation of the models were carried out on an NVIDIA RTX A6000 GPU with 48 GB of memory.

Supplementary file 1

Table S1: Tumor volumes and inter-quartile range (IQR) of motion across the datasets.

	LMU (training)	LMU (validation)	LMU (testing)	Gemelli
Min. volume [cm ³]	0.9	0.7	0.6	4.2
Max. volume [cm ³]	326.8	165.4	205.8	65.4
Mean volume [cm ³]	31.5	29.8	52.1	34.4
Std volume [cm ³]	52.7	41.6	68.9	30.6
Mean IQR _{SI} [mm]	7.1	7.6	8.5	4.3
Mean IQR _{AP} [mm]	3.5	3.8	3.5	1.5

Supplementary file 2

Table S2: Mean and standard deviation of different metrics for the 500 ms forecasted masks for the validation set when repeating the optimization to assess model stability (compare to Table 1 in main text).

Model	HD _{50%} [mm]	HD _{95%} [mm]	DSC	RMSE _{SI} [mm]	RMSE _{AP} [mm]
<i>offline</i> LSTM-shift	1.1 ± 0.4	2.6 ± 1.0	0.91 ± 0.05	1.2 ± 0.7	0.7 ± 0.3
<i>online</i> LSTM-shift	1.1 ± 0.4	2.6 ± 1.0	0.91 ± 0.04	1.2 ± 0.7	0.7 ± 0.3
ConvLSTM	1.3 ± 0.7	3.2 ± 1.5	0.89 ± 0.06	1.7 ± 1.0	0.8 ± 0.4
ConvLSTM-STL	1.3 ± 0.6	3.2 ± 1.6	0.89 ± 0.06	1.6 ± 1.0	0.8 ± 0.4

Supplementary file 3

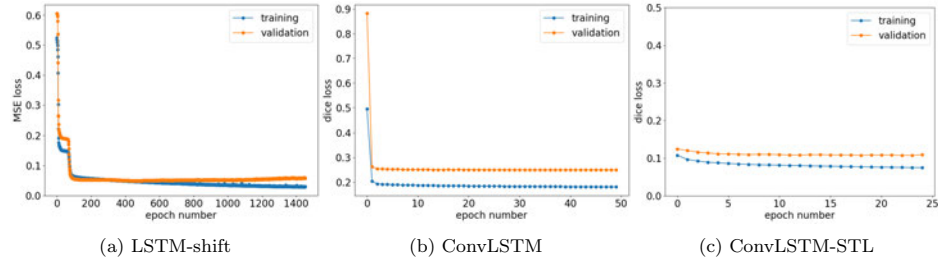


Figure S2: Training and validation loss for the different models showing convergence of the optimization process. The best model for the LSTM-shift was found at epoch 452, for the ConvLSTM at epoch 46 and for the ConvLSTM-STL at epoch 22.

Supplementary file 4

Table S3: P-values obtained from the post-hoc Nemenyi test for the indicated testing set for all possible pairwise model comparisons. Significant p-values (< 0.05) are denoted with an asterisk.

Comparison		LMU	Gemelli
Model 1	Model 2	p-value	
<i>offline</i> LSTM-shift	<i>online</i> LSTM-shift	0.08	0.001*
<i>offline</i> LSTM-shift	ConvLSTM	0.001*	0.001*
<i>offline</i> LSTM-shift	ConvLSTM-STL	0.001*	0.14
<i>offline</i> LSTM-shift	No-predictor	0.001*	0.001*
<i>online</i> LSTM-shift	ConvLSTM	0.001*	0.001*
<i>online</i> LSTM-shift	ConvLSTM-STL	0.001*	0.001*
<i>online</i> LSTM-shift	No-predictor	0.001*	0.001*
ConvLSTM	ConvLSTM-STL	0.90	0.001*
ConvLSTM	No-predictor	0.001*	0.005*
ConvLSTM-STL	No-predictor	0.001*	0.001*

References

- [1] E. Lombardo, et al., Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy, *Phys. Med. Biol.* 67 (2022) 095006. URL: <https://doi.org/10.1088/1361-6560/ac60b7>. doi:10.1088/1361-6560/ac60b7.
- [2] Y. Xiong, M. Rabe, L. Nierer, M. Kawula, S. Corradini, C. Belka, M. Riboldi, G. Landry, C. Kurz, Assessment of intrafractional prostate motion and its dosimetric impact in MRI-guided online adaptive radiotherapy with gating, *Strahlentherapie und Onkologie* (2022) 1–10. URL: <https://doi.org/10.1007/s00066-022-02005-1>.
- [3] A. E. Bourque, J.-F. Carrier, É. Filion, S. Bedwani, A particle filter motion prediction algorithm based on an autoregressive model for real-time MRI-guided radiotherapy of lung cancer, *Biomedical Physics & Engineering Express* 3 (2017) 035001. doi:10.1088/2057-1976/aa6b5b.
- [4] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo, Convolutional LSTM network: A machine learning approach for precipitation nowcasting, *Advances in neural information processing systems* 28 (2015). URL: <https://doi.org/10.48550/arXiv.1506.04214>.
- [5] B. Sautermeister, Deep learning approaches to predict future frames in videos, Master Thesis. Department of Informatics. Technische Universität München (2016) 129. URL: https://bsautermeister.de/research/docs/msc_thesis.pdf.
- [6] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: *International conference on machine learning*, PMLR, 2015, pp. 448–456. URL: <https://doi.org/10.48550/arXiv.1502.03167>.
- [7] S. Bengio, O. Vinyals, N. Jaitly, N. Shazeer, Scheduled sampling for sequence prediction with recurrent neural networks, *Advances in neural information processing systems* 28 (2015). URL: <https://doi.org/10.48550/arXiv.1506.03099>.
- [8] L. V. Romaguera, R. Plantefève, F. P. Romero, F. Hébert, J.-F. Carrier, S. Kadoury, Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks,

- Medical image analysis 64 (2020) 101754. doi:10.1016/j.media.2020.101754.
- [9] M. Jaderberg, K. Simonyan, A. Zisserman, et al., Spatial transformer networks, *Advances in neural information processing systems* 28 (2015). URL: <https://doi.org/10.48550/arXiv.1506.02025>.
- [10] L. V. Romaguera, T. Mezheritsky, R. Mansour, J.-F. Carrier, S. Kadoury, Probabilistic 4D predictive model from in-room surrogates using conditional generative networks for image-guided radiotherapy, *Medical Image Analysis* 74 (2021) 102250. doi:<https://doi.org/10.1016/j.media.2021.102250>.
- [11] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: *2016 fourth international conference on 3D vision (3DV)*, IEEE, 2016, pp. 565–571. URL: <https://doi.org/10.48550/arXiv.1606.04797>.
- [12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988. URL: <https://doi.org/10.48550/arXiv.1708.02002>.
- [13] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: *3rd International Conference for Learning Representations*, 2015.
- [14] T. M. Consortium, Project monai (2020). URL: <https://monai.io/>.

2.3 Paper III: Experimental comparison of linear regression and LSTM motion prediction models for MLC-tracking on an MRI-linac

During a two-month research stay at the University of Sydney, the LSTM and linear regression models from Paper I were experimentally validated at the Australian MRI-linac, which resulted in the following article. The models' code was adapted to work with a real-time stream of data and developed in a standalone software environment (container) to facilitate the integration with the existing setup. By performing repeated MLC-tracking experiments using an MRI-compatible phantom driven by eight unseen patient respiratory motion traces (lung cancers), it was shown that both the LSTM and the linear regression successfully compensated for the measured system latency of 390 ms. The continuously re-optimized LSTM reduced the geometric MLC-tracking error significantly more than the population-based LSTM and the continuously updated linear regression model, confirming the previous in-silico results and convincingly demonstrating the real-time applicability of the developed models.

Reprinted with permission from "Lombardo, E., Liu, P. Z., Waddington, D. E., Grover, J., Whelan, B., Wong, E., ... & Keall, P. J. (2023) Experimental comparison of linear regression and LSTM motion prediction models for MLC-tracking on an MRI-linac. *Medical Physics*, 50(11), 7083-7092".

Received: 9 May 2023 | Revised: 30 August 2023 | Accepted: 17 September 2023

DOI: 10.1002/mp.16770

RESEARCH ARTICLE

MEDICAL PHYSICS

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

Elia Lombardo¹ | Paul Z. Y. Liu^{2,3} | David E. J. Waddington^{2,3} | James Grover^{2,3} |
 Brendan Whelan^{2,3} | Esther Wong³ | Michael Reiner¹ | Stefanie Corradini¹ |
 Claus Belka^{1,4,5} | Marco Riboldi⁶ | Christopher Kurz¹ | Guillaume Landry¹ |
 Paul J. Keall^{2,3}

¹Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, Germany

²Image X Institute, University of Sydney Central Clinical School, Sydney, New South Wales, Australia

³Department of Medical Physics, Ingham Institute of Applied Medical Research, Liverpool, New South Wales, Australia

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

⁵Bavarian Cancer Research Center (BZKF), Munich, Germany

⁶Department 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

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 conformality 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 ± 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 ± 1.3) mm, compared to the (3.2 ± 1.9) mm of the *offline* LSTM, the (3.3 ± 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 ± 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.^{1,2} MRI-linac systems, which are linear accelerators with an embedded MRI unit, are increasingly being used clinically over the past years.³ To adapt to intra-fractional changes, for instance due to respiratory motion, current clinical systems rely on motion monitoring⁴ with gated beam delivery.⁵ 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 being reported in clinics, depending on patient compliance.^{6,7} 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.⁸ However, it has been shown that a factor which is critical to the accuracy of MRIGRT with MLC-tracking is the system latency.⁹

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 ≤ 500 ms is necessary to meet the definition of real-time motion compensation.¹⁰ For MRI-linacs capable of MLC-tracking it has been experimentally measured to range from 205 ms to 411 ms,^{9,11} 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^{12–14} or RT in general.^{15–17} 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.¹⁷ 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^{18,19} and in MRIGRT.^{20,21} 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.²⁰

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.²², 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.²³ 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 plans²⁴ and with a hybrid 2D/4D-MRI methodology.²⁵ 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 models²⁰ for MLC-tracking with a prototype MRI-linac. Compared to previous MRIGRT studies with MLC-tracking and motion prediction,^{22–25} 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).²⁶ 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.⁹) 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.⁹: 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.²⁷ 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.²⁸ 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). 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.²⁹ 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.,²⁷ 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 rescaled 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

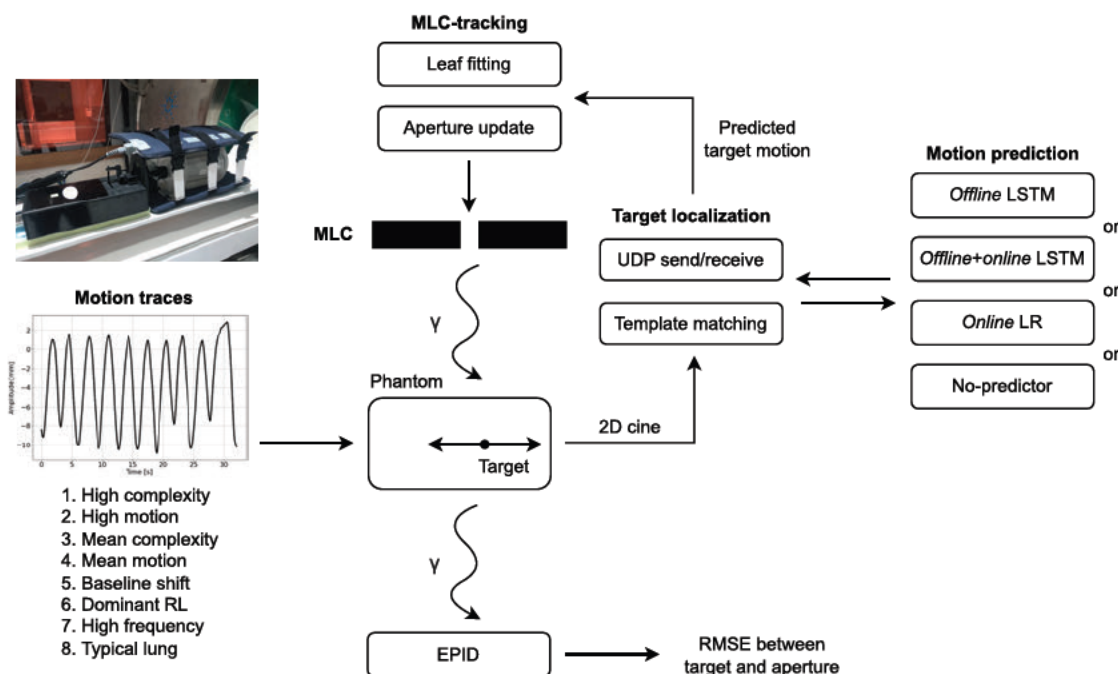


FIGURE 1 Experimental setup for MRI-guided MLC-tracking with motion prediction. A target inside a 1D motion phantom (photo on the top left) is moved following one of the eight patient respiratory motion traces. The moving target is imaged at 4 Hz using cine MRI. The target centroid position is then extracted 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.²⁷ 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 mm³. Using these sequence parameters, an imaging frequency of 4 Hz was obtained. As

in Liu et al.,⁹ 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.

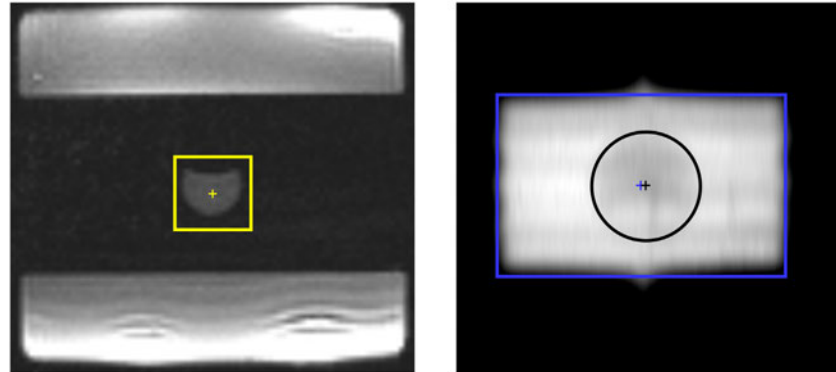


FIGURE 2 (Left) Cine MRI frame acquired during irradiation with real-time target localization (yellow) using template matching. (Right) EPID frame acquired during irradiation with post-irradiation template matching analysis to obtain the target centroid positions (black) and the MLC aperture centroid 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-silico²⁰ 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).²⁰ 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.²⁰ 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.³⁰ 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.²⁰ 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.³¹ 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.^{8,32}

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.⁹ 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.³³ If the Friedman test was significant (p -value < 0.05), a post-hoc Nemenyi test³⁴ 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 ± 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 ± 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 ± 1.0) mm, the *offline+online* LSTM to (2.8 ± 0.7) mm, the *online* LR to (3.3 ± 0.7) mm and the no-predictor to (4.5 ± 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	1.9; 2.0	2.6; 2.9	3.2; 3.4
High motion	3.5; 3.8	3.0; 3.0	3.6; 4.2	4.8; 4.7
Mean complexity	2.3; 2.4	2.0; 2.0	2.8; 2.3	3.2; 3.3
Mean motion	5.1; 4.4	3.6; 3.7	4.4; 4.1	6.3; 6.2
Baseline shift	2.2; 2.2	2.1; 2.1	2.3; 2.4	2.9; 3.0
Dominant RL	3.4; 3.5	2.8; 3.1	4.0; 3.3	4.9; 5.0
High frequency	4.8; 4.9	3.8; 3.6	4.2; 3.8	6.9; 7.0
Typical lung	3.0; 3.0	2.8; 2.7	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.

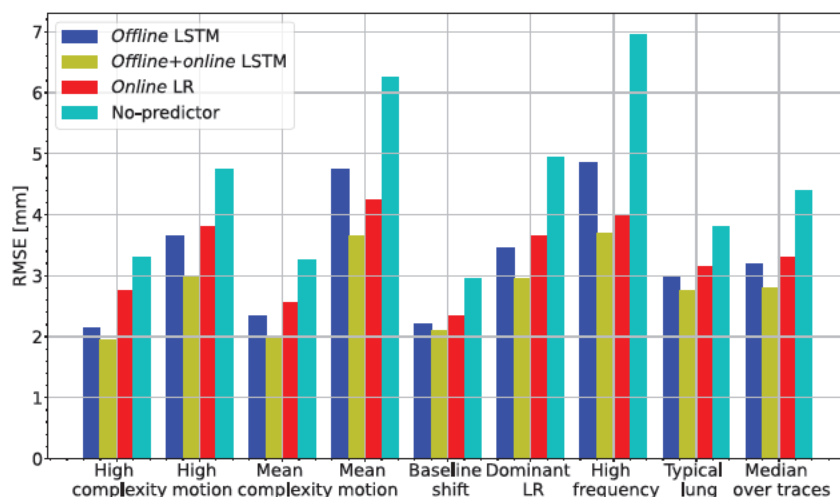


FIGURE 3 Plot with the mean (bar) and standard deviation (error bar) of the RMSE obtained from the two measurements for the four models and eight motion traces used in this study. Additionally, the mean and standard deviation of the RMSE over all traces for each model is shown. RMSE, root-mean-square 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 ± 15) ms, which is slightly increased compared to the latency of (328 ± 44) ms

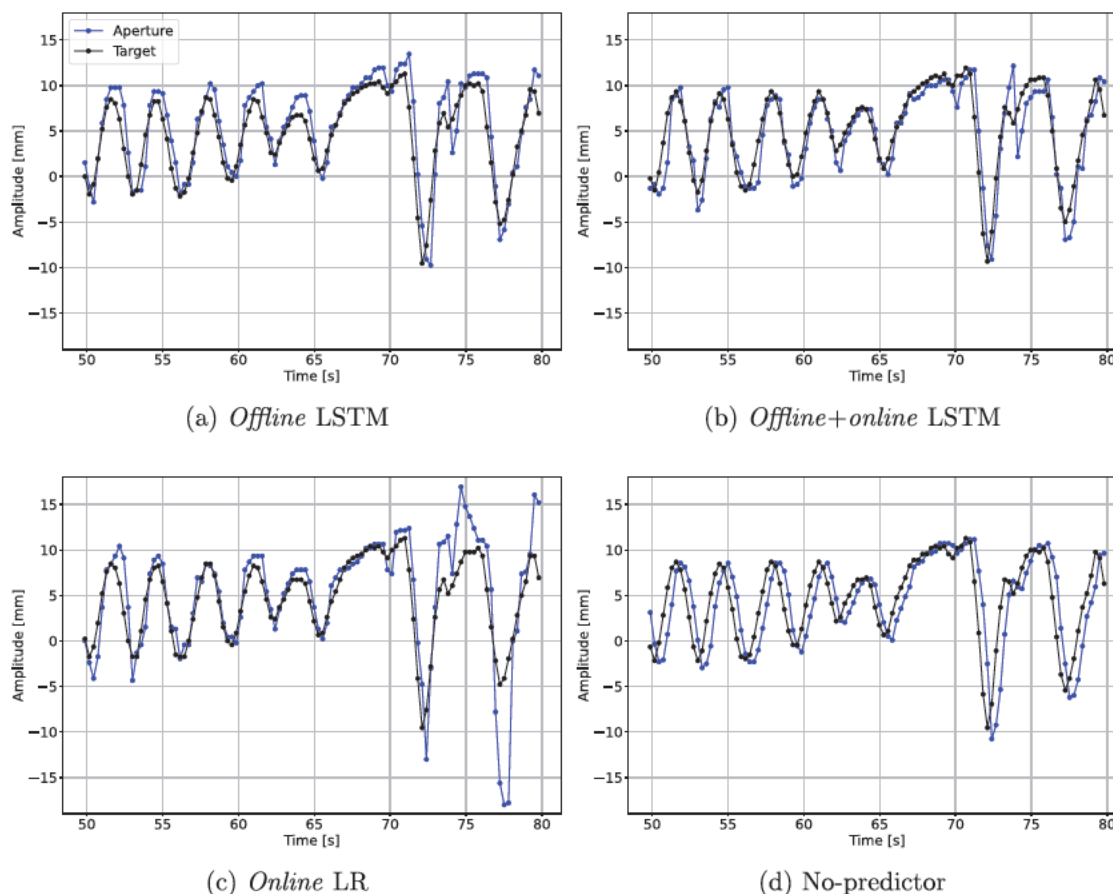


FIGURE 4 MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the “high complexity” motion trace for one of the two measurements. The MLC is moved according to one of the four motion prediction models. The same four models but for the other seven traces are shown in the online supplementary ([mp16770-sup-0001-SupMat.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,⁹ but still within what is expected for MLC-tracking on MRI-linacs in other studies.^{11,22} The latency corrected baseline RMSE of (1.1 ± 0.1) mm is comparable with the RMSE of (1.2 ± 0.1) found by Liu et al.⁹ 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.³⁵ 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.^{36,37} 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,³⁸ a 2D contour prediction algorithm²¹ and a deformable MLC-tracking algorithm.³⁹ Beam's eye view 2D cine MRI with tumor-volume projection might be used to ensure better beam conformality.^{40,41} To fully compensate motion in all three directions, time-resolved volumetric MRIs would be needed, which are currently being investigated by several groups^{42–44} 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, Andrasschke 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, Legendijk 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 E, 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. *Radiation 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. *Radiation Oncol*. 2022;174:149-157.
 25. Keijnemans 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. *Radiation 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's-eye-view 2D-cine with tumor-volume projection. *Phys Med Biol*. 2021;66:045020.
 42. Stemkens B, Tjissen RH, de Senneville BD, Legendijk 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>

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

E. Lombardo¹, P.Z.Y. Liu^{2,3}, D.E.J. Waddington^{2,3}, J. Grover^{2,3}, B. Whelan^{2,3}, E. Wong³, M. Reiner¹, S. Corradini¹, C. Belka^{1,4,5}, M. Riboldi⁶, C. Kurz¹, G. Landry^{1,+} & P.J. Keall^{2,3,a,+}

¹ Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, 81377, Germany

² Image X Institute, University of Sydney Central Clinical School, Sydney, NSW, 2015, Australia

³ Department of Medical Physics, Ingham Institute of Applied Medical Research, Liverpool, NSW, 2170, Australia

⁴ German Cancer Consortium (DKTK), partner site Munich, a partnership between DKFZ and LMU University Hospital Munich, 81377, Germany

⁵ Bavarian Cancer Research Center (BZKF), Munich, 81377, Germany

⁶ Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Garching b. München, 85748, Germany

^a Author to whom correspondence should be addressed: paul.keall@sydney.edu.au

⁺ Contributed equally as senior author

Supplementary material

Table S1: Hyper-parameters found in the previous in-silico study¹ and used without any changes in this work. A detailed explanation of why the batch size of the *offline+online* LSTM is fixed can be found there. The training with the *offline* LSTM took about one hour to converge (early stopping) using the described cohort and the *offline+online* LSTM was re-optimized for 10 epochs every time a new data point was made available, as it was found out that this takes about 150 ms and is therefore compatible with 4 Hz cine MRI¹.

Hyper-parameter	<i>offline</i> LSTM	<i>offline+online</i> LSTM	<i>online</i> LR
Number of layers	5	5	-
Dropout rate	0	0	-
Learning rate	5×10^{-4}	1×10^{-6}	-
Batch size	128	47 (fixed)	-
L2 weight	0	1×10^{-6}	1×10^{-4}
Input window length	32	32	8

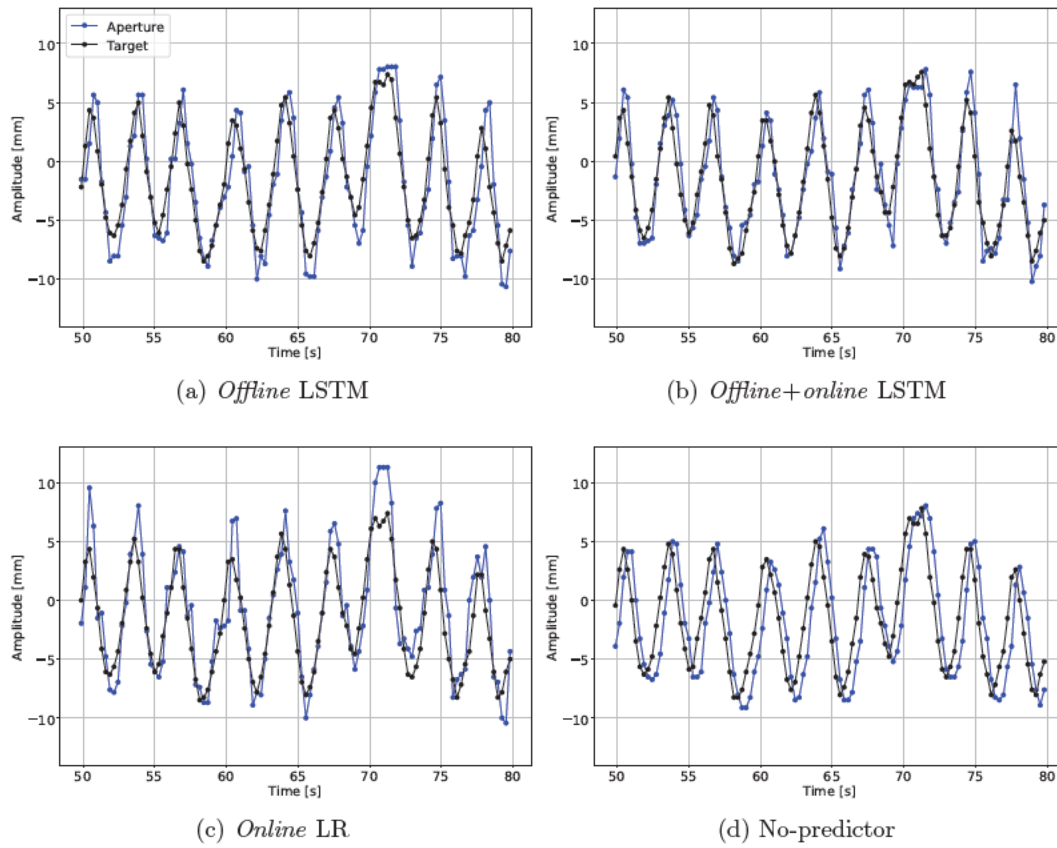


Figure S1: MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the "mean complexity" motion trace.

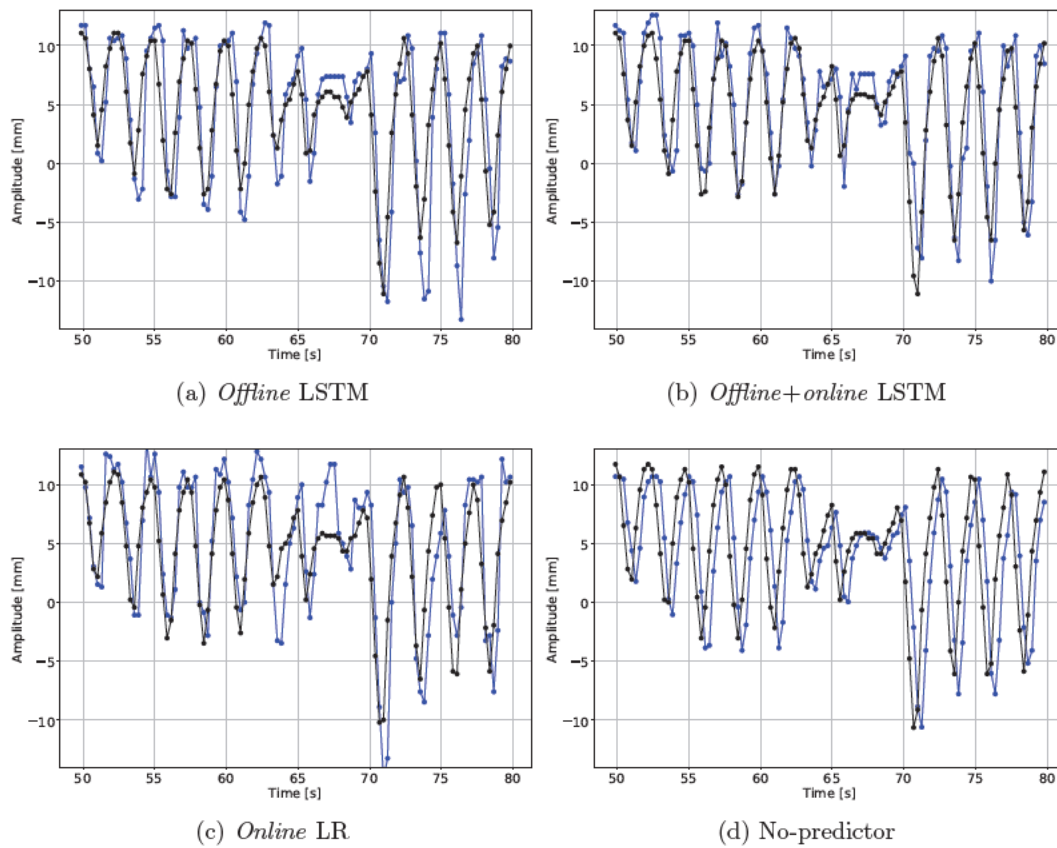


Figure S2: MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the "high motion" motion trace.

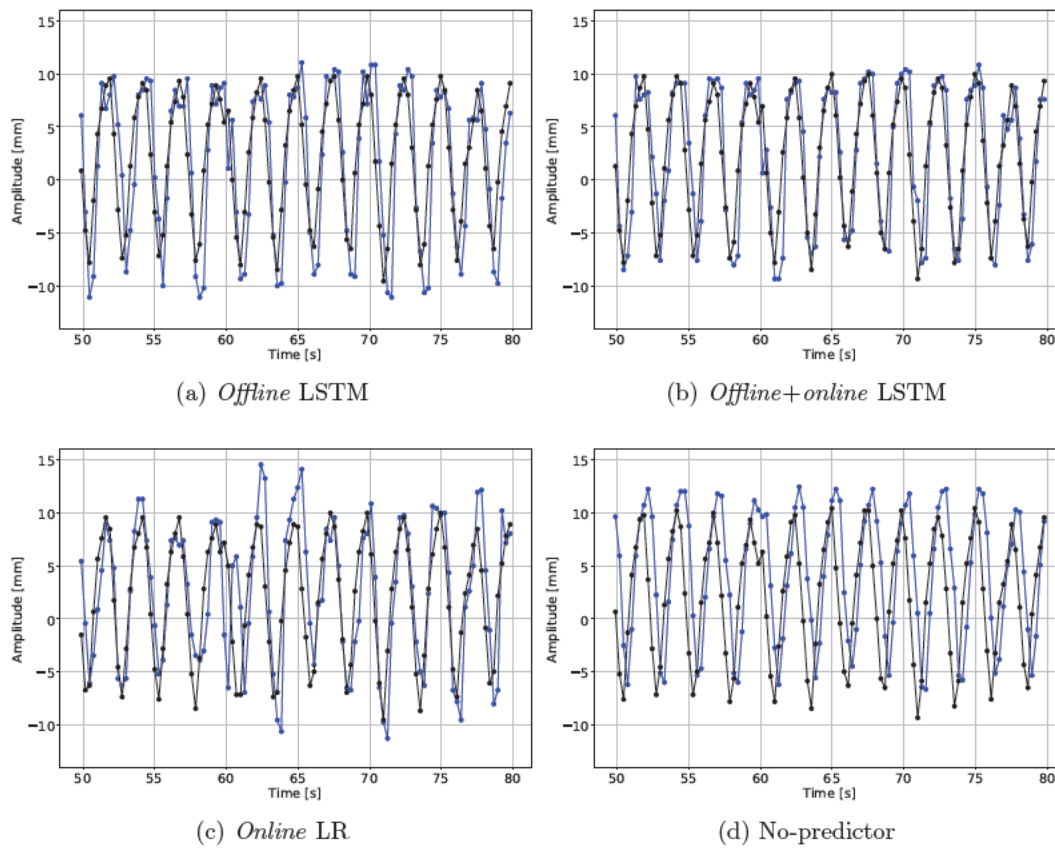


Figure S3: MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the "mean motion" motion trace.

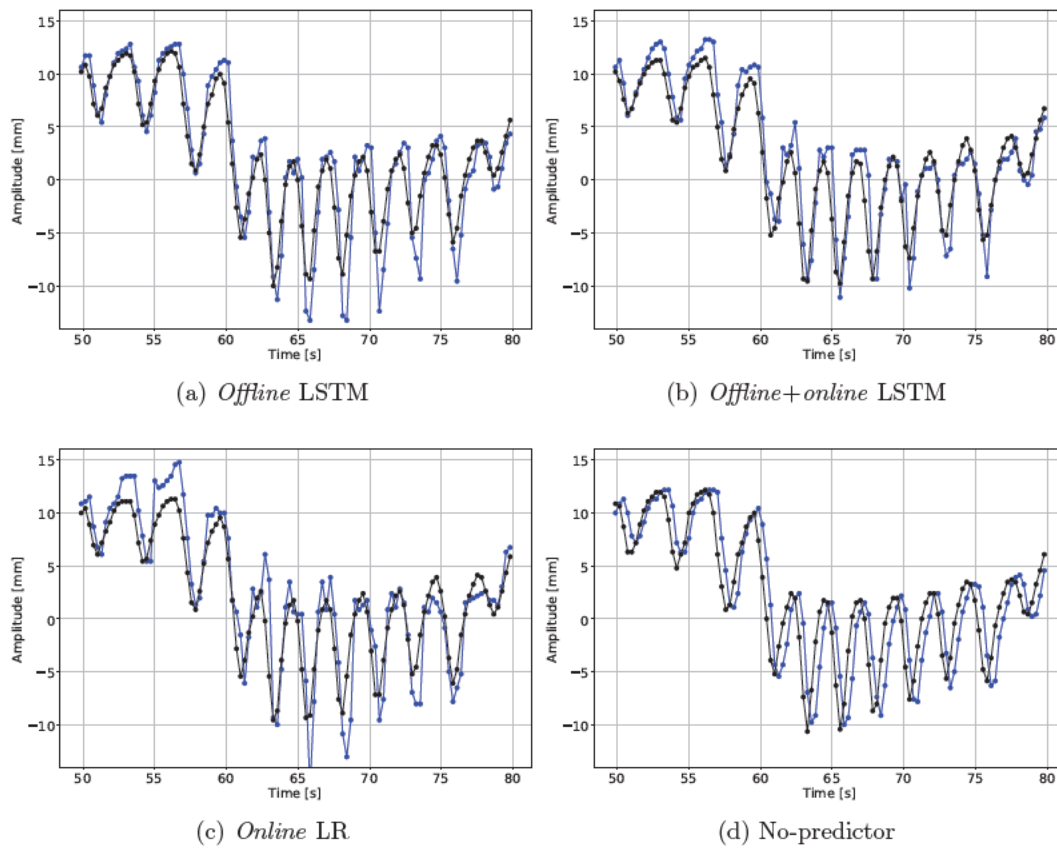


Figure S4: MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the "base-line shift" motion trace.

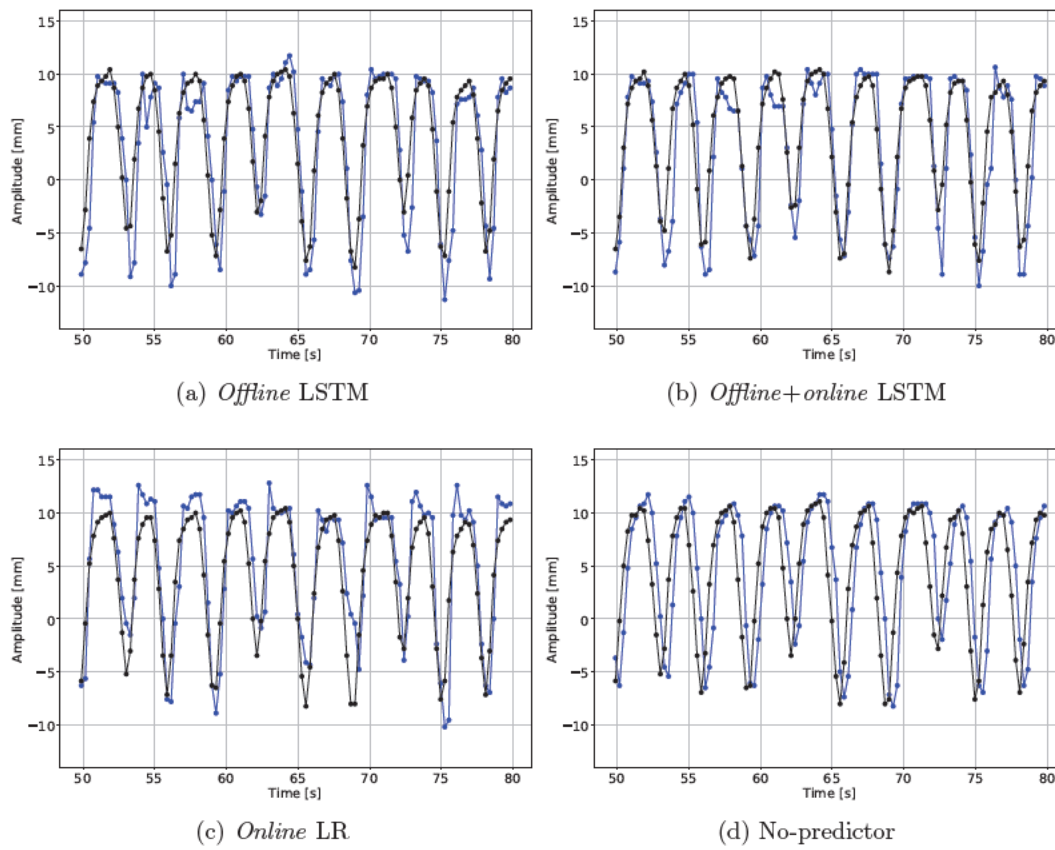


Figure S5: MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the "dominant RL" motion trace.

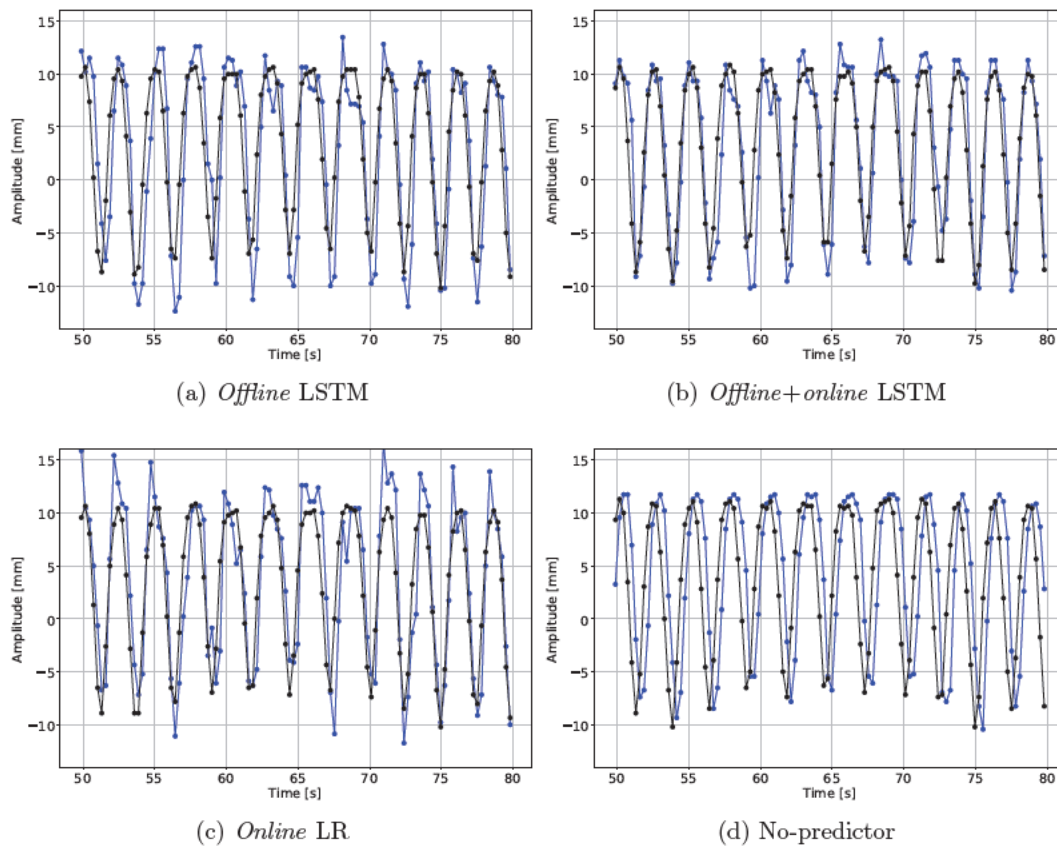


Figure S6: MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the "high frequency" motion trace.

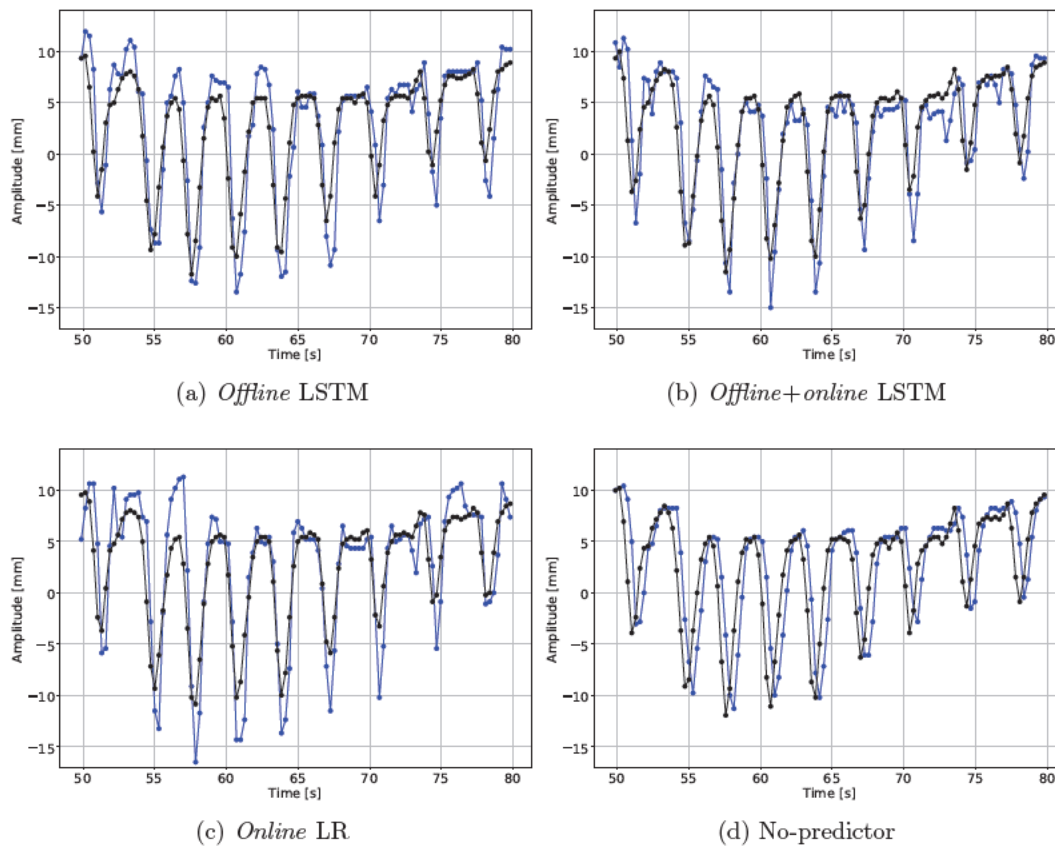


Figure S7: MLC aperture centroid (blue) and target centroid (black) obtained from the EPID for a selected part of the "typical lung" motion trace.

References

- ¹ E. Lombardo et al., Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy, *Phys. Med. Biol.* **67**, 095006 (2022).

2.4 Paper IV: Patient-specific deep learning tracking framework for real-time 2D target localization in magnetic resonance imaging-guided radiation therapy

In this study the focus was put on the step before motion prediction, i.e., on localizing the tumor/irradiation target on the current 2D cine MRI frame. This study was done last as it took the candidate some time to set up the extraction of the un-processed cine MRI data from the MRIdian, previous studies using post-processed cine MRI video data which could easily be exported from the system. Using unlabelled and manually labeled un-processed data from the MRIdian at the LMU University Hospital in Munich, different DL and conventional algorithms were compared in-silico. Specifically, a transformer DIR, a U-net auto-segmentation and a B-spline DIR acting as the slow baseline were implemented. The transformer was trained in an unsupervised (without labels), a supervised (with labels) and a patient-specific (with labels) fashion while for the U-net only patient-specific models were trained. The patient-specific transformer was found to outperform both the U-net and the B-spline without introducing a significant additional latency (inference time of ≈ 36 ms). The U-net was found to perform better than the best transformer only on a single testing set patient with a very large target exhibiting large motion, suggesting that a combination of the two models could be used to optimally localize the target for all investigated patients.

Reprinted with permission from "Lombardo, E., Velezmoro, L., Marschner, S.N., Rabe, M., Tejero, C., Papadopoulou, C.I., ... & Landry, G. (2025) Patient-specific deep learning tracking framework for real-time 2D target localization in magnetic resonance imaging-guided radiation therapy. *International Journal of Radiation Oncology*Biology*Physics*, 122(4), 827-837".

PHYSICS CONTRIBUTION

Patient-Specific Deep Learning Tracking Framework for Real-Time 2D Target Localization in Magnetic Resonance Imaging-Guided Radiation Therapy



Elia Lombardo, MSc,* Laura Velezmoro,* Sebastian N. Marschner, MD,* Moritz Rabe, PhD,* Claudia Tejero, MSc,*
Christianna I. Papadopoulou,* Zhuojie Sui, MSc,[†] Michael Reiner, PhD,* Stefanie Corradini, MD,*
Claus Belka, MD,*^{‡,§} Christopher Kurz, PhD,* Marco Riboldi, PhD,[‡] and Guillaume Landry, PhD*

*Department of Radiation Oncology, LMU University Hospital, LMU Munich; [†]Department of Medical Physics, Faculty of Physics, LMU Munich; [‡]German Cancer Consortium (DKTK), Partner Site Munich, a Partnership between DKFZ and LMU University Hospital Munich; and [§]Bavarian Cancer Research Center (BZKF), Partner Site Munich, Munich, Germany

Received Dec 9, 2023; Accepted for publication Oct 11, 2024

Purpose: We propose a tumor tracking framework for 2D cine magnetic resonance imaging (MRI) based on a pair of deep learning (DL) models relying on patient-specific (PS) training.

Methods and Materials: The chosen DL models are: (1) an image registration transformer and (2) an auto-segmentation convolutional neural network (CNN). We collected over 1,400,000 cine MRI frames from 219 patients treated on a 0.35 T MRI-linac plus 7500 frames from additional 35 patients that were manually labeled and subdivided into fine-tuning, validation, and testing sets. The transformer was first trained on the unlabeled data (without segmentations). We then continued training (with segmentations) either on the fine-tuning set or for PS models based on 8 randomly selected frames from the first 5 seconds of each patient's cine MRI. The PS auto-segmentation CNN was trained from scratch with the same 8 frames for each patient, without pre-training. Furthermore, we implemented B-spline image registration as a conventional model, as well as different baselines. Output segmentations of all models were compared on the testing set using the Dice similarity coefficient, the 50% and 95% Hausdorff distance ($HD_{50\%}/HD_{95\%}$), and the root-mean-square-error of the target centroid in superior-inferior direction.

Results: The PS transformer and CNN significantly outperformed all other models, achieving a median (interquartile range) dice similarity coefficient of 0.92 (0.03)/0.90 (0.04), $HD_{50\%}$ of 1.0 (0.1)/1.0 (0.4) mm, $HD_{95\%}$ of 3.1 (1.9)/3.8 (2.0) mm, and root-mean-square-error of the target centroid in superior-inferior direction of 0.7 (0.4)/0.9 (1.0) mm on the testing set. Their inference time was about 36/8 ms per frame and PS fine-tuning required 3 min for labeling and 8/4 min for training. The

Corresponding author: Elia Lombardo, MSc; E-mail: el.lombardo@med.uni-muenchen.de

Elia Lombardo and Laura Velezmoro made equal contributions to this study.

Disclosures: E.L.—German Research Foundation (DFG): support within research training group GRK 2274 “Advanced Medical Physics for Image-Guided Cancer Therapy.” M. Rabe—German Research Foundation (DFG): support within research training group GRK2274 “Advanced Medical Physics for Image-Guided Cancer Therapy.” Deutsche Krebshilfe: short-term fellowship unrelated to this study. Z.S.—China Scholarship Council: funded by the China Scholarship Council within the joint LMU-CSC program. S.C.—Brainlab: research grant; Brainlab, Viewray, and Elekta:

speaker honoraria and travel expenses. C.B.—ViewRay, Elekta, Brainlab, and C-RAD: research agreement (institutional); ViewRay, Elekta, Opasca, and C-RAD: honoraria outside of this work. The Department of Radiation Oncology of the LMU University Hospital has research agreements with Brainlab and Elekta. The study was supported by DFG, research training group GRK 2274, and Bavarian Cancer Research Center (BZKF).

Data Sharing Statement: Imaging data can be shared upon request. Code has been publicly shared.

Acknowledgments—Seyed-Ahmad Ahmadi is thanked for fruitful discussions and support with MONAI (Medical Open Network for AI) Label.

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.ijrobp.2024.10.021](https://doi.org/10.1016/j.ijrobp.2024.10.021).

transformer was better than the CNN in 9/12 patients, the CNN better in 1/12 patients, and the 2 PS models achieved the same performance on the remaining 2/12 testing patients.

Conclusions: For targets in the thorax, abdomen, and pelvis, we found 2 PS DL models to provide accurate real-time target localization during MRI-guided radiotherapy. © 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Introduction

Magnetic resonance imaging (MRI)-guided radiation therapy (MRIgRT) using integrated MRI-linacs is an established technique for high-precision stereotactic body radiation therapy of tumor sites affected by motion.¹⁻³ MRI-linacs allow intra-fractional motion management in real-time through a cine MRI-based gating approach.^{4,5} Cine MRIs in either a single 2D plane or interleaved orthogonal 2D planes are continuously acquired throughout the irradiation session, with the tumor or a surrogate tracking target localized on the current frame for beam gating. Although this method decreases target volume margins, it may extend the treatment duration, with duty cycle efficiencies of 20% and 51% for low-field MRI-linacs^{6,7} and 68% for a high-field MRI-linac.⁵ Multileaf collimator (MLC)-tracking in free-breathing is investigated as a dosimetrically equivalent yet more time efficient alternative to gating.⁸⁻¹⁰ For MLC-tracking, higher accuracy is required because the target must be precisely localized in every state of motion for accurate tracking.

Several target tracking algorithms utilizing cine MRI have been proposed. The first methods made use of rigid template matching.¹¹⁻¹³ Because template matching cannot provide deformation or rotation information, deformable image registration (DIR) or auto-segmentation methods can be used instead. In DIR-based algorithms, the deformation between the current and a reference frame is calculated, and then used to adapt the available reference segmentation. Alternatively, paired samples of cine MRI frames and corresponding manual segmentations are exploited to derive a model that learns how to segment the subsequent frames (auto-segmentation).^{14,15} More recently, approaches have focused on the use of deep learning (DL) methods such as convolutional neural networks (CNNs) for both DIR and auto-segmentation. Once trained, their execution takes little time, while showing superior performance compared with conventional methods.¹⁶⁻¹⁸ Friedrich et al¹⁶ demonstrated this potential by proposing a CNN for real-time auto-segmentation of liver tumors for both radial and Cartesian 2D cine MRIs from a 0.35 T MRI-linac. They introduced a novel supervised patient-specific (PS) training strategy, which trained CNNs from scratch using the initial 10 manually segmented frames of each patient. However, only 2 patients and a single tumor site were included in their study. More recently, Hunt et al¹⁸ presented a CNN for 2D cine MRI DIR, using 21 patients with liver, lung, or pancreas tumors. By framing the task as DIR, they trained the CNN in an unsupervised manner, avoiding the need for manual

segmentations. However, they solely evaluated registration accuracy on an image-level and did not assess tracking accuracy on segmentations.

For clinical use, a tracking approach must be reliable in a wide variety of tracking scenarios, covering targets of various sizes and contrasts. We thus propose to develop a pair of DL models with different underlying mechanisms and evaluate them on manually segmented ground truth data. Recently, the transformer DL architecture has become state-of-the-art in many fields, from natural language processing¹⁹ and computer vision²⁰ to medical imaging.²¹ As shown by Chen et al,²² one advantage of transformers compared with CNNs is their larger effective receptive field, which makes them capable of capturing longer-range spatial information to improve the performance in low-contrast settings. This is especially relevant for models based on image registration, as correlated motion from more distant locations can be leveraged to aid deformation of a (possibly) low-contrast target in cine MRI. Therefore, we implemented a transformer for DIR as first model. The second model we implemented is based on auto-segmentation using a CNN. Our hypothesis is that such a model will perform well for larger, high-contrast targets with large motion. We trained the models with different unsupervised, supervised and PS strategies compatible with current MRIgRT clinical workflows and applied them to targets in the thorax, abdomen, and pelvis from a total of 254 patients treated on a 0.35 T MRI-linac.

Methods and Materials

Data

2D+t cine MRIs from 219 patients treated with a 0.35 T MRI-linac at the LMU University Hospital were retrospectively collected with ethics board approval, totaling over 1,400,000 2D frames from different tumor sites (13 abdomen, 58 liver, 93 lung, 2 mediastinum, 28 pancreas, and 25 prostate). All cine MRIs were acquired at either 4 or 8 Hz in a single sagittal plane with a balanced steady-state free precession sequence (TrueFISP; in-plane resolution of 2.4×2.4 or 3.5×3.5 mm²; field-of-view of 270×270 or 350×350 mm²; slice thickness of 5, 7, or 10 mm) and exported in a binary format containing the un-processed images (without segmentations). The cine MRIs included data acquired during both treatment simulation and delivery, and exhibited a mix of breath-holds and periods of free-breathing. All frames were then

resampled to $1 \times 1 \text{ mm}^2$ using linear interpolation and cropped to 224×224 pixels. We rescaled the MRI values 0 and 1000 to 0 and 1. We refer to this cohort as the unlabeled training set.

Additionally, over 7500 cine MRI frames from 35 different patients were similarly collected (4 abdomen, 8 liver, 10 lung, 1 mediastinum, 6 pancreas, 4 prostate plus 2 patients with cardiac tumor). These patients were selected to represent a wide range of tumor sizes and motions, and varying levels of tumor visibility as assessed visually. Only treatment simulation cine MRIs were selected for these patients, to avoid image degradation due to gantry movements in this dataset.¹⁸ Data acquisition, export, and pre-processing were the same as for the first dataset. Additionally, the gross tumor volume or a surrogate tracking target was manually segmented on each frame by one observer based on the tracking targets used clinically. We segmented on average 240 frames per patient (range, 89-545 frames). This cohort was then subdivided into 3 sets: the labeled fine-tuning set (12 patients), validation set (10 patients), and testing set (13 patients). We then asked a second observer to manually segment the gross tumor volume/target for the 13 testing set patients following the same guidelines as the first observer. Figure 1 shows an overview of the 13 testing set patients with segmentations from both observers. Table 1 reports the number of labeled frames, whether the tumor or a surrogate was localized, the acquisition frame rate, statistics of their respective motion, prescribed dose as well as the patient's sex and age.

In summary, we had 4 datasets in this study: (1) the unlabeled training set, (2) the labeled fine-tuning set (observer 1), (3) the labeled validation set (observer 1), and (4) the labeled testing set (both observers).

Target localization models

The 2 different DL models were based on registration and auto-segmentation. The registration model was not optimized to obtain smooth deformations, but accurate target segmentations on the current frame. We thus tolerated a certain amount of folding (negative Jacobian) if the segmentation accuracy was high. We also compared our DL models to a conventional B-spline DIR. A more detailed description of the models follows here:

- TransMorph: we adapted the 3D DIR transformer-based model developed by Chen et al²² to 2D DIR. The model leverages shifted window transformer blocks²³ in the encoder and successive up-sampling and convolutional layers in the decoder. The input to the encoder are the fixed and moving unlabeled 2D cine MRI frames and the output of the decoder is a 2D deformation vector field (DVF), followed by a spatial transformation layer.²⁴ This layer warps the segmentation of the moving cine MRI frame to yield the output binary segmentation. We consistently used the first frame of each cine MRI as moving image/segmentation and the remaining frames as fixed

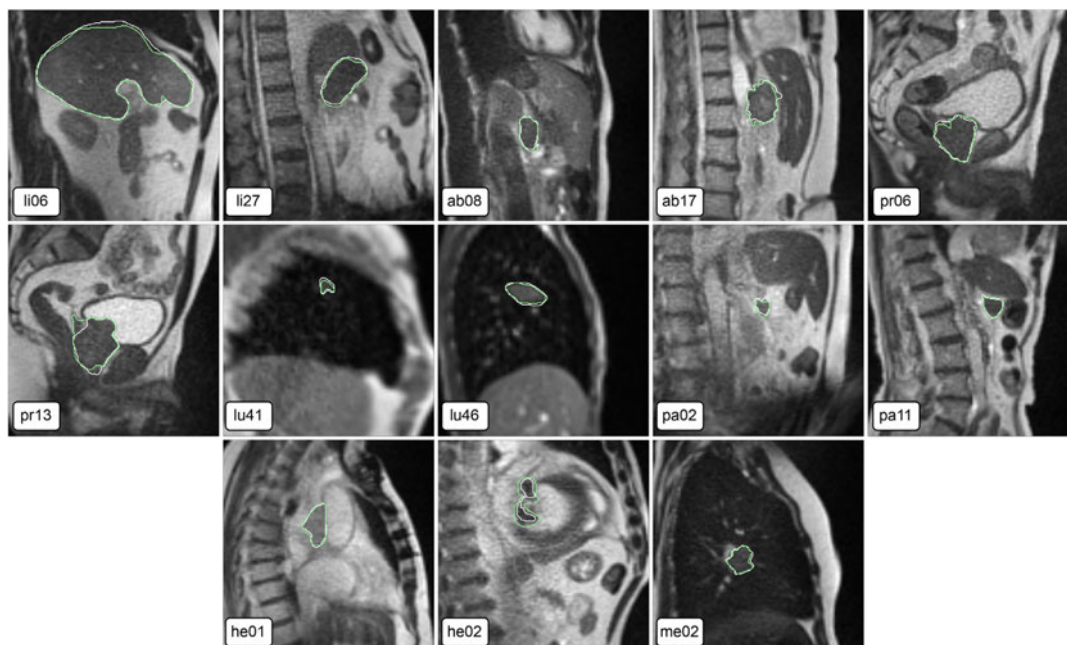


Fig. 1. Segmentation of both observers (observer 1 in white and observer 2 in green) for one cine MRI frame of each testing patient included in this study. Abbreviation: MRI = magnetic resonance imaging.

Table 1 Patient characteristics for the labeled testing set

Patient ID	Labeled frames	Tracking target	Tumor type	FPS	IQR AP/SI (mm)	Max motion AP/SI (mm)	Breath-hold (%)	Dose (Gy)	Sex	Age
liver06 (li06)	240	Surrogate	Metastasis	8	7.1/5.6	35.0/57.0	74	3 × 15	M	56
liver027 (li27)	240	Tumor	Metastasis	8	12.4/19.0	20.0/32.2	47	5 × 8	M	68
abdomen08 (ab08)	545	Tumor	Metastasis	8	0.9/1.7	4.3/7.3	13	5 × 8	F	46
abdomen17 (ab17)	106	Tumor	Metastasis	8	1.0/7.6	4.2/12.5	25	5 × 7	F	61
prostate06 (pr06)	329	Tumor	Primary	8	0.9/1.9	6.9/10.6	-	20 × 3	M	77
prostate13 (pr13)	240	Tumor	Primary	8	0.5/0.9	4.1/5.6	-	5 × 7.5	M	60
lung41 (lu41)	114	Tumor	Primary	4	2.3/1.6	20.4/26.8	84	3 × 13.5	M	77
lung46 (lu46)	89	Tumor	Metastasis	4	6.5/3.7	10.4/9.4	44	10 × 5	F	54
pancreas02 (pa02)	240	Tumor	Primary	8	1.2/8.9	5.8/19.4	59	5 × 8	M	81
pancreas11 (pa11)	240	Tumor	Primary	8	4.8/13.3	11.6/25.9	42	5 × 8	F	91
heart01 (he01)	213	Tumor	Metastasis	8	1.9/8.5	6.8/24.1	37	5 × 7	F	65
heart02 (he02)	477	Surrogate	Primary	8	4.6/8.7	13.8/26.1	11	5 × 6	M	73
mediastinum02 (me02)	189	Tumor	Metastasis	8	8.4/11.2	13.5/20.7	79	10 × 5	F	62

The patient ID also indicates the anatomical area imaged during the cine magnetic resonance imaging (MRI). The number of labeled frames, the tracking target that was segmented, the tumor type, the cine imaging frames per seconds (FPS), the interquartile range (IQR) of motion in anterior-posterior (AP) and superior-inferior (SI) directions, the maximum peak-to-peak motion in AP and SI directions, the relative number of breath-hold frames per patient cine MRI, the prescribed dose (#fractions × dose per fraction in Gy), as well as the patient's sex and age are shown. No breath-holds performed for prostate cases.

images. A more detailed description of the TransMorph architecture can be found in the original publication.²² We did not use the affine transformer included by Chen et al²² before the deformable registration to keep the model as fast as possible. A link to a GitHub repository with the adapted code can be found in section 1.1 of the [supplementary material](#).

- SegResNet: similarly to Friedrich et al,¹⁶ we used a 2D CNN for tumor tracking by means of auto-segmentation. We employed a U-net Segmentation Residual Network (SegResNet). SegResNet takes an unlabeled 2D MRI frame as input and generates a binary tumor segmentation as output. A more detailed description of the model can be found in the original publication.²⁵
- B-spline: we used the cubic B-spline DIR implementation provided by the Plastimatch software.²⁶ This conventional method is based on the iterative optimization of B-spline functions.²⁷ The registration process is made up of N consecutive stages going from coarser to finer image grids in which pixel-wise intensity differences between the 2 frames are minimized to find the most optimized corresponding B-splines. B-spline DIR was implemented to serve as a non-real-time but state-of-the-art benchmark.

Furthermore, 2 baselines were computed:

- No-reg: we computed for each frame the evaluation metrics between the moving segmentation and the fixed segmentation.
- Interobserver: we computed for each frame the evaluation metrics between the ground truth segmentations of the first observer and the second observer.

Model training

Figure 2 summarizes the data subdivision and DL model training strategies. The 4 datasets were used for unsupervised, supervised, and PS strategies:

- TransMorph: we trained the 2D TransMorph model from scratch on the large unlabeled training set while using the labeled validation set to continuously monitor training performance. Although the validation metrics (same as testing set metrics, see next subsection) used labels, the loss function was solely based on image and DVF metrics, making this training unsupervised. We considered mean-square-error (MSE) and local normalized cross-correlation as image losses L_{IMG} , and the first spatial derivative (diffusion regularizer)²⁸ or the second derivative of the DVFs (bending energy regularizer)²⁷ for DVF losses L_{DVF} . DVF regularization was solely included to improve the validation metrics and not to obtain smooth deformations. The relative importance of the image and DVF losses was balanced by the hyperparameters γ and λ :

$$L_{\text{TOT}} = \gamma L_{\text{IMG}} + \lambda L_{\text{DVF}} \quad (1)$$

Other hyperparameters investigated included the usage of image augmentations (random affine, bias field, MRI motion artifacts, elastic deformations, Gibbs noise, and Gaussian smoothing) and the optimizer learning rate. We trained for 30 epochs and model training was stopped after 5 epochs if the validation loss did not improve (early stopping of 5 epochs). The batch size was 192. The proposed 2D model has 31 million trainable weights but is still fast enough for real-time inference, as shown in the results.

- TransMorph-Sup: the TransMorph model was then used as a starting point for continued supervised training on the labeled fine-tuning set, again using the labeled validation set to find the optimal hyperparameters. We explored the same image and DVF losses as for TransMorph plus the segmentation Dice loss L_{SEG} ,²⁹ balanced by an additional hyperparameter β ²²:

$$L_{\text{TOT}} = \gamma L_{\text{IMG}} + \lambda L_{\text{DVF}} + \beta L_{\text{SEG}} \quad (2)$$

Additional hyperparameters included image augmentations (the same as for 2D TransMorph training) and the learning rate. We trained for 500 epochs with an early stopping of 100 epochs and a batch size of 16.

- TransMorph-PS: the TransMorph model was also used as a starting point for continued PS training. We randomly picked 8 frames out of the first 5 seconds of cine MRI of each patient and asked observer 1 to segment the tumor/target. These 8 frames were then used to continue training TransMorph separately for each patient. During training, the PS model was validated on the remaining frames of the patient. This was repeated on the labeled validation set patients to find the best hyperparameters (same parameters investigated as for TransMorph and TransMorph-Sup) while employing the same L_{TOT} as in Eq. 2. The PS models were trained for 300 epochs without early stopping and a batch size of 2. The best set of hyperparameters was chosen according to the labeled validation set and all TransMorph-PS models shared the same set of hyperparameters.
- SegResNet-PS: the same PS procedure was applied to the auto-segmentation SegResNet model. However, as this model is lightweight (6 million weights) and needs less training data compared with TransMorph, it was trained for each patient without starting from pre-trained weights. Substantial data augmentation was used (see below) to avoid overfitting the 8 training frames. This model was solely trained with a segmentation loss, which was a combination of equally weighted Dice and cross-entropy loss.³⁰ Investigated hyperparameters included the usage of image augmentations (random affine, bias field, Gibbs noise, and Gaussian smooth transformations) and the dropout rate, whereas for the learning rate, the default value of 0.0001 was used. Like TransMorph-PS, this model was trained for 300 epochs without early stopping and with a batch size of 2.

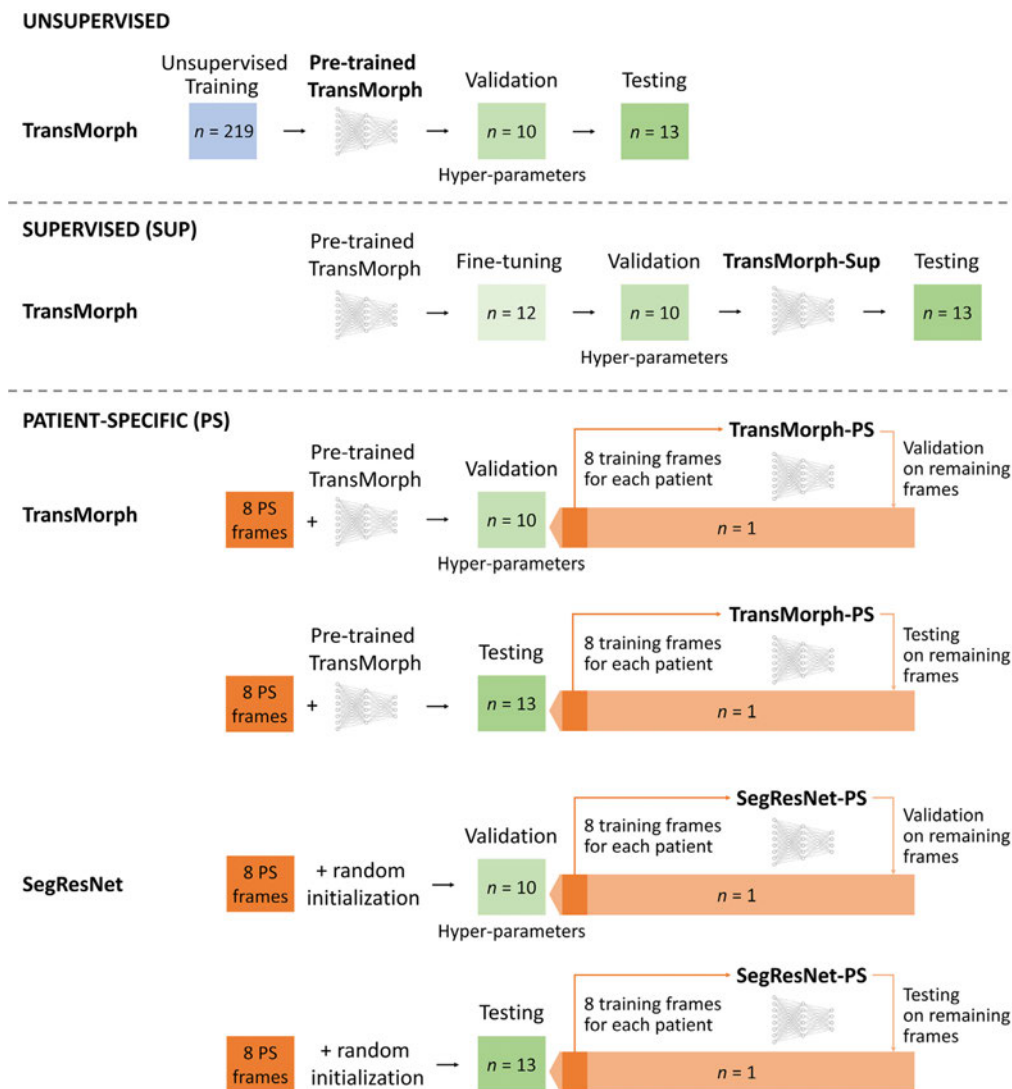


Fig. 2. DL model training strategies. The number of patients in each dataset is indicated by n. PS training is performed 1 patient at a time for both the validation and testing sets. For the transformers, we used the TransMorph pre-trained on the unsupervised training set as a starting point, whereas for SegResNet, we performed PS training from scratch. Abbreviations: DL = deep learning; PS = patient-specific.

For all DL models the optimal set of hyperparameters (see Table E1) was found empirically and independently by repeating training and validation several times. The same 10 patients from the labeled validation dataset were used for all models and model-specific hyperparameters were selected for maximizing the overlap between the segmentations. For all models except the ones based on PS training, the weights of the epoch achieving the best validation performance were used to apply the models to the labeled testing set for both observers. For the TransMorph models, B-spline, and No-reg, the moving segmentation of each observer was

subsequently deformed and/or compared with the ground truth segmentation of the same observer when evaluating performance metrics. For TransMorph-PS and SegResNet-PS, we trained one model per testing patient only once with 8 frames of each patient, using segmentations from observer 1 only, with the set of hyperparameters determined during validation. The model weights of the 300th epoch were then used for testing on the remaining frames, deforming the moving segmentation of each observer as described above. For DIR-based models, performance was evaluated against the same observer as for the moving segmentation.

The optimal parameters for the B-spline model were found in a previous internal investigation and were used also for this study. The registration comprised 4 stages, starting with an initial grid spacing of 90 mm, then 60 mm, 30 mm, and a final stage with 23 mm spacing. For all stages, we set the resolution to 1 mm, the MSE as optimization metric, a regularization weight of 5.0 and the maximum number of iterations to 500.

Evaluation metrics and statistics

Evaluation metrics were calculated based on the model output and the ground truth segmentation for each frame in the validation and testing sets. For fair model comparison, the 8 frames that were used for PS training were excluded from evaluation for all models. Metrics included the Dice similarity coefficient (DSC)³¹ and the Hausdorff distance (HD) 50% and 95%, defined as 50th and 95th percentile of the distances between boundary points, respectively.³² Additionally, the root MSE in superior-inferior direction (RMSE_{SI}) between the centroid of the output and the ground truth segmentations was also computed. For the TransMorph models and the B-spline, the percentage of negative Jacobian determinant (NegJ) of the obtained DVFs was calculated, which represents the fraction of pixels for which folding occurs. For the 2 best real-time models, we reported on a patient-by-patient basis which model achieved better metrics, mimicking clinical workflows where the treatment team may select the best performing tracking algorithm.¹⁵

To determine if a statistically significant difference existed among the testing metrics of the models (excluding interobserver), we combined all the metrics (for all frames of both observers) into a single vector and applied a non-parametric Friedman test,³³ with significance defined as a *P* value less than 0.05. If the Friedman test was significant,

we conducted a post-hoc Nemenyi test³⁴ to identify which model outperformed others in a pairwise comparison.

Results

Table 2 reports testing set evaluation metrics for both observers combined. The metrics obtained for individual observers can be found in Tables E2 and E3. We found TransMorph-PS to outperform all other models in all metrics for both observers, achieving a median (interquartile) DSC of 0.92 (0.03), HD_{50%} of 1.0 (0.1) mm, HD_{95%} of 3.1 (1.9) mm and RMSE_{SI} of 0.7 (0.4) mm. It was slightly better than SegResNet-PS and substantially better than B-spline, TransMorph-Sup and TransMorph. When looking at NegJ, we found B-spline to lead to the smallest folding of pixels, whereas TransMorph-PS led to the largest folding with up to 2.5% NegJ. Both TransMorph-PS and SegResNet-PS achieved an accuracy comparable with the interobserver variability. The No-reg baseline showed that patients exhibited varying amounts of motion.

The Friedman test was found to be significant, and the results of the Nemenyi test are shown in Table E4. Only the B-spline/TransMorph, B-spline/TransMorph-Sup, and the TransMorph/TransMorph-Sup pairs did not show a significant difference.

Figure 3 shows B-spline, SegResNet-PS, and TransMorph-PS results for a series of selected frames for 3 challenging testing set patients. We see how for small targets or targets with limited contrast such as lung41 or heart01, TransMorph-PS performed the best, whereas for large/clearly visible targets with large motion such as liver06, SegResNet-PS was better. Figure E1 shows how TransMorph-PS achieved not only significantly but also substantially better performance than TransMorph and TransMorph-Sup. The trends observed qualitatively were confirmed

Table 2 Dice similarity coefficient (DSC), 50% and 95% Hausdorff distances (HD_{50%} and HD_{95%}), and root-mean-square-error in superior-inferior (RMSE_{SI}) achieved on the labeled testing set for the different models (both observers combined), and interobserver

Model	DSC	HD _{50%} (mm)	HD _{95%} (mm)	RMSE _{SI} (mm)	NegJ (%)
No-reg	0.66 (0.44)	4.2 (3.9)	9.0 (7.1)	6.7 (7.6)	-
B-spline	0.88 (0.06)	1.4 (1.0)	5.0 (2.2)	1.3 (1.0)	<0.001
TransMorph	0.86 (0.12)	1.4 (1.0)	5.0 (2.0)	1.6 (0.9)	0.1 (0.2)
TransMorph-Sup	0.87 (0.08)	1.4 (0.4)	4.5 (2.1)	1.4 (1.1)	0.2 (0.2)
TransMorph-PS	0.92 (0.03)*	1.0 (0.1)*	3.1 (1.9)*	0.7 (0.4)*	2.5 (2.2)
SegResNet-PS	0.90 (0.04)†	1.0 (0.4)†	3.8 (2.0)†	0.9 (1.0)†	-
Interobserver	0.89 (0.03)	1.2 (0.3)	3.9 (2.3)	1.2 (0.4)	-

The median (interquartile range) over all frames of both observers for all patients pooled together is shown. The percentage of negative Jacobian is additionally reported.

Abbreviation: PS = patient-specific.

* Best values for each metric.

† Second-best values.

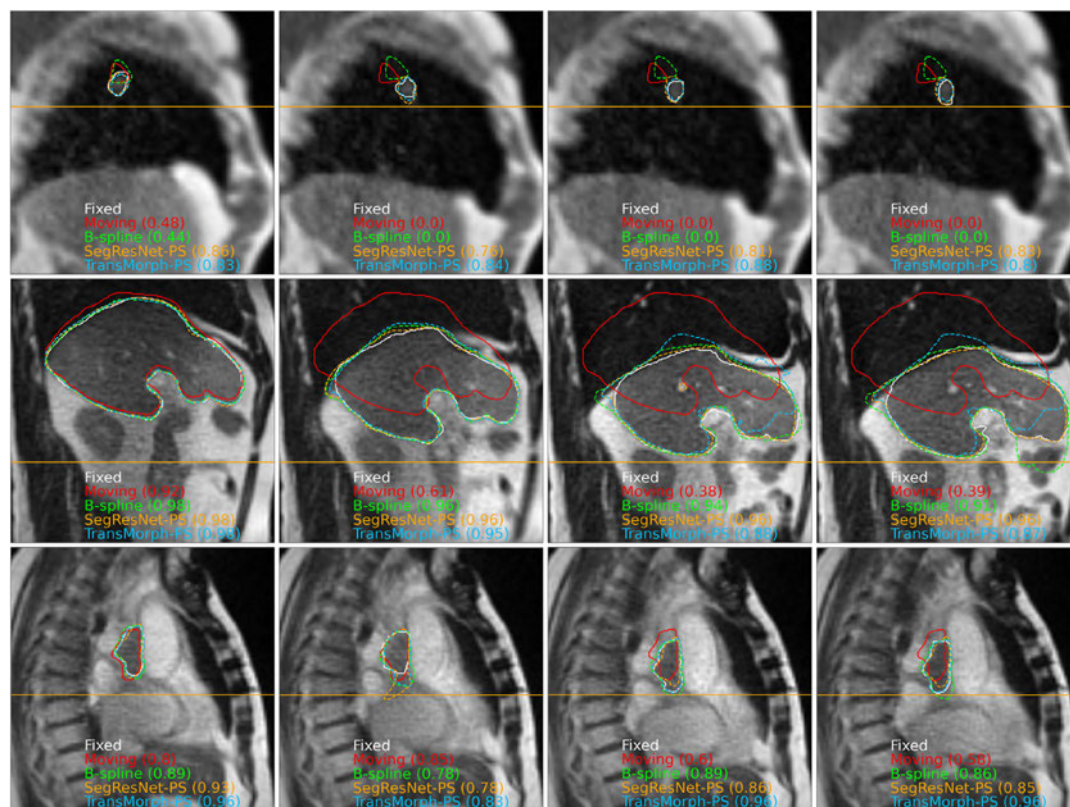


Fig. 3. Comparison of output segmentations. Moving segmentation (red solid, equivalent to No-reg), fixed segmentation (white solid, ground truth segmentation by observer 1), and outputs by different models (dashed) overlaid with the fixed cine MRI for 4 frames representative of the breathing cycle. To improve visibility, only the best performing transformer (TransMorph-PS) is shown. Top: lung tumor (lung41). Middle: liver tumor (liver06). Bottom: cardiac tumor (heart01). The orange line denotes the maximum inhalation phase. The numbers in brackets denote the Dice similarity coefficient (DSC) between current fixed segmentation and the segmentation of the corresponding model. Abbreviations: MRI = magnetic resonance imaging; PS = patient-specific.

quantitatively for all testing patients, as can be seen in Figure 4 for the DSC and the $HD_{95\%}$ and in Figure E2 for the $RMSE_{SI}$ and $HD_{50\%}$. TransMorph-PS achieved a median DSC above 0.80 for all patients and a median $HD_{50\%}$ below 2 mm, $RMSE_{SI}$ below 2 mm, and a median $HD_{95\%}$ below 5 mm for all patients but liver06 (large max motion of 35 mm in anterior-posterior and 57 mm in superior-inferior, see Table 1). A similar observation can be made for SegResNet-PS, which achieved a median $HD_{50\%}$ below 2 mm, $RMSE_{SI}$ below 2 mm and a median $HD_{95\%}$ below 6 mm for all patients but heart02 (limited contrast). We found SegResNet-PS to perform better than TransMorph-PS in 1/12 testing set patients (liver06 where the DSC improved from 0.86 to 0.96 and the $HD_{95\%}$ from 17 to 6 mm). The 2 models achieved the same performance on 2 patients (mediastinum02 and liver27), whereas TransMorph-PS was better on the remaining 9/12. Performance over a temporal series of frames can be appreciated for TransMorph-PS and SegResNet-PS from the animations

provided as Videos E1-E6. The metrics obtained on the validation set are shown in Table E5.

An inference time of about 36 ms was measured for TransMorph and about 8 ms for SegResNet. Registration of 2 frames with B-spline took on average 3 seconds. Segmentation of 8 frames for the PS training took 3 min on average, whereas PS optimization for 300 epochs took an additional 4 min for SegResNet and 8 min for TransMorph.

Discussion

A framework with 2 DL models for target localization with an MRI-linac was investigated and different training strategies were compared. PS training leveraging 8 manually segmented frames substantially increased the performance compared with both supervised and unsupervised training for transformer-based DIR models. Additionally, PS auto-segmentation with SegResNet trained from scratch on

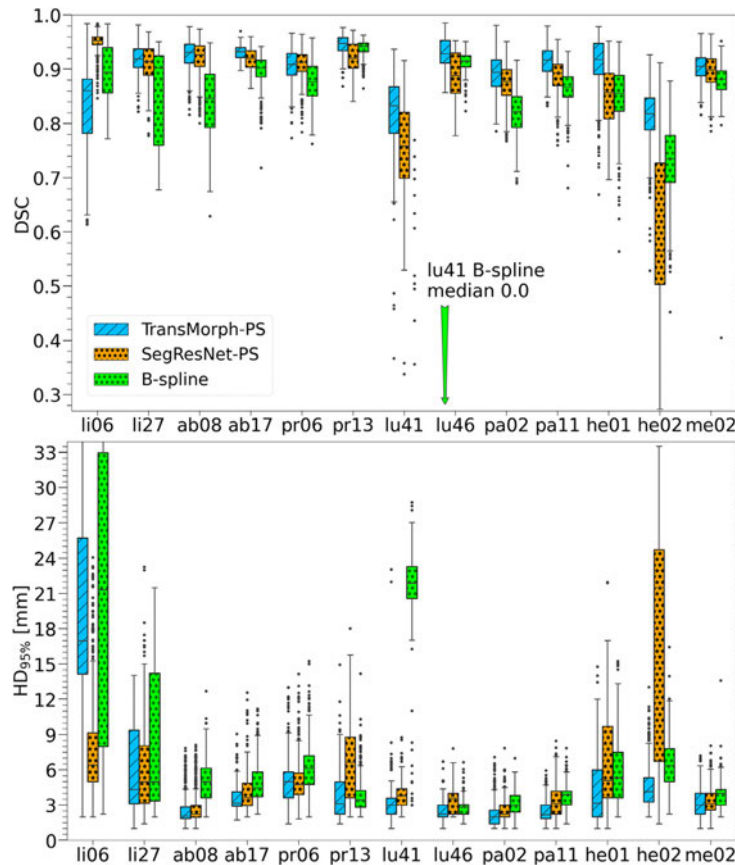


Fig. 4. Box plots of evaluation metrics. Top: Dice similarity coefficient (DSC) and bottom: 95% Hausdorff distance (HD_{95%}) achieved by different models on the labeled testing set (observers combined). To improve visibility, TransMorph, TransMorph-Sup, No-reg, and Interobserver are not shown. For the same reason, we set the y-axis limit for the DSC plot in such a way that the box plot for B-spline for patient lung41 is not visible. Abbreviation: PS = patient-specific.

8 segmented frames can achieve results close to those of TransMorph-PS. Both models achieved accuracy on-par with interobserver variability on the testing set (see Table 2). This shows how the choice of the training strategy (unsupervised, supervised or PS) leads to larger performance differences than the choice of the model (transformer or CNN for DIR or auto-segmentation) for the architectures investigated in this study. Overall, TransMorph-PS was the best model; however, there were some important patient-dependent differences. TransMorph-PS worked better on patients with small or low-contrast targets, whereas SegResNet-PS was better for clearly visible targets with large motion. DIR models can likely use motion from surrounding tissue to correctly deform a low-contrast target but may have problems with very large deformations (compared with the moving image). Auto-segmentation models may be less impacted by large motion as they output the segmentation based on the current frame but may benefit from high-contrast. A pre-treatment workflow could be foreseen in which both TransMorph-PS and SegResNet-PS are trained in parallel using

2 GPUs, and the best one for the current patient is picked by the clinical staff. This would be similar to current clinical practice with different conventional DIR model variants available on the 0.35 T MRI-linac.¹⁵ We found the conventional B-spline model to achieve the same performance as TransMorph and TransMorph-Sup; however, it was significantly worse than the 2 PS models. PS training improved tracking from very inaccurate to an accuracy sufficient for clinical use in patient lung41, for which the DSC/RMSE_{SI} improved from 0.0/17 mm with B-spline to better than 0.7/2 mm with TransMorph-PS or SegResNet-PS.

In section 1.2 of supplementary material, we describe the implementation of TransMorph-lung-PS, for which we performed the unsupervised training based only on the 93 available lung cases. It achieved a slightly worse performance on the testing set than TransMorph-PS, also for the 2 lung patients, which may be caused by using a smaller number of patients for pre-training.

For the 13 testing patients we additionally collected and manually segmented data from the first treatment fraction

to investigate whether the 2 PS models are robust against inter-fractional changes without further re-training (see [supplementary material](#) section 1.3). We found TransMorph-PS to retain high accuracy, whereas a larger drop was observed for SegResNet-PS (which can be expected as the tracking target was sometimes changed). This is clinically relevant because it shows that PS training of TransMorph might be needed only during the simulation session, avoiding a prolongation of the online adaptive workflow for the fractions. On the other hand, if larger changes in the tumor characteristics are observed or if the tracking target is changed by the clinical staff, both PS models could readily be re-trained before the fraction's start.

Finally, we also retrospectively compared the proposed PS framework with clinically used segmentations from the vendor's target localization algorithm and found TransMorph-PS to perform better for all testing patients but for liver06 ([Table E6](#)). A more detailed presentation of the segmentation extraction from the vendor-provided video files and the results is provided in section 1.4 of the [supplementary material](#).

A direct comparison with the study by Hunt et al.¹⁸ is difficult because they did not segment targets. The authors selected the regularization λ for the DVF leading to a percentage of NegJ below 0.5% to obtain physically realistic deformations. Our best model TransMorph-PS had a higher NegJ of 2.5%, which was higher than TransMorph and TransMorph-Sup with NegJ values of 0.1% and 0.2%, respectively. It is likely that the high tracking accuracy achieved with TransMorph-PS comes at the cost of less physically realistic DVFs. This is not problematic when we only require accurate target localization such as in current real-time motion management approaches.³⁵

The DSC of about 0.90 obtained on our liver patients is comparable with the DSC obtained by Friedrich et al.¹⁶ and Frueh et al.¹⁷ on their liver cases. Also, the fact that our best models are comparable with interobserver variability agrees with the finding by Palacios et al.¹⁵ A direct comparison with the models from these studies would require usage of the same patients, because tumor size, visibility, degree of motion, and number of breath-holds are all factors that affect the performance substantially. In addition to the geometric metrics reported in this study, in section 1.5 of the [supplementary material](#) we performed a static dose cloud approximation in 2D to estimate relative changes in the $D_{98\%}$ of the tracking target for different models and confirmed that TransMorph-PS performs best ([Table E7](#)). Similarly, in section 1.6 of the [supplementary material](#) we estimated the tracking efficiency by the proportion of frames for which the center-of-mass error was below 3 mm ([Table E8](#)), and found TransMorph-PS to perform best (99%), followed by SegResNet-PS (97%) and then the vendor's algorithm (86%).

Both TransMorph and SegResNet were found to be fast enough for real-time tracking compared with measured MRI-linac latencies⁸⁻¹⁰ whereas B-spline was, as expected, too slow and was considered in this study as the benchmark

conventional registration model. The 3 minutes required for manual segmentation of 8 frames for PS training would only slightly prolong current online adaptive workflows, which can take up to 30 minutes, excluding irradiation.³⁶ The additional 4/8 minutes for PS training of SegResNet/TransMorph could run automatically in the background. As done currently for the 0.35 T system, the staff could switch model or even improve it by contouring additional frames during the treatment, if needed. Using more frames would probably increase the performance; however, it would also require more time for annotation and training. Future studies could investigate the trade-off between the number of segmented frames used for PS training and model accuracy.

A limitation of this study is that target tracking was limited to a single 2D sagittal plane. This can decrease localization accuracy when motion in the left-right direction is present, causing the target to move in and out of the imaging plane. Out-of-plane motion was observed for 1 patient in this study (lung41), and observers left the segmentation empty for those frames such that they could be excluded from the calculation of the evaluation metrics. As vendors have recently implemented multi-plane cine MRI acquisition clinically,^{13,15} future studies will investigate modifications of the models to track in multiple planes, with the final goal to perform 3D+t target tracking and motion management in real-time.³⁵ Another limitation is that TransMorph-PS requires pre-training on a large unlabeled dataset. However, it can be argued that such datasets are typically available in clinics, and that we made TransMorph's weights publicly available. Finally, only cine MRIs from a 0.35 T MRI-linac were used in this work. In future studies, we plan to investigate accuracy and transferability of the implemented models to MRI-linacs with a different field strength.

Conclusions

A transformer-based DIR model and an auto-segmentation CNN model were found to achieve accurate performance for real-time target localization during MRIGRT for a wide range of tumor sites when using a PS training approach. Personalization to a specific patient only takes a few minutes and would be feasible in current online adaptive pretreatment workflows.

References

1. Hall WA, Paulson ES, van der Heide UA, et al. The transformation of radiation oncology using real-time magnetic resonance guidance: A review. *Eur J Cancer* 2019;122:42-52.
2. Kurz C, Buizza G, Landry G, et al. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol* 2020;15:93.
3. Keall PJ, Glide-Hurst CK, Cao M, et al. ICRU REPORT 97: MRI-guided radiation therapy using MRI-linear accelerators. *J ICRU* 2022;22:1-100.
4. Green OL, Rankine LJ, Cai B, et al. First clinical implementation of real-time, real anatomy tracking and radiation beam control. *Med Phys*

- 2018;45(8). <https://doi.org/10.1002/mp.13002>. Published online May 28.
5. Grimbergen G, Hackett SL, van Ommen F, et al. Gating and intrafraction drift correction on a 1.5 T MR-Linac: Clinical dosimetric benefits for upper abdominal tumors. *Radiother Oncol* 2023;189 109932.
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;17:43.
7. Finazzi T, van Sörnsen de Koste JR, Palacios MA, et al. Delivery of magnetic resonance-guided single-fraction stereotactic lung radiotherapy. *Phys Imaging Radiat Oncol* 2020;14:17-23.
8. Uijtewaal P, Borman PTS, 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.
9. Liu PZY, 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. 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.
11. Mazur TR, Fischer-Valuck BW, Wang Y, Yang D, Mutic S, Li HH. SIFT-based dense pixel tracking on 0.35 T cine-MR images acquired during image-guided radiation therapy with application to gating optimization. *Med Phys* 2016;43:279.
12. Keiper TD, Tai A, Chen X, et al. Feasibility of real-time motion tracking using cine MRI during MR-guided radiation therapy for abdominal targets. *Med Phys* 2020;47:3554-3566.
13. 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 1.5T MR-Linac. *Med Phys* 2023;50:3103-3116.
14. Feng Y, Kawrakow I, Olsen J, et al. A comparative study of automatic image segmentation algorithms for target tracking in MR-IGRT. *J Appl Clin Med Phys* 2016;17:441-460.
15. Palacios MA, Gerganov G, Cobussen P, et al. Accuracy of deformable image registration-based intra-fraction motion management in magnetic resonance-guided radiotherapy. *Phys Imaging Radiat Oncol* 2023;26 100437.
16. Friedrich F, Hörner-Rieber J, Renkamp CK, et al. Stability of conventional and machine learning-based tumor auto-segmentation techniques using undersampled dynamic radial bSSFP acquisitions on a 0.35 T hybrid MR-linac system. *Med Phys* 2021;48:587-596.
17. Frueh M, Kuestner T, Nachbar M, Thorwarth D, Schilling A, Gatidis S. Self-supervised learning for automated anatomical tracking in medical image data with minimal human labeling effort. *Comput Methods Programs Biomed* 2022;225 107085.
18. Hunt B, Gill GS, Alexander DA, et al. Fast deformable image registration for real-time target tracking during radiation therapy using cine MRI and deep learning. *Int J Radiat Oncol Biol Phys* 2023;115:983-993.
19. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need; *Advances in Neural Information Processing Systems* 30. Curran Associates, Inc; 2017. Available at: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html. Accessed November 14, 2023.
20. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv. Published online June 3, 2021. doi:10.48550/arXiv.2010.11929.
21. Tang Y, Yang D, Li W, et al. Self-supervised pre-training of Swin transformers for 3D medical image analysis. arXiv. Published online March 28, 2022. doi:10.48550/arXiv.2111.14791.
22. Chen J, Frey EC, He Y, Segars WP, Li Y, Du Y. TransMorph: Transformer for unsupervised medical image registration. *Med Image Anal* 2022;82 102615.
23. Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. arXiv. Published online August 17, 2021. doi:10.48550/arXiv.2103.14030.
24. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. arXiv. Published online February 4, 2016. doi:10.48550/arXiv.1506.02025.
25. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. In: Crimi A, Bakas S, Kuijff H, Keyvan F, Reyes M, van Walsum T, eds. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Lecture Notes in Computer Science*. Springer International Publishing; 2019:311-320.
26. Shackleford J, Kandasamy N, Sharp G. Chapter 6 - Plastimatch—an open-source software for radiotherapy imaging. In: Shackleford J, Kandasamy N, Sharp G, eds. *High Performance Deformable Image Registration Algorithms for Manycore Processors*. Morgan Kaufmann; 2013:107-114.
27. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Trans Med Imaging* 1999;18:712-721.
28. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Trans Med Imaging* 2019;38:1788-1800.
29. Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*. 565-571.
30. Good IJ. Rational decisions. *J R Stat Soc Series B* 1952;14:107-114.
31. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol* 2004;11:178-189.
32. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 1993;15:850-863.
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 P. *Distribution-Free Multiple Comparisons*. Princeton University; 1963.
35. Lombardo E, Dhont J, Page D, et al. Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects. *Radiother Oncol* 2024;190 109970.
36. Sahin B, Zoto Mustafayev T, Gungor G, et al. First 500 fractions delivered with a magnetic resonance-guided radiotherapy system: Initial experience. *Cureus* 2019;11:e6457.

Patient-specific deep learning tracking framework for real-time 2D target localization in MRI-guided radiotherapy

Supplementary material

Table of contents

1. Supplementary information
 - 1.1 Implementation details
 - 1.2 Lung-only unsupervised TransMorph
 - 1.3 Fractional data
 - 1.4 Comparison with vendor algorithm
 - 1.5 Dosimetric evaluation
 - 1.6 Tracking efficiency analysis
 - 1.7 Demons algorithm
 - References
2. Supplementary tables
 - Table E1: Optimal hyper-parameters found based on the labeled validation set.
 - Table E2: Metrics obtained on labeled testing set (observer 1).
 - Table E3: Metrics obtained on labeled testing set (observer 2).
 - Table E4: P-values obtained from the post-hoc Nemenyi test for the labeled testing set (observers combined) for all possible model pairs.
 - Table E5: Metrics obtained on labeled validation set.
 - Table E6: Metrics obtained on the labeled testing set (observer 1) with the vendor algorithm and TransMorph-PS.
 - Table E7: Relative decrease in tracking target $D_{98\%}$ per testing set patient.
 - Table E8: Tracking efficiency per testing set patient.
3. Supplementary figures
 - Figure E1: Comparison of output segmentations.
 - Figure E2: Box plots of evaluation metrics.
4. Supplementary videos
 - Video E1: TransMorph-PS and SegResNet-PS for patient abdomen17.
 - Video E2: TransMorph-PS and SegResNet-PS for patient heart01.
 - Video E3: TransMorph-PS and SegResNet-PS for patient liver06.
 - Video E4: TransMorph-PS and SegResNet-PS for patient lung41.
 - Video E5: TransMorph-PS and SegResNet-PS for patient lung46.
 - Video E6: TransMorph-PS and SegResNet-PS for patient mediastinum02.

1. Supplementary information

1.1 Implementation details

The code employed in this study was developed using Python 3.8.13 and is accessible at the following URL: https://github.com/LMUK-RADONC-PHYS-RES/transformer_target_localization. All DL models were trained using the Adam optimizer [1]. TransMorph models were optimized using the PyTorch-based framework MONAI 1.1.0 [2], while for SegResNet we used the MONAI Label [3] graphical user interface available in 3D Slicer. Training and inference of the DL models was conducted on an NVIDIA Quadro RTX 8000 GPU equipped with 48 GB of memory while B-spline was executed on an Intel Xeon Gold 6254 (Cascade Lake-EP) 18-Core CPU. Inference times for all models were measured and are reported in the results. For the PS models we also measured the time to manually segment the eight training frames and the time to complete the 300 epochs, as these must be compatible with the duration of the online adaptive treatment workflow on MRI-linacs.

1.2 Lung-only unsupervised TransMorph

Additionally to the unsupervised training of TransMorph using all 219 available patients as described in the main article, we also trained TransMorph in an unsupervised fashion using only the lung patients (93 in total) to investigate whether this could lead to a better performance for this tumor site. The unsupervised-lung TransMorph was refined in a PS fashion in the same way as TransMorph-PS. TransMorph-lung-PS achieved the following median (IQR) results over all testing set patients: DSC of 0.90 (0.04), HD50% of 1.0 (0.1) mm, HD95% of 3.6 (1.6) mm and RMSE_{SI} of 0.8 (0.5) mm, which is slightly worse than TransMorph-PS from the main article. Looking only at the two lung testing patients we found that for lung41 TransMorph-lung-PS achieved a DSC of 0.78 while TransMorph-PS a DSC of 0.84. For lung46 both achieved a DSC of 0.95. These findings confirm our hypothesis that with limited labelled data, unsupervised training should be performed with as many patients as possible as to better learn generic patterns of motion. During the PS refinement phase, the motion patterns of the current tumor can be learned. Finally, we also applied TransMorph-lung (without any PS) to the two lung testing patients and compared it with TransMorph trained on all patients (without any PS). The former obtained a DSC of 0.07 and 0.92 while the latter obtained a DSC of 0.15 and 0.81 on lung41 and lung46, respectively. As already discussed in the main article, including PS training makes the largest difference but it is interesting to note how TransMorph-lung without any PS worked better on a ‘typical lung patient’ such as lung46 compared to TransMorph.

1.3 Fractional data

Additionally to the datasets described in the main article, we collected for the test set up to 100 frames per patient from the first treatment fraction and asked an additional (third) observer to segment this data following the same tracking target used clinically. For some patients we collected less than 100 frames as a gantry movement started within the first 100 frames and disrupted image quality (74 for abdomen08 and 49 for lung46). We then applied TransMorph-PS and SegResNet-PS without re-training to the fractional data and compared with the manually obtained segmentations. This allowed us to investigate whether intra-fractional changes can affect the performance of the PS models. TransMorph-PS retained a high accuracy on the fractions, achieving a median (IQR) over all patients for the DSC, $HD_{50\%}$, $HD_{95\%}$ and $RMSE_{SI}$ of 0.89 (0.06), 1.5 (0.5) mm, 4.8 (3.6) mm and 1.2 (1.2) mm, respectively. It worked well (DSC above 0.80 and $HD_{50\%}$ below 2.5 mm) for all patients but for lung41, for which it achieved a DSC of 0.22 and a $HD_{50\%}$ of 7.0 mm, which would not be acceptable for clinical use. On the other hand, we observed a drop in performance for SegResNet-PS, which achieved a median (IQR) over all patients for the DSC, $HD_{50\%}$, $HD_{95\%}$ and $RMSE_{SI}$ of 0.74 (0.36), 2.2 (3.0) mm, 11.5 (25.6) mm and 3.5 (6.6) mm, respectively. It achieved substantially lower performance for several patients, but in particular for liver27, abdomen17 and pancreas02, for which it achieved a DSC smaller than 0.1 and an $HD_{50\%}$ larger than 50 mm. Interestingly, for these three patients the clinicians had changed the tracking target from the pre-treatment imaging to the first fraction (often a different sagittal slice or another surrogate target is selected to improve tracking performance). It is expected that the performance of SegResNet-PS tends to zero for such cases as it was trained on a substantially different target based on the pre-treatment selection. TransMorph-PS is affected less by such changes as the first frame from the current fraction is deformed and even if the PS training focused on a different target, as long as the pre-treatment and the fractional target are close, we hypothesize that the learned deformations are still representative enough for the fractional tracking target.

1.4 Comparison with vendor algorithm

To enable a comparison with the vendor's target localization algorithm (ViewRay MRIdian software version 5.3.6.11 for all data used in this study), each cine MRI was exported twice: one time in .OGV file format, which is a video containing resampled and interpolated cine MRI frames with the vendor-generated target and gating boundary segmentation overlaid using the color channels of the video. The second export was in binary file format, which contains the un-processed images without any segmentation. The former was processed with in-house software to extract the vendor target segmentations, as described in more detail in [4,5]. The segmentations extracted from each testing patient were visually inspected for correctness, which lead to the exclusion of patients heart01, heart02, mediastinum02, prostate06 and pancreas11 due to a substantial number of frames with segmentations not being correctly extracted from the OGV video. For the sake of transparency, results for these patients are still displayed in the results tables E6, E7 and E8 but are not considered when computing averages/medians and for discussion. The un-processed cine MR images were converted from binary into .MHA using scripts provided by the vendor and used to build all models described in the main article. The observers were asked to manually segment the un-processed frames but could have a look at the frames from the video having the vendor segmentations as a clinical reference. To correctly compare the vendor segmentations with the segmentations obtained with the models of the main article two additional steps were needed: the alignment of the frames both in time and in space. Misalignment between the video and the binary data in time arises due to two reasons: first, the video is saved only starting from the 6th acquired frame, so the first six frames in the binary data need

to be excluded. Second, the frames during which the gantry rotates are not included in the video and the number of skipped frames is not available. While, as shown by Hunt et al. [6], image analysis techniques can be used to approximately find the frames where gantry rotation occurs in the binary data, we decided to use only simulation cine MRI (having no gantry rotations as no treatment is performed) for the testing set patients to ensure a perfect correspondence in time after exclusion of the first six frames. The spatial misalignment arises from the fact that in the video files additional information such as the orientation of the cine, beam status, tracking confidence, etc. are included at the four sides of each image in the video, leading to a small translation and scaling of the actual imaging data. To correct for this, we optimized the parameters of a scaling image transformation followed by a translation transformation to match each video frame and its corresponding extracted vendor segmentation to the binary imaging data and the corresponding segmentations of the algorithms and the ground-truth segmentations described in the main article. The scaling and translation transformations were the same for all patients.

As can be seen in Table E6, TransMorph-PS obtained better performance than the vendor algorithm for all testing set patients for which the segmentations could be correctly obtained (8/13), excluding patient liver06. For that patient, the first frame selected for deformation by TransMorph-PS was in a motion state distant from the breath-hold state, while in clinical practice the clinical algorithm uses a key frame in breath-hold. As shown in Table 1, 74 % of this cine MRI was acquired during a breath-hold. Unfortunately, it is not possible for us to obtain the moving frame used by the vendor as this information is not available in the system and we therefore consider taking the first frame of each video the fairest solution possible. When SegResNet-PS was used for this case instead of TransMorph-PS, superior performance was obtained (mean DSC, HD_{50%}, HD_{95%} and RMSE_{SI} of 0.96, 1.4 mm, 7.5 mm and 1.1 mm, respectively).

1.5 Dosimetric evaluation

We performed a static dose cloud approximation in 2D to get a first order estimate of relative changes in the $D_{98\%}$ of the tracking target for the different models implemented in this study and for the vendor algorithm. For each testing patient, the segmentation of the moving frame (the first frame), was expanded to simulate a 3 mm expansion of the GTV to the CTV, and smoothed with a Gaussian kernel of 6 mm standard deviation for targets in the lung (simulates a dose fall-off similar to those observed for lung patients in our treatment planning system) and of 4 mm for all other targets (dose fall-off for targets in higher density tissue). We considered this the ground truth dose distribution for each patient. We then shifted this dose distribution by the difference between the ground truth centroid position of the tracking target and the centroid position obtained by the investigated model for each frame. We then averaged these shifted distributions to get a centroid-error shifted dose. For each patient the relative difference between the GTV (or tracking target) $D_{98\%}$ for the ground truth distribution and for the final shifted distribution was calculated.

The results of the dosimetric evaluation are shown in Table E7. They confirm that TransMorph-PS is the overall best model. It was slightly better than SegResNet-PS and substantially better than B-spline and the No-reg baseline. When looking at the subset of testing patients for which a comparison with vendor was possible, we found TransMorph-PS to perform better for all patients but liver06. If SegResNet-PS was used instead of TransMorph-PS for liver06, the vendor algorithm would be outperformed for all investigated patients.

1.6 Tracking efficiency analysis

To investigate tracking efficiency, we computed the number of frames for which the distance between the center of mass of the ground truth segmentation and the center of mass of the model output segmentation was less than 3 mm in either superior-inferior or anterior-posterior direction. We chose 3 mm as this expansion is often applied in the clinic to create the boundary margin for gating [7]. The results shown in Table E8 show that TransMorph-PS has the highest tracking efficiency. When looking at the subset of patients for which a comparison with the vendor is possible, we find TransMorph-PS to perform better for all patients but for patient liver06. The vendor algorithm achieved the best tracking efficiency for this case, being slightly better than SegResNet-PS and substantially better than TransMorph-PS. When looking at the median tracking efficiency over the usable subset of testing patients we found TransMorph-PS to be 2% better than SegResNet-PS and 13% better than the vendor algorithm, with larger differences present when looking at specific patients.

1.7 Demons algorithm

Following Hunt et al., we also implemented the Demons algorithm as a second conventional image registration method next to B-spline. We used the “fast symmetric forces” variant implemented in the open-source Python-based library SimpleITK and performed, based on validation patients, a hyper-parameter search for the number of iterations and the standard deviation of the Gaussian filter used to smooth the DVF. We found 200 iterations and a standard deviation of 2.0 pixels to lead to highest validation accuracy and therefore used these values for the testing set patients. Registration of one frame took approximately 7 seconds and the following results were obtained (median and IQR over all testing frames for both observers): DSC = 0.83 (0.09), HD_{50%} = 2.0 (1.0) mm, HD_{95%} = 5.9 (1.7) mm, RMSE_{SI} = 1.9 (2.5) mm. As this model is slower than B-spline and has a worse performance, we presented B-spline as the benchmark conventional model in the main article.

References

- [1] Kingma D.P., Ba J. (2017) Adam: A method for stochastic optimization. *ArXiv*, [doi:10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980)
- [2] Cardoso M.J., Li W., Brown R., ... & Feng, A. (2022) MONAI: An open-source framework for deep learning in healthcare. *ArXiv*, [doi:10.48550/arXiv.2211.02701](https://doi.org/10.48550/arXiv.2211.02701)
- [3] Diaz-Pinto A., Alle S., Nath V., ... & Cardoso M.J. (2024) MONAI Label: A framework for AI-assisted Interactive Labeling of 3D Medical Images, *Med Image Anal*, 95, 103207.
- [4] Xiong, Y., Rabe, M., Nierer, L., Kawula, M., Corradini, S., Belka, C., ... & Kurz, C. (2023). Assessment of intrafractional prostate motion and its dosimetric impact in MRI-guided online adaptive radiotherapy with gating. *Strahlentherapie und Onkologie*, 199(6), 544-553.
- [5] Lombardo, E., Rabe, M., Xiong, Y., Nierer, L., Cusumano, D., Placidi, L., ... & Landry, G. (2022). Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. *Physics in Medicine & Biology*, 67(9), 095006.
- [6] Hunt, B., Gill, G. S., Alexander, D. A., Streeter, S. S., Gladstone, D. J., Russo, G. A., ... & Zhang, R. (2023). Fast Deformable image registration for real-time target tracking during radiation therapy using

cine MRI and deep learning. *International Journal of Radiation Oncology* Biology* Physics*, 115(4), 983-993.

[7] Eze, C., Lombardo, E., Nierer, L., Xiong, Y., Niyazi, M., Belka, C., ... & Corradini, S. (2022). 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. *Radiation Oncology*, 17(1), 43.

2. Supplementary tables

Table E1. Optimal hyper-parameters found for each model based on the labeled validation set. The variable γ stands for the image loss (MSE) weight, λ stands for the deformation loss (diffusion) weight and β stands for the segmentation loss (Dice) weight. The variable p stands for the probability of that given augmentation to be applied during training.

Parameter	TransMorph	TransMorph-Sup	TransMorph-PS	SegResNet-PS
γ	1.0	0.05	0.05	-
λ	0.1	0.05	0.01	-
β	-	1.0	1.0	-
Affine p	0.0	0.5	0.75	0.5
Bias field p	0.0	0.5	0.75	0.5
Motion artifact p	0.0	0.5	0.75	-
Elastic deformation p	0.0	1.0	1.0	-
Gibbs noise p	0.0	0.5	0.75	0.5
Gaussian smooth p	0.0	0.5	0.75	0.5
Learning rate	1e-4	1e-5	1e-5	1e-4
Dropout rate	0.0	0.0	0.0	0.25

Table E2. Metrics obtained on labeled testing set (observer 1) for the different models. Median (inter-quartile range) over patients is shown. Best values for each metric are shown in bold and second-best values in italics.

Model	DSC	HD _{50%} [mm]	HD _{95%} [mm]	RMSE _{SI} [mm]	NegJ [%]
No-reg	0.60 (0.35)	3.9 (4.8)	9.2 (6.6)	8.0 (6.9)	-
B-spline	0.87 (0.06)	1.6 (0.9)	4.6 (2.2)	1.5 (1.3)	< 0.001
TransMorph	0.86 (0.07)	1.5 (1.0)	5.1 (2.5)	1.7 (0.9)	<i>0.1 (0.2)</i>
TransMorph-Sup	0.87 (0.06)	1.4 (0.2)	4.8 (2.6)	1.6 (1.2)	0.2 (0.2)
TransMorph-PS	0.92 (0.03)	1.0 (0.5)	3.2 (2.1)	0.8 (0.7)	2.5 (2.2)
SegResNet-PS	<i>0.91 (0.04)</i>	<i>1.1 (0.3)</i>	<i>4.1 (1.4)</i>	<i>1.0 (0.8)</i>	-

Table E3. Metrics obtained on labeled testing set (observer 2) for the different models. Median (inter-quartile range) over patients is shown. Best values for each metric are shown in bold and second-best values in italics.

Model	DSC	HD _{50%} [mm]	HD _{95%} [mm]	RMSE _{SI} [mm]	NegJ [%]
No-reg	0.58 (0.38)	4.3 (3.8)	10.6 (6.8)	7.1 (6.3)	-
B-spline	0.86 (0.07)	1.6 (1.0)	5.4 (4.0)	1.7 (2.4)	< 0.001
TransMorph	0.82 (0.10)	1.7 (0.7)	5.7 (2.2)	1.6 (1.3)	<i>0.1 (0.2)</i>
TransMorph-Sup	0.86 (0.07)	<i>1.4 (0.3)</i>	<i>4.4 (2.9)</i>	1.6 (1.1)	0.2 (0.2)
TransMorph-PS	0.90 (0.03)	1.1 (0.3)	3.8 (2.7)	1.0 (0.9)	2.5 (2.2)
SegResNet-PS	<i>0.90 (0.06)</i>	1.4 (0.5)	4.8 (4.7)	<i>1.2 (1.4)</i>	-

Table E4. P-values obtained from the post-hoc Nemenyi test for the labeled testing set (observers combined) for all possible model pairs. Significant p-values (<0.05) are denoted with an asterisk.

Model 1	Model 2	p-value
No-reg	B-spline	0.001*
No-reg	TransMorph	0.001*
No-reg	TransMorph-Sup	0.001*
No-reg	TransMorph-PS	0.001*
No-reg	SegResNet-PS	0.001*
B-spline	TransMorph	0.9
B-spline	TransMorph-Sup	0.9
B-spline	TransMorph-PS	0.001*
B-spline	SegResNet-PS	0.001*
TransMorph	TransMorph-Sup	0.9
TransMorph	TransMorph-PS	0.001*
TransMorph	SegResNet-PS	0.001*
TransMorph-Sup	TransMorph-PS	0.001*
TransMorph-Sup	SegResNet-PS	0.001*
TransMorph-PS	SegResNet-PS	0.001*

Table E5. Metrics obtained on labeled validation set for the different models. Median (inter-quartile range) over patients is shown. Best values for each metric are shown in bold and second-best values in italics.

Model	DSC	HD _{50%} [mm]	HD _{95%} [mm]	RMSE _{SI} [mm]	NegJ [%]
No-reg	0.73 (0.18)	5.2 (7.0)	10.0 (8.0)	7.6 (9.2)	-
B-spline	0.82 (0.07)	2.6 (1.9)	10.3 (7.2)	1.4 (3.7)	< 0.001
TransMorph	0.80 (0.12)	2.1 (5.1)	10.9 (8.6)	2.4 (6.6)	0.1 (0.3)
TransMorph-Sup	0.83 (0.08)	2.0 (1.3)	9.7 (5.7)	2.6 (4.3)	0.2 (0.3)
TransMorph-PS	0.90 (0.04)	1.4 (0.5)	5.5 (3.0)	1.3 (1.7)	1.3 (3.6)
SegResNet-PS	0.90 (0.06)	1.3 (0.5)	4.8 (1.3)	1.1 (0.9)	-

Table E6. Metrics obtained on the labeled testing set (observer 1) with the vendor algorithm (VA) and TransMorph-PS (TM-PS). Mean (standard deviation) over the frames of each patient is shown. Median of the means and IQR is shown in the last row. The patients with an asterisk were excluded for fairness from the computation of the median performance for both models due to issues in the extraction of the vendor segmentations.

Patient	DSC		HD _{50%} [mm]		HD _{95%} [mm]		RMSE _{SI} [mm]	
	VA	TM-PS	VA	TM-PS	VA	TM-PS	VA	TM-PS
liver06	0.87(0.15)	0.85(0.08)	4.5(5.7)	5.2(3.2)	12.1(11.6)	18.6(9.6)	5.4(9.5)	9.6(6.0)
liver027	0.88(0.04)	0.92(0.03)	2.1(0.09)	1.2(0.5)	6.6(2.2)	5.9(3.6)	2.5(1.6)	1.3(1.1)
abdomen08	0.88(0.03)	0.94(0.03)	1.1(0.3)	0.8(0.4)	3.2(1.1)	2.1(1.0)	1.1(0.9)	0.5(0.5)
abdomen17	0.90(0.01)	0.93(0.01)	1.3(0.3)	1.0(0.1)	4.2(0.9)	3.6(1.0)	1.0(0.7)	0.8(0.6)
prostate06*	0.82(0.04)	0.92(0.03)	2.3(0.8)	1.2(0.3)	8.9(2.3)	4.5(2.0)	1.7(1.3)	1.4(1.2)
prostate13	0.89(0.02)	0.95(0.01)	1.9(0.4)	1.0(0.1)	7.7(2.2)	3.2(1.7)	2.3(1.1)	0.8(0.8)
lung41	0.70(0.16)	0.84(0.11)	2.1(1.8)	1.3(1.7)	5.0(2.3)	3.0(2.3)	2.9(2.2)	1.0(2.2)
lung46	0.88(0.03)	0.95(0.02)	1.3(0.4)	0.7(0.5)	3.7(0.8)	2.1(1.4)	0.8(0.6)	0.4(0.4)
pancreas02	0.79(0.05)	0.91(0.04)	1.3(0.4)	0.7(0.5)	3.7(1.2)	2.0(0.9)	1.4(0.9)	0.6(0.4)
pancreas11*	0.76(0.04)	0.92(0.03)	1.9(0.5)	0.8(0.4)	7.9(1.8)	2.2(1.0)	1.0(0.6)	0.6(0.4)
heart01*	0.37(0.12)	0.94(0.03)	7.9(3.5)	0.7(0.5)	25.1(4.8)	2.2(1.6)	12.3(5.8)	0.7(0.8)
heart02*	0.62(0.09)	0.80(0.05)	3.7(1.2)	1.2(0.3)	9.0(1.6)	4.3(1.5)	1.9(1.5)	1.5(1.3)
mediastinum02*	0.74(0.04)	0.91(0.03)	2.4(0.9)	1.0(0.2)	12.4(1.6)	3.4(1.2)	3.0(1.0)	0.7(0.5)
<i>median</i>	0.88(0.04)	0.93(0.05)	1.6(0.7)	1.0(0.4)	4.6(1.2)	3.0(1.6)	1.9(0.9)	0.8(0.9)

Table E7. Relative decrease in tracking target $D_{98\%}$ in % obtained for each patient in the labeled testing set (observer 1). The patients with an asterisk were excluded for fairness from the computation of the median performance for all models due to issues in the extraction of the vendor segmentations. Lower values are better.

Patient	No-Reg	B-spline	TransMorph-PS	SegResNet-PS	Vendor
$D_{98\%}$ decrease (%)					
liver06	86	28	62	12	20
liver027	46	27	2	10	5
abdomen08	1	18	0	0	5
abdomen17	53	4	1	1	2
prostate06*	4	17	5	4	28
prostate13	2	3	2	7	15
lung41	99	99	1	7	20
lung46	33	0	0	1	1
pancreas02	81	7	0	5	7
pancreas11*	28	12	0	1	27
heart01*	55	10	0	10	82
heart02*	72	7	5	65	7
mediastinum02*	67	2	0	1	56
<i>median</i>	50	13	1	6	7

Table E8. Tracking efficiency in % obtained for each patient in the labeled testing set (observer 1). The patients with an asterisk were excluded for fairness from the computation of the median performance for all models due to issues in the extraction of the vendor segmentations. Higher values are better.

Patient	No-Reg	B-spline	TransMorph-PS	SegResNet-PS	Vendor
tracking efficiency (%)					
liver06	7	42	18	66	70
liver027	31	61	91	75	60
abdomen08	99	77	100	100	96
abdomen17	28	86	100	100	100
prostate06*	90	72	90	93	44
prostate13	98	96	98	91	77
lung41	1	2	96	93	68
lung46	20	100	100	100	100
pancreas02	12	100	100	100	95
pancreas11*	29	98	100	100	52
heart01*	6	90	98	90	0
heart02*	15	86	88	25	77
mediastinum02*	14	100	100	100	1
<i>median</i>	24	82	99	97	86

3. Supplementary figures

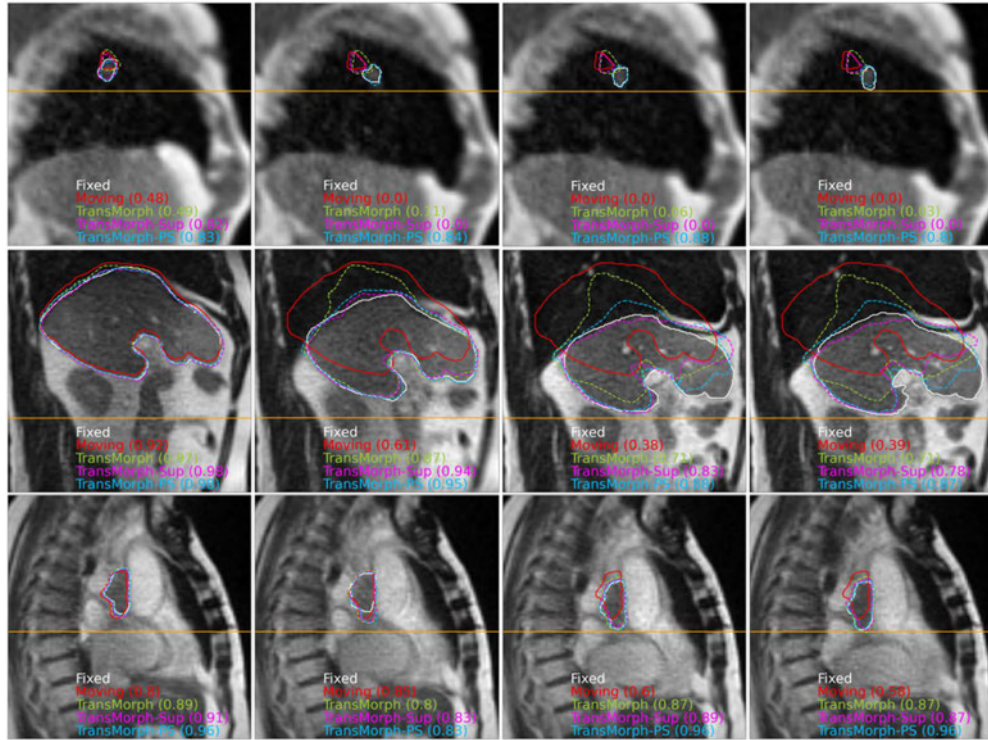


Figure E1. Comparison of output segmentations. Moving segmentation (red solid, equivalent to No-reg), fixed segmentation (white solid, ground truth segmentation by observer 1) and outputs by different TransMorph models (dashed) overlaid with the fixed cine MRI for four frames representative of the breathing cycle. To improve visibility, only the transformers with different training strategies are shown. (Top) Lung tumor (lung41). (Middle) Liver tumor (liver06). (Bottom) Cardiac tumor (heart01). Orange line denotes the maximum inhalation phase. Number in brackets denotes the Dice similarity coefficient (DSC) between current fixed segmentation and segmentation of corresponding model.

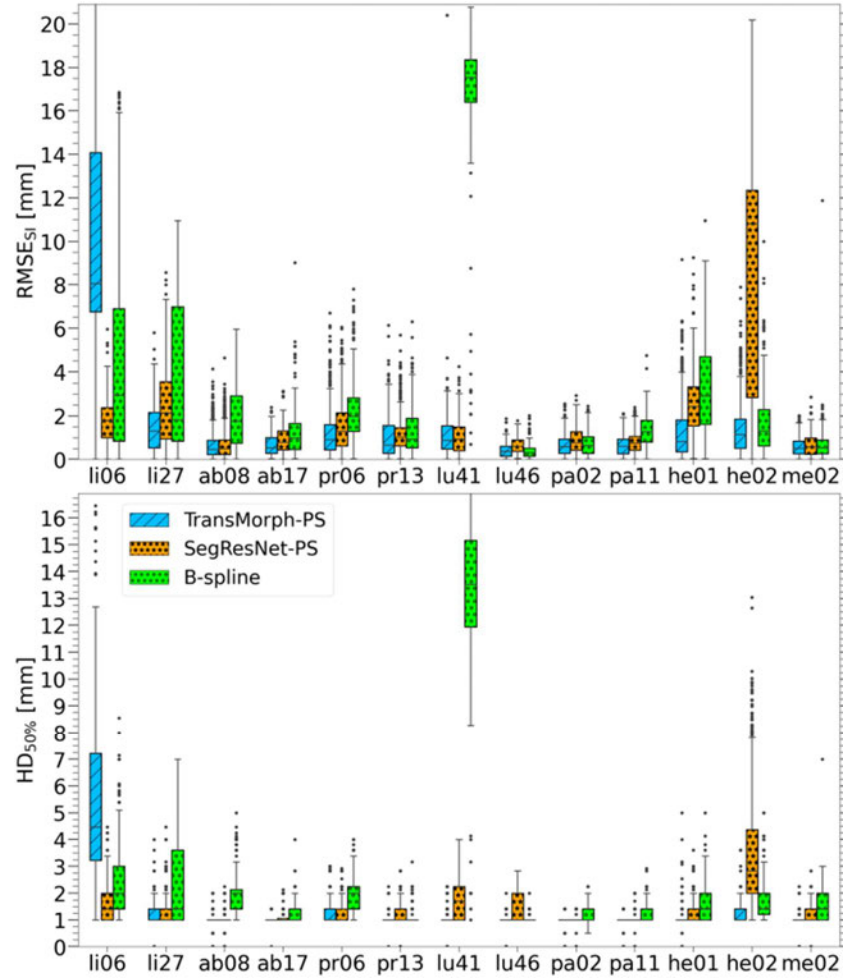
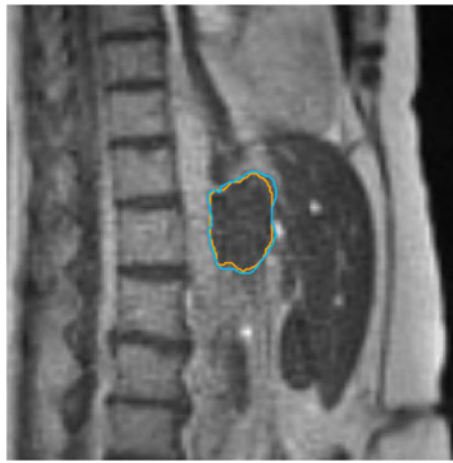


Figure E2. Box plots of evaluation metrics. (Top) Root-mean-square-error in superior-inferior (RMSE_{si}) and (Bottom) 50% Hausdorff distance (HD_{50%}) achieved by different models on the labeled testing set (observers combined). To improve visibility, TransMorph, TransMorph-Sup, No-reg and Inter-observer are not shown.

4. Supplementary videos



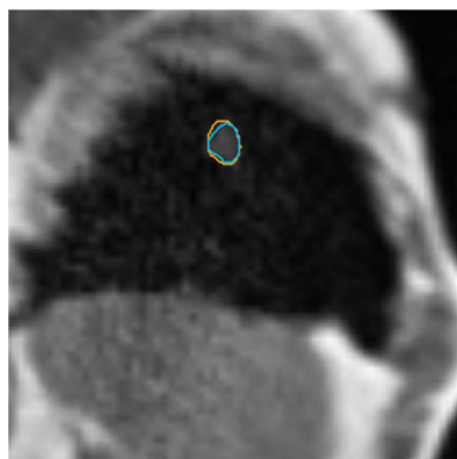
Video E1: TransMorph-PS and SegResNet-PS performance for patient abdomen17. Output segmentation overlaid with cine MRI at the original frame rate of 8 Hz. Video file can be found online under Appendix Supplementary material.



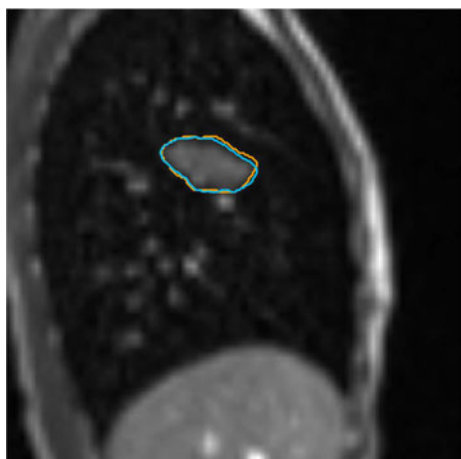
Video E2: TransMorph-PS and SegResNet-PS performance for patient heart01. Output segmentation overlaid with cine MRI at the original frame rate of 8 Hz. Video file can be found online under Appendix Supplementary material.



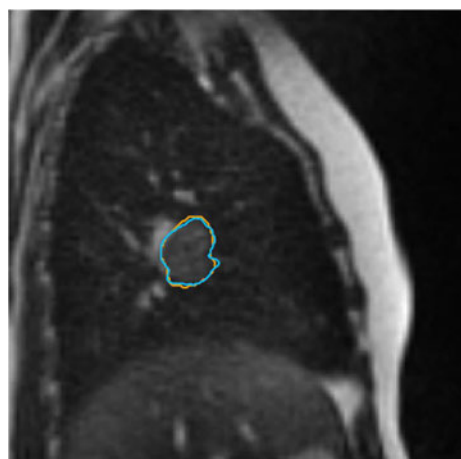
Video E3: TransMorph-PS and SegResNet-PS performance for patient liver06. Output segmentation overlayed with cine MRI at the original frame rate of 8 Hz. Video file can be found online under Appendix Supplementary material.



Video E4: TransMorph-PS and SegResNet-PS performance for patient lung41. Output segmentation overlayed with cine MRI at the original frame rate of 4 Hz. Video file can be found online under Appendix Supplementary material.



Video E5: TransMorph-PS and SegResNet-PS performance for patient lung46. Output segmentation overlayed with cine MRI at the original frame rate of 4 Hz. Video file can be found online under Appendix Supplementary material.



Video E6: TransMorph-PS and SegResNet-PS performance for patient mediastinum02. Output segmentation overlayed with cine MRI at the original frame rate of 8 Hz. Video file can be found online under Appendix Supplementary material.

Chapter 3

Outlook

Motion occurring during treatment, for instance due to respiration, remains one of the major challenges in modern radiotherapy. MRI-linacs represent the state-of-the-art for real-time motion management and currently rely on gated beam delivery with the patients holding their breath. However, this comes at the cost of prolonged treatments, which leads to less patients being able to benefit from this technology and therefore higher costs. MLC-tracking in free-breathing as a more time efficient alternative to gating is currently being investigated, however, it requires, among others, more accurate target localization in all breathing states and compensation of the system latency to avoid lagging behind the moving tumor.

In this thesis, different DL algorithms were developed to tackle the above mentioned problems. To compensate for the MRI-linac latency, different LSTM motion prediction models have been developed in-silico and then experimentally validated for MLC-tracking on the Australian MRI-linac. Furthermore, a transformer DIR model was developed for real-time target localization based on cine MRI and shown to outperform both conventional DIR and DL auto-segmentation models in-silico. The experimental validation of the proposed models in an MLC-tracking study with a Viewray-MRIdian system is limited by the fact that the vendor does not provide leaf control/access to the motion management framework and further complicated by the hardware constraint of the MLC leaves not moving parallel to the main component of motion, i.e., the superior-inferior direction. For these reasons, future MLC-tracking experiments will likely be performed on the Australian MRI-linac or an Elekta-Unity. A potential implementation issue regarding the presented target localization transformer is that the manual labelling of multiple frames needed for the patient-specific training requires 3 minutes and the training itself another 4 minutes, which adds additional complexity to the online adaptive MRIGRT workflow. A possible solution to get rid of the patient-specific training is currently being investigated in a follow-up study by another PhD student and involves the usage of foundation models, large networks that are trained on vast amounts of data to perform a wide range of tasks across different domains and that can be readily adapted to a specific application. Whether transformers can replace LSTMs for motion prediction and whether they are going to remain the first-choice architecture for most tasks remains to be seen, with a recent publication by the inventor of the LSTM suggesting that xLSTM, a new LSTM model with attention, can theoretically outperform state-of-the-art transformers both in performance and scaling [62]. A well-known issue for science in general and in particular for comparative DL studies is confirmation bias: authors involuntarily tend to optimize the performance of the most interesting/newest model more, therefore making

the comparison potentially unfair. Challenges in which participants from all over the world can test their (AI or non-AI) algorithm on large curated datasets are considered a better way to objectively compare models for specific tasks. Such a challenge for target localization based on 2D cine MRI from commercial MRI-linacs is currently being co-organized by the candidate together with researchers from the LMU, Italy, Belgium, the Netherlands, the Czech Republic, Australia and China.

Finally, while MRI-guided MLC-tracking in free-breathing would lead to a substantial improvement in terms of efficiency compared to gating, it still has drawbacks such as information about motion being limited to a single 2D slice and OAR movement with respect to the target being potentially different from assumptions made during treatment planning on the static anatomy. The topical review published by the candidate and presented in section A.1 proposes a 3D+t dosimetric MLC-tracking workflow in which the MLC leaves are moved to account for current under/over-dosage with respect to the reference treatment plan as a possible improvement to the current state-of-the-art 2D+t geometric MLC-tracking. The proposed workflow requires AI in many of its steps, including but not limited to extending the motion prediction models presented in this thesis to the prediction of future 3D deformations or the extension of the target localization 2D DIR model to the real-time estimation of 3D+t deformation fields from 2D+t multi-plane input data. While all datasets in the candidate's publications comprised 2D cine MRIs acquired in a single sagittal plane, multi-plane 2D cine MRI has been collected from the MRIdian at the LMU University Hospital recently. This data could be used in future studies to evaluate both conventional [63, 64] and AI-based [65, 66] 2D+t to 3D+t motion estimation models.

3.1 Conclusions

This thesis provides convincing evidence on the role DL can play for real-time motion management in MRIGRT. In in-silico studies using large clinical datasets from two MRI-linacs it was shown that DL algorithms significantly outperform conventional ones both for motion prediction and for target localization while achieving inference times below 50 ms, which is necessary for real-time deployment. Of note, for motion prediction the in-silico results were additionally confirmed in a real-time experimental study with a prototype MRI-linac. Several follow-up projects resulting from this work are already ongoing and include the application of the developed motion prediction model to cardio-respiratory motion traces acquired with an Elekta-Unity in the Netherlands, the in-silico validation of the developed target localization model on 2D cine MRI data from an Elekta-Unity in China, the experimental validation of the developed target localization model or a follow-up foundational model on the Australian MRI-linac and the organization of a target localization challenge using multi-institutional 2D cine MRI data from different MRI-linacs, with the ultimate goal of improving the accuracy of MRIGRT for the treatment of cancer.

Appendix A

Additional publication

A.1 Topical review: Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects

This topical review was published as the result of a workshop organized by the European Society for Radiotherapy and Oncology in 2022 in Lisbon. Leveraging the expertise from top researchers from multiple countries, the candidate summarized the existing literature on the topic of real-time motion management with MRI-linacs and proposed a future AI-enabled 3D+t dosimetric MLC-tracking workflow as a potentially more accurate alternative to the current 2D+t geometric MLC-tracking.

Reprinted with permission from "Lombardo, E., Dhont, J., Page, D., Garibaldi, C., Künzels, L. A., Hurkmans, C., ... & Placidi, L. (2023) Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects. *Radiotherapy and Oncology*, 109970".



Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects

Elia Lombardo^a, Jennifer Dhont^{b,c}, Denis Page^d, Cristina Garibaldi^e, Luise A. Künzel^{f,g,h},
Coen Hurkmansⁱ, Rob H.N. Tijssenⁱ, Chiara Paganelli^j, Paul Z.Y. Liu^{k,l}, Paul J. Keall^{k,l},
Marco Riboldi^m, Christopher Kurz^a, Guillaume Landry^{a,n,o}, Davide Cusumano^p,
Marco Fusella^{q,*}, Lorenzo Placidi^r

^a Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, Germany^b Université libre de Bruxelles (ULB), Hôpital Universitaire de Bruxelles (H.U.B.), Institut Jules Bordet, Department of Medical Physics, Brussels, Belgium^c Université Libre De Bruxelles (ULB), Radiophysics and MRI Physics Laboratory, Brussels, Belgium^d University of Manchester, Division of Cancer Sciences, Manchester, United Kingdom^e IBO, Unit of Radiation Research, European Institute of Oncology IRCCS, Milan, Italy^f National Center for Tumor Diseases (NCT), Dresden, Germany; German Cancer Research Center (DKFZ), Heidelberg, Germany; Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; Helmholtz-Zentrum Dresden - Rossendorf (HZDR), Dresden, Germany^g Department of Radiotherapy and Radiation Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany^h OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden - Rossendorf, Dresden, Germanyⁱ Department of Radiation Oncology, Catharina Hospital, Eindhoven, the Netherlands^j Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy^k Image X Institute, University of Sydney Central Clinical School, Sydney, NSW, Australia^l Department of Medical Physics, Ingham Institute of Applied Medical Research, Liverpool, NSW, Australia^m Department of Medical Physics, Faculty of Physics, LMU Munich, Munich, Germanyⁿ German Cancer Consortium (DKTK), Partner Site Munich, a Partnership between DKFZ and LMU University Hospital Munich, Germany^o Bavarian Cancer Research Center (BZKF), Partner Site Munich, Munich, Germany^p Mater Olbia Hospital, Olbia, Italy^q Department of Radiation Oncology, Abano Terme Hospital, Italy^r Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Roma, Italy

ARTICLE INFO

Keywords:

MRI-linac
Intra-fractional motion
Motion estimation
Tumour tracking
Motion prediction
AI

ABSTRACT

MRI-guided radiotherapy (MRIGRT) is a highly complex treatment modality, allowing adaptation to anatomical changes occurring from one treatment day to the other (inter-fractional), but also to motion occurring during a treatment fraction (intra-fractional). In this vision paper, we describe the different steps of intra-fractional motion management during MRIGRT, from imaging to beam adaptation, and the solutions currently available both clinically and at a research level. Furthermore, considering the latest developments in the literature, a workflow is foreseen in which motion-induced over- and/or under-dosage is compensated in 3D, with minimal impact to the radiotherapy treatment time. Considering the time constraints of real-time adaptation, a particular focus is put on artificial intelligence (AI) solutions as a fast and accurate alternative to conventional algorithms.

Introduction

Continuous technological advancements over nearly two decades have led to the commercialization of MRI-guided radiotherapy (MRIGRT), introducing what may be the epitome of image-guided

radiotherapy into the clinic [1,2]. By combining a linear accelerator with the superior soft tissue contrast of MRI, MRIGRT offers higher geometric accuracy and treatment efficacy through online adaptation [3–5]. Furthermore, the availability of functional MRI enables exploration of novel strategies such as targeting aggressive or resistant tumour

* Corresponding author.

E-mail address: mfusella@casacura.it (M. Fusella).<https://doi.org/10.1016/j.radonc.2023.109970>

Received 18 August 2023; Received in revised form 19 October 2023; Accepted 22 October 2023

Available online 26 October 2023

0167-8140/© 2023 Elsevier B.V. All rights reserved.

A.1. Topical review: Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects

107

E. Lombardo et al.

Radiotherapy and Oncology 190 (2024) 109970

subregions [6,7]. Recent feasibility studies have also demonstrated the possibility to monitor and predict treatment response on the MR images acquired during RT, a promising development for personalized treatment approaches [8].

Next to the anatomical and functional imaging of changes occurring from one treatment day to the other (inter-fractional), MRIgRT also provides direct and real-time monitoring of changes occurring during a treatment fraction (intra-fractional) [109]. Through cine MRI, intra-fraction motion such as slow drifts, sudden shifts or (semi-)periodic motion of the target but also of nearby organs at risk (OARs) can be observed [9-12]. These observations open the door to real-time adaptation, but the requirements for real-time motion management go beyond the availability of images. Fast and accurate localization, motion prediction to overcome latency, and the actual adaptation through gating or tracking are equally critical [13]. Of further consideration in the context of real-time motion management is the currently available spatial and temporal resolution of cine MRI on MRI-linacs, and the extent of the artefact-free field-of-view which is often limited to two dimensions at high temporal resolution. With online adaptation and real-time motion management added to the task list, MRIgRT has become a highly complex treatment modality [14,107].

During the ESTRO Physics workshop 2022 on “Next generation MR-guided radiotherapy: AI applications for planning and image guidance”, we built a working group on real-time motion management in MRIgRT, from which we present the results of our discussion in this article. Our aim is to describe the status and limitations of intra-fractional motion management in MRIgRT based on the authors’ experience, communication with the different vendors and a review of the existing literature. To address limitations of current intra-fractional motion management approaches, an envisioned workflow is proposed. Considering the amount of data generated during an MRIgRT treatment fraction and the need for short computational times in real-time motion management, it is not surprising that an often-proposed solution to some of the current hurdles in MRIgRT is artificial intelligence (AI) [15]. We therefore cast a particular focus on these approaches and the benefit they can bring.

Terminology and overview

Over the last decade, the terms motion tracking, prediction, and estimation have been used interchangeably in the scientific literature [10,16-22]. In this work, we adopt the terms real-time or intra-fractional motion management to refer to the overall procedure going from the real-time visualization of target/tumour motion via dedicated imaging up to the adaptation of the radiation beam occurring during an RT treatment fraction [109,110]. From this perspective, tracking, prediction, and other related terms can be viewed as specific components or steps within this comprehensive management procedure, as we will elaborate upon in subsequent sections. Although this work focuses on MRI-guidance, these terms can also be applicable to image-guided radiotherapy in general. The current intra-fractional motion management procedure in MRIgRT can be conceptually divided into four essential steps (see Fig. 1):

Real-time imaging

The first step involves real-time visualization of the patient anatomy using cine MRI with a dedicated MRI sequence which is fast and robust to motion, providing continuous and consistently high-quality imaging needed for intra-fractional motion management [23]. Compared to other fast imaging techniques like fluoroscopy, cine MRI offers superior soft-tissue contrast, enabling accurate localization of various targets, including lung tumours, pancreatic lesions, and primary and secondary hepatic lesions [5]. Furthermore, cine MRI carries no additional radiation dose, allowing safe and continuous verification throughout all treatment fractions [24].

Target/OAR localization

Once a new cine MRI frame is acquired, the next step involves localizing the treatment target and if needed nearby OARs. This step is often referred to as motion tracking, however, due to etymological similarity with MLC-tracking, we prefer the term localization. In this work, the term motion estimation is used to refer to the more general computation of displacement fields (no contours involved). Various algorithms have been proposed for real-time target localization on cine MRIs and will be discussed in the next section. While some methods focus on robustness by providing the updated centre of mass (COM) position of the target [16] other methods update the target contour itself [22,25], thus considering both rigid displacements as well as rotations and deformations.

Motion prediction

Each step of the motion management workflow requires a certain amount of time. The resulting time delay between the physical motion of the target and the execution of the adaptation is called system latency. Compared to other motion monitoring techniques, the acquisition of cine MRIs is comparably slow and represents the primary contributor to system latency. Measurements have shown that system latencies can range from approximately 200 ms to 440 ms, depending on imaging frequency, beam adaptation technique and MRI-linac system [17,25-27]. For comparison, measured latencies of 115 ms for a robotic linac system (optical marker localization and robotic adaptation) [104,111], 48 ms for a gimbaled linac system (infra-red marker localization and MLC-tracking) [105] and 350 ms for a Kv intrafractional monitoring system (kilovoltage intrafraction monitoring for implanted fiducial localization and gating/couch-shifts or MLC-tracking) [106] have been reported. To mitigate the dosimetric effects of latency, temporal prediction algorithms have been explored. These algorithms use a sequence of previous inputs, which can be target coordinates, contours or entire cine MRI frames to forecast the future target location to be used in the beam adaptation step.

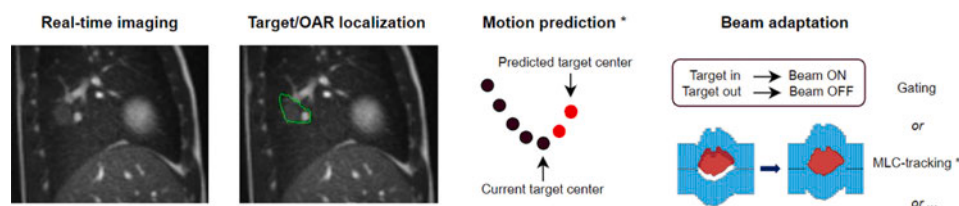


Fig. 1. Workflow of real-time motion management in MRIgRT. 1. First, a 2D cine MRI is acquired to visualise the patient’s moving anatomy. 2. Then, the position of the moving target is localized on the current cine MRI. 3. Prediction of the future target position can be used to overcome system latency. 4. Finally, the beam is adapted to compensate for the target’s motion using one or a combination of different techniques. (*) Asterisks denote that motion prediction and MLC-tracking have currently been implemented only on prototype MRI-linacs. Adapted with permission from [93].

E. Lombardo et al.

Radiotherapy and Oncology 190 (2024) 109970

Beam adaptation

In the final step of intra-fractional motion management, the actual adaptation of the radiation beam based on the observed or predicted motion takes place. In this work, the term adaptation is used to describe either a change in the spatial distribution of the beam, for instance when the MLC aperture is re-shaped, or a change in time, for instance when the beam is paused during gated delivery. Several techniques have been developed and clinically applied for x-ray guidance in recent years [86–88]. While certain modifications may be required to adapt these techniques for use on MRI-linacs, the majority can be directly applied to MRIGRT.

Clinical motion management, challenges and research solutions

In the following section, a detailed description of the different methods which are available clinically and at a research level for the different steps of real-time motion management in MRIGRT will be provided. A focus will be given to open challenges and where AI can provide advancements.

Real-time imaging

Ideally, MRI for intra-fractional motion management would involve real-time 3D cine MRI, i.e., continuous 3D imaging at a framerate of few Hz and a voxel size of few mm³. However, achieving this ideal scenario remains a significant challenge due to the inherent trade-off between spatial and temporal resolution [28,29]. As a result, current techniques for time-resolved MRI in motion monitoring typically rely on acquiring 2D information, which enables image acquisition at sub-second frame rates using specialized 2D cine MRI sequences. These can be acquired either on a single plane or multiple parallel planes (commonly in the sagittal direction), or, more recently, in an orthogonal sagittal/coronal/axial interleaved fashion [30]. Imaging is typically centred in the tumour to achieve pseudo 3D information. Cine MRI used in clinical practice today is based on balanced steady-state free precession sequences with cartesian or radial k-space read-out. With voxel-sizes in the order of few millimetres (2–4 mm in-plane with slice thicknesses of 5–15 mm), a temporal resolution of up to 8 Hz is achieved [26,31], with reduced frame rates for multi-slice acquisition.

As mentioned previously, the main contributor to the MRIGRT system latency is the time for image acquisition, which has been investigated extensively for gated beam delivery [34]. Alternative k-space trajectories have been proposed on a prototype 1.5 T MRI-linac to mitigate latencies directly during imaging, without relying on temporal motion prediction [36]. Further options exploited to speed up imaging in 2D include parallel imaging and under-sampling schemes [37]. Developments to go from multi-slice 2D cine imaging to 3D imaging are currently ongoing. 3D cine MRI at frame rates of about 2 Hz, yet at a coarse spatial resolution of about 5 mm isotropic has been shown feasible using current commercial acceleration methods [38]. While this approach may be sufficient for motion management of pelvic tumour sites, more advanced acceleration methods are needed for respiratory resolved imaging of the abdomen and thorax. AI may provide the necessary tools to achieve the required temporal and spatial resolution. AI models for MRI reconstruction are rapidly evolving and have the benefit to be computationally fast compared to reconstruction methods like compressed sensing [39,92]. Super-resolution AI models trained on pairs of low- and high-resolution MRI images can up-sample acquired low resolution images to a higher spatial resolution without increasing latency significantly [82,89]. In addition, one can adopt motion modelling techniques to generate 3D data with acceptable spatial resolution from 2D cine MRI [86], by creating a correlation model between a-priori 3D or respiratory-correlated 4D imaging and in-room 2D cine MRI; fast implementations are however required to achieve high frame rates. The estimation of 3D MRI at 3.65 Hz using a conventional motion

modelling method has been demonstrated in a phantom study on a 0.35 T MRI-linac [40,103], and AI solutions to obtain real-time 3D information are also emerging [41,57], as discussed in the subsection *Target/OAR localization*.

Target/OAR localization

Target localization algorithms can either solely rely on the current image or more typically on prior knowledge such as the target shape on a planning image.

Over the past decade, various methods have been explored at a research level, an artificial neural network (ANN) already being used for this task in 2011 [16]. The ANN took as input the principal components of the image values from a relatively small region of interest fixed around the target and encompassing its range of motion, thus reducing the input size considerably. The ANN was compared to template matching using the cross-correlation metric, where a manually selected template image of about 10 × 10 pixels was selected from the first frame. Shi et al. [42] also leveraged template matching to localize lung tumours on 2D cine MRI, achieving accuracy at a similar level to manual annotation. Mazur et al. [43] published one of the first studies using cine MRI from the 0.35 T MRI-linac, using a scale invariant feature transform (SIFT) approach, where a deformable spatial pyramid matching approach was employed to find corresponding image features. They reported good tracking accuracy, but their method took 250 ms per frame and was not used clinically. Paganelli et al. [44] also used SIFT but applied it to interleaved orthogonal (sagittal and coronal planes) 2D images from a 3 T scanner. Bourque et al. [45] used a diagnostic 1.5 T scanner to localize lung tumours in four lung cancer patients with a sequential Monte Carlo method called particle filter. Several other groups have explored similar approaches as listed above [46–49]. A few research papers have reported on using MRI tumour localization at the 1.5 T MRI-linac to drive MLC-tracking approaches. Glitzner et al. [26] used a simple COM approach to demonstrate MLC-tracking feasibility using a phantom, and Uijtewaal et al. [50,51] used template matching based on cross-correlation in a similar setup. It should be noted that while template matching is a fast and robust localization algorithm, it cannot take into account target deformation and rotations.

Clinically, since early 2023 a gating functionality based on the tracking method by Keiper et al. [52] has been released on the 1.5 T MRI-linac. The approach requires a daily reference 3D MRI with a target segmentation along with sagittal and coronal 2D cine MRI pairs. Template image pairs are generated from 60 images acquired over 12 s before beam-on and these are registered to extracted slices from the 3D-MRI. Subsequently the target mask is moved by rigid registration using template matching within 50 ms, as explained in greater detail in Jassar et al. [53]. The clinical use of the 0.35 T MRI-linac's tumour-tracking-based gating functionality was reported in 2017 [23,54] and is applicable to lesions of differing contrast including outside the lung. The target localization in these papers was based on multi-scale DIR of a key frame which minimizes a cross-correlation cost function using gradient magnitude regularization. The problem is solved in both directions to ensure robustness of the obtained 2D deformation vector fields (DVF). A detailed description can be found in Appendix A of Feng et al. [55], where a runtime of 20 ms is reported for 100 × 100 pixels. In the latest version, several trackers are implemented with their own cost functions and regularization, and can operate locally and globally, as described in Palacios et al. [56]. The final localized target is the result of the consensus among the trackers. Four algorithms are available, combining different numbers of local and global trackers. The localization also reports a confidence value calculated as the intersection of all tracker contours over the union of all tracker contours, weighted by a function of the cross correlation of the deformed image from each tracker with the target frame.

DIR for target localization can also be achieved using AI, the approaches described in the next section having only been investigated at a

A.1. Topical review: Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects 109

E. Lombardo et al.

Radiotherapy and Oncology 190 (2024) 109970

research level. The key challenges for such approaches are a low additional latency and the availability of training and evaluation sets, all methods being currently available only at a research level. Most approaches employ convolutional neural networks (CNNs) for the task. For instance, Terpstra et al. [21] showed that a CNN outperformed a conventional optical-flow algorithm for 2D DVF estimation in abdominal cancer patients and extended their work to a 3D lung dataset using a cascade CNN at multiple resolutions [57]. Hunt et al. [58] employed a VoxelMorph-based CNN to rapidly produce DFVs, showing an improvement over (slow) B-spline and Demons approaches. A ResNet-18 model with self-supervised learning was found to outperform optical flow approaches on three different datasets (cardiac cine MRI from

diagnostic scanner, abdominal cine MRI from 1.5 T MRI-linac and echocardiography) by Frueh et al. [59]. Real-time estimation of time-resolved 3D MRI was implemented by Wei et al. [41], who employed a multi-branch CNN. Specifically, their model takes the daily setup 3D MRI and the current coronal 2D cine MRI frame and outputs a 3D DVF which is used to warp the setup MRI in approximately 100 ms. As ground truth real-time 3D cine MRI with high resolution is not available, the authors validated their approach by comparing a slice of the estimated 3D MRI with 2D slices acquired orthogonally to the coronal cine MRI frame. Shao et al. [60] also proposed a U-Net based architecture for 3D DVF estimation based on under-sampled MRI, demonstrating generalisability of liver and cardiac patients, though reporting a latency slightly

Table 1

A summary of AI methods for target localization on cine MRI. For each publication, a categorisation of the approach is provided, alongside the details of the data set and metrics used and a condensed review of their performance (results and latency). Note that both DIR and auto-contouring methods can by definition account for deformations/rotations of the target.

Author	Approach	Network	Treatment site	Dataset size	Metrics	Results	Inference time (ms)
Terpstra et al. (2020) [21]	DIR (2D)	SpyNET (CNN cascade)	Abdomen (including liver, kidney, pancreas)	In total, 135 patients. With data augmentation, this yielded 130,471 DVFs for training and validation, and 28,275 for testing	Structural Similarity (SSIM), Root mean squared error (RMSE)	For under-sampling factors of up to 25, SSIM > 0.8, RMSE up to 1 mm. SpyNET performs better in both metrics for under-sampling factors > 10 compared to optical flow method.	60
Frueh et al. (2022) [59]	DIR (2D)	ResNet-18 (CNN) with self-supervised learning	Heart, abdomen	Cardiac: 1140 patients, 372,810 frames. Abdominal: 50 patients, 165,261 frames	Forward and backward propagated DSC	Cardiac – forward DSC: 0.89, backward DSC: 0.90 Abdominal – forward DSC: 0.95, backward DSC: 0.96	Not reported
Hunt et al. (2023) [58]	DIR (2D)	VoxelMorph (CNN)	Liver, lung, pancreas	In total, 21 patients with about 21 h of cine MRI	RMSE	CNN: 0.032 B-spline: 0.040 Demons: 0.036	8
Terpstra et al. (2021) [57]	DIR with under-sampling (3D)	Multiresolution CNN	Lung	In total, 27 patients. Training on 17 patients equalling 2108 DVFs	End-point-error (EPE)	EPE: (1.87 ± 1.65) mm	200
Shao et al. (2022) [60]	DIR with under-sampling (3D)	U-Net (CNN)	Heart, liver	8 cardiac patients and 9 liver patients	Dice similarity coefficient (DSC), COM error	DSC between (0.88 ± 0.03) and (0.89 ± 0.02) , COM error between (1.29 ± 1.22) mm and (1.01 ± 0.86) mm	600
Wei et al. (2023) [41]	DIR (3D)	Multi-branch CNN	Liver	In total, 8 patients	DSC, Hausdorff distance (HD)	Mean DSC between (96.5 ± 1.1) and (98.7 ± 0.6) across a range of imaging planes. Mean HD between (3.0 ± 1.9) mm and (2.1 ± 1.1) mm across a range of imaging planes	100
Huttinga et al. (2023) [90]	DIR with uncertainty estimation (3D)	Gaussian Process	Abdomen	Digital phantom and 5 healthy volunteers	EPE	EPE: 75th percentile below 1 mm when applying rejection criterion based on uncertainty estimation. Almost 6 mm without application of the rejection criterion	0.1
Yun et al. (2015) [22]	Auto-contouring (2D)	PCNN	Lung	4 patients, 650 images per patient	DSC, HD, centroid error	Mean DSC: 0.87–0.92 Mean HD: 3.12–4.35 mm Mean centroid error: 1.03–1.35 mm	20
Yun et al. (2016) [62]	Auto-contouring (2D)	PCNN + intelligent parameter optimization	Lung	6 patients, 130 images per patient	DSC, HD, centroid error.	Compared with Yun et al. (2015) method, DSC increased up to 3 %, HD decreased up to 1.9 mm and centroid error decreased up to 0.5 mm	20
Yip et al. (2018) [63]	Auto-contouring (2D)	PCNN + intelligent parameter optimization	Lung	6 patients, 130 images per patient	DSC, HD	Mean DSC: (0.87 ± 0.04) . Mean HD: (4.8 ± 1.7) mm. Both were similar to the inter-observer variability	20
Bourque et al. (2016) [45]	Auto-contouring (2D)	Particle filtering (sequential Monte Carlo)	Lung	4 patients, 240 images per patient	DSC, precision, recall, centroid error	Mean DSC: 0.89–0.91 Mean precision: 0.88–0.91 Mean recall: 0.89–0.95 Mean centroid error: 0.6–2.0 mm	2 ms every 500 particles plus 14 ms
Friedrich et al. (2021) [64]	Auto-contouring with under-sampling (2D)	U-Net (CNN)	Liver	2 patients (3 tumours), 50 reconstructed images per tumour	DSC, HD, contour distance	When 30 out of a possible 128 spokes are used for reconstruction, DSC ≥ 0.83 , HD ≤ 1.4 pixel and mean contour distance ≤ 0.5 pixel. Outperforming B-spline in all metrics	53

longer than 500 ms. On the other hand, Huttinga et al. [90] recently showed how a motion model based on Gaussian Processes can be used to obtain 3D DVFs from three mutually orthogonal k-space read outs at a frame rate up to 69 Hz, including data acquisition and reconstruction. Unlike CNN-based approaches, general adversarial networks (GANs) can be used to perform DIR in multi-modality workflows which include MRI. Whilst demonstrating a greater generalisability across treatment sites, current GAN implementations introduce greater latency, with an inference time of 1.3 s for 3D (static) MRI registration [61] or about 3 s for unsupervised CT-MRI DIR [108] and are thus not yet suited for real-time target localization.

AI approaches for automatic segmentation were also investigated as a method of localizing targets in real-time, the short inference time requirement posing the main challenge compared to offline auto-segmentation. Yun et al. propose a pulse-coupled neural network (PCNN) trained on at least two full breathing cycles in four patients imaged with 3 T MRI, where each input image pixel corresponds to a neuron in the network, for localizing lung tumours [22]. An extension of this work, which added intelligent parameter optimization [62], was evaluated against manual contour accuracy and was reported to be similar to inter-observer variability by Yip et al. [63]. Friedrich et al. [64] noted that their U-Net-based auto-segmentation outperforms B-spline DIR for localizing liver tumours in under-sampled cine MRI over a range of under-sampling factors; however only two patients were used.

The AI methods discussed above are summarised in Table 1.

Motion prediction

Motion prediction must take place in real-time and be fast to avoid the introduction of further latency. While for certain tumour sites motion prediction is not possible due to the random nature of motion, e.g., for prostate cancer [65], for all sites affected by respiration such as lung, liver or pancreas, prediction is possible (even for irregular breathing motion) and relevant, as motion typically ranges from 1 cm to 4 cm [66].

Over the past two decades, several motion prediction algorithms have been implemented and compared in-silico [19,67]. Most of the studies focused on the prediction of the 1D, 2D or 3D target COM positions, meaning that developments were independent of the imaging technique used for localization, provided that similar accuracy and latency were guaranteed. The conventional algorithm which has been reported to be the most successful in predicting future target positions is linear regression (LR). An advantage of the LR is that an analytical solution to compute its parameters exists [19], which requires little time to compute and has led to the development of improved LR models which can be continuously updated based on current motion during treatment [67].

While LR models are very promising for latencies up to 250 ms, it has been shown that for larger forecasts they are outperformed by AI models [68,69]. Specifically, long short-term memory (LSTM) networks, which are a type of AI specifically designed to deal with sequential inputs have been proven successful for the prediction of target COM positions 500 ms and 750 ms into the future [70], even in challenging cases such as motion with breath-holds or irregular free-breathing [69]. MRI is particularly suited for such challenging cases as it provides higher quality localization compared to x-rays, which in turn leads to better prediction performance. Another strength of LSTMs is that they can be extended via the use of convolutions to the prediction of target contours [71] or cine MRI frames [72], thus allowing to account for rotations or deformations. In particular Romaguera et al. [72] developed a convolutional LSTM for the prediction of 2D DVFs up to 1.6 s into the future, which achieved submillimetre vessel position prediction accuracy on (diagnostic) cine MRI.

Currently, no clinical MRI-linac is equipped with motion prediction. This is likely related to the fact that the gain in accuracy from prediction depends on the beam adaptation strategy, MLC-tracking being the one benefitting the most but currently available only on research machines

(see subsection *Beam adaptation*). The robotic linac system is currently the only example of usage of motion prediction in clinical practice and uses an LR model [111]. At a research level, motion prediction has been experimentally shown to improve MLC-tracking accuracy during MRI-guided irradiations in phantom studies. Yun et al. [27] demonstrated for sinusoidal motion on a prototype 0.2 T MRI-linac that MLC-tracking with motion prediction using an ANN leads to minimal difference in beam width and beam penumbra compared to the static scenario. Uijtewaal et al. [50] showed on a prototype 1.5 T MRI-linac, that MLC-tracking with motion prediction using a continuously updated LR model leads to accurate dose delivery in the presence of sinusoidal motion with baseline drifts compared to the reference dose. In a follow-up study with a more complex 2D/4D hybrid MRI methodology, the authors showed for a sinusoidal and a patient-derived motion trace that MLC-tracking with LR based motion prediction was still possible with a latency of up to 850 ms [73]. However, they observed a decrease in accuracy possibly caused by this large latency. Recently, Lombardo et al. [100] showed in a prototype 1.0 T MRI-linac that a continuously updated LSTM significantly outperformed LR for MLC-tracking with eight different patient-derived motion traces. The authors also compared the LSTM to a scenario without any motion prediction and found the geometrical error to be nearly halved thanks to a better MLC to target alignment.

Beam adaptation

MRI-linacs have three systems of adaptation to improve the beam-target alignment based on the MRI guidance information. These are beam gating, trailing, and MLC-tracking. These systems are all independently controlled and can therefore be used in parallel. For example, in MLC-tracking, beam gating can be enabled if the leaves cannot track the target and indeed is a patient safety feature [74].

Beam gating allows treatment of the target in a fixed position in space at the cost of increased treatment times. The beam is automatically switched on as soon as the tumour or a surrogate volume is in a predefined position and automatically switched off as soon as the target is outside of a predefined boundary. This can be performed in free breathing or using breath-hold techniques (e.g., for lung treatments). Using repeated breath-holds during gated treatment has been shown to lead to higher dosimetric accuracy compared to free-breathing gating in a phantom study by Charters et al. [96]. However, while gating can achieve a dose conformity close to the static scenario, it leads to lower duty cycle efficiency (the ratio between beam-on time and treatment duration) and hence longer treatments, with values of 20 % and 55 % having been reported in the literature [32,33]. Direct tumor position visual feedback has been proposed to increase the duty cycle efficiency, with Kim et al. reporting an increase from 44 % to 48 % in a clinical comparison with 13 patients [95]. For sites which are not affected by respiratory motion such as the prostate, it has been shown that while adaptation to inter-fractional changes plays an important role [97], real-time adaptation to intra-fractional movement of the prostate via gating has a negligible dosimetric effect (interruptions of the treatment due to shifts excluded) [35,99]. To account for target localization uncertainty, some systems allow to define a confidence level which, if unmet, leads to automatic switching off of the beam. To reduce this uncertainty, authors have also proposed the usage of contrast agents to enhance the visibility of, for example, liver lesions [94]. The gating latency can be divided in two components: detection of the target inside/out of the boundary and time between detection and beam on/off. The time to switch the beam on or off is the same and relatively low, with typical values around 10 ms for the 0.35 T system. Beam on or off is enforced by switching of the radiofrequency pulse of the beam with pulse repetition rates of 100–120 Hz. The beam-on latency is usually larger than the beam-off latency because additional data is acquired before switching on to ensure the target is in the desired position. While the beam-on latency contributes to the reduction of the duty cycle efficiency, the beam-off latency may

A.1. Topical review: Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects

111

E. Lombardo et al.

Radiotherapy and Oncology 190 (2024) 109970

lead to a less accurate irradiation and is thus of more concern.

In an early study carried out at the 0.35 T MRI-Cobalt system, Lamb et al. [75] measured a beam-off gating latency of 436 ms (4 Hz) using a motion phantom with a radiochromic film insert. Similar results were observed by Green et al. [23], who measured an average beam-off latency of 394 ms with a range of 246–527 ms (4 Hz). As for the 0.35 T MRI-linac version, Kim et al. [31] evaluated both beam-off and beam-on gating latency considering four localization algorithms and two imaging frequencies. They measured beam-off latencies of 128–243 ms at 4 Hz and 47–302 ms at 8 Hz and beam-on latencies of 342–464 ms at 4 Hz and 664–785 ms at 8 Hz imaging.

An adaptation solution with potential for drifting targets is a mid-position delivery in combination with trailing [76]. Trailing is a technique in which the beam aperture is continuously adjusted according to the target's last available time-averaged position, limiting the impact of tumour baseline drift. The advantage of this approach is the simplicity since it requires a low imaging frequency, and it is insensitive to latency. Uijtewaal et al. [50] experimentally demonstrated that for central and peripheral lung SBRT, MRI-guided trailing was highly effective in mitigating baseline motion while MLC-tracking maximized the sparing of healthy tissue. In their study, the PTV for MLC-tracking was obtained by a 3 mm isotropic expansion of the GTV while the PTV for trailing was obtained by adding anisotropic margins of 3 mm left-right, 4 mm anterior-posterior and 5 mm superior-inferior direction to the GTV based on the anticipated motion. More recently, Grimbergen et al. reported on the first seven patients treated with gating in combination with trailing on the 1.5 T system [98]. They showed that the employed beam adaptation strategy improved congruence to the planned static dose with duty cycle efficiencies ranging from 41 % to 93 %.

MLC-tracking consists in the continuous realignment of the MLC positions to compensate for tumour motion [93]. The feasibility of MRI-guided MLC-tracking for clinically acceptable lung SBRT was proven with an in-silico study by Menten et al. [77] showing that MLC-tracking increases the dose delivery accuracy, allowing to use smaller treatment margins. While being the most complex of the discussed beam adaptation strategies, it is also the most flexible: the leaf positions can be optimized to compensate for geometrical differences with respect to the static plan, e.g., when the current target position is different from the expected target position (geometry-optimized MLC-tracking) or to compensate for over/under-dosage compared to the static plan (dose-optimized MLC-tracking).

System latency is more problematic during MLC-tracking because it may induce residual tracking errors, causing a missing of the target and unplanned irradiation of the healthy tissues/organs at risks. Although MLC-tracking consists of several steps that all introduce latency, the main contributor to the end-to-end latency is the relatively low imaging frequency of MRI acquisition process. Glitzner et al. [26] found for the isolated MLC a latency of 20.67 ms, while the end-to-end system latency was 347.45 ms at 4 Hz imaging and 204 ms at 8 Hz on the 1.5 T MRI-linac. Their study represented the first experimental demonstration of geometry-optimized MRI-guided MLC-tracking on the 1.5 T MRI-linac for a single gantry angle with a conformal treatment field. In 2021, Uijtewaal et al. [50] confirmed these results in an experimental phantom study for lung SBRT IMRT plans, as described above. Using films inserted in the moving phantom, they could show that compensating for the latency using motion prediction resulted in an increase in 2 %/2mm gamma pass rates from 83 % (no prediction) to 100 % (continuously updated LR) compared to the reference static plan. In a follow-up study, the authors also provided a first experimental demonstration of VMAT combined with MRI-guided MLC-tracking for a range of lung SBRT fractionation schemes [51]. Geometry-optimized MLC-tracking combined with VMAT delivery resulted in highly conformal dose distributions with similar target coverage as their respective static reference.

Adaptation becomes more critical when two or multiple targets that move independently need to be treated simultaneously (e.g., primary lung tumour plus mediastinal lymph nodes). Liu et al. [17]

experimentally showed geometry-optimized MLC-tracking with simultaneous monitoring and compensation of independent motion of 2 targets on a 1.0 T prototype MRI-linac. This technology would allow treatment margins to be reduced if multiple targets are simultaneously treated.

Dose-optimized MLC-tracking has currently only been shown in-silico. Mejnertsen et al. [78] showed using a simplified dose calculation algorithm that dose-optimized MLC-tracking achieved significantly lower failure rates than geometry-optimized MLC-tracking for VMAT deliveries in a prostate cancer patient monitored with intra-fractional kilovoltage imaging. In a follow-up simulation study, Hewson et al. [79] confirmed the dosimetric advantage when performing multi-target dose-optimized MLC-tracking over geometry-optimized when irradiating moving prostate and static lymph node targets.

For geometry-optimized MLC-tracking instructions can all be performed with analytic solutions of the order of milliseconds, even for solving an iterative MLC optimization problem, e.g., the Ruan algorithm [80]. It could be argued that AI applications are not needed for geometry-optimized beam adaptation given there are alternative analytical solutions. However, when performing dose-optimized adaptation, there is the additional computation of the patient anatomic state and the optimal dose to give to the patient within the constraints of the adaptation system. This complexity lends itself to AI solutions, particularly if the optimization includes future estimates of the patient state. An example of an AI model calculating the 3D dose delivered by individual MLC shapes on the current patient anatomy is the CNN proposed by Kontaxis et al. [83]. However, their model takes 900 ms for input generation and 600 ms for inference, making it too slow for real-time dose calculation. An ideal candidate could be the transformer based model by Xiao et al. [101], which computes the 3D beam dose in 310 ms and can account for magnetic fields. On a testing set of 20 patients treated with the 1.5 T MRI-linac, the authors obtained 3 %/2mm gamma pass rates larger than 98 % compared to ground truth Monte Carlo simulations.

Discussion

Envisioned workflow

Given the current MRI-guided intra-fractional motion management workflow described above, we would like to describe an envisioned workflow which in our opinion has the greatest potential for dosimetric accuracy without reducing the duty cycle efficiency of treatments (see Fig. 2).

In terms of imaging, multi-slice 2D cine MRI with a frequency of 2–3 Hz as implemented by vendors [53], combined with fast AI based 2D to 3D motion estimation (about 100 ms) [41], could be used to obtain 3D DVFs to be applied to the setup 3D MRI in real-time. Alternatively, AI reconstruction [81] and super-resolution [82] methods could be tailored to make 3D cine MRI suitable for respiratory resolved imaging of the abdomen and thorax. Real-time CNN-based 3D motion estimation from 3D cine MRI as proposed by Terpstra et al. [57] could then be used to obtain the 3D DVFs within 200 ms. To compensate for the latencies introduced during imaging, motion estimation and the subsequent dose calculation and MLC-tracking steps, prediction of future DVFs with LSTMs could be leveraged similarly to the study by Romaguera et al. [72] described above but extended to 3D. The predicted DVFs could then be applied to the setup 3D CT (currently generated by vendors using image registration based on the setup 3D MRI and used to adapt the plan to the anatomy of the day) to generate an estimated 3D CT with target/OARs structures in the current motion state. The fast transformer-based dose calculation framework proposed by Xiao et al. [101], could then be used to derive the 3D dose delivered by individual MLC shapes on the current patient anatomy. This dose should then be warped (using the inverse DVF obtained during motion estimation) to the static anatomy to accumulate and compare it with the planned dose. In the final step, the

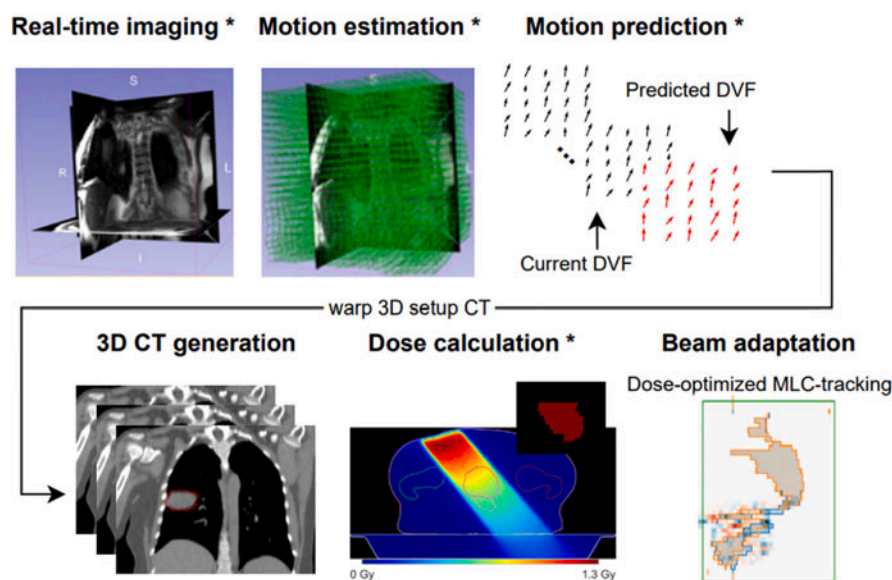


Fig. 2. Envisioned workflow of real-time motion management in MRIGRT. First, multiple 2D cine MRI slices are acquired in real-time. A motion estimation model is then used to compute 3D DVFs with respect to the 3D MRI acquired prior to treatment start (not shown). Motion prediction is employed to compensate for the overall system latency by obtaining a predicted DVF. The predicted DVF is used to generate an estimated 3D CT (with contours) by warping the 3D setup CT obtained prior to treatment start (not shown). The dose on the estimated 3D CT is then calculated in real-time, accumulated and used to perform MLC-tracking based on computed over/under-dosage with respect to the static plan. (*) Asterisks denote where AI can play a role. Adapted with permission from [79,83] and [85].

difference between the accumulated delivered dose and the planned dose could be given as input to an MLC-tracking algorithm similar to the one proposed by Mejnertsen et al. [78], which optimizes the MLC aperture in about 40 ms to actively target regions of underdose while avoiding overdose. Alternatively, an open loop adaptive sequencing planning system as the one proposed by Kontaxis et al. could be used [102]. Using AI instead of simplified dose calculations would have the advantage that also changes in density such as organ filling could potentially be taken into account.

An advantage of the envisioned workflow is that little to no changes to the hardware of current clinical MRI-linacs are needed. A challenge is posed by the usage of many different AI models for different steps and the resulting necessity for accurate end-to-end QA tests. To reduce the number of models, some of the steps could be combined by using a single AI model. Such a model could for instance perform DVF estimation and prediction at the same time.

QA and AI uncertainty quantification

One of the main physical questions to be addressed in the context of the modern MRIGRT systems is the ability of these machines to effectively adapt the beam delivery to tumour motion. Both clinical MRIGRT systems are now equipped with gating, the 1.5 T MRI-linac being oriented to MLC-tracking in the future. In both cases, studies to evaluate the software ability in localizing the lesion and the physical latency of the machine have been reported in the literature. As regards the 0.35 T MRI-linac, three experiences have been reported, combining dynamic phantoms with 1D or 2D radiation detectors (ion chambers, gafchromic films) [23,31,75]. To determine whether the dosimetric accuracy of the used beam adaptation technique is preserved over time, end-to-end tests of the real-time adaptive workflow should be performed. Stark et al. [84] developed such an end-to-end test for gating at a 0.35 T MRI-linac using clinically realistic treatment times and gating windows.

For AI in general and especially target localization and motion prediction models, it is important to understand that the uncertainty in an AI model can be divided into aleatoric (statistical) uncertainty, which

refers to the notion of randomness, and epistemic (systematic) uncertainty, which refers to uncertainty caused by a lack of knowledge [91]. As opposed to aleatoric uncertainty, epistemic uncertainty can in principle be reduced based on additional information. For example, the target localization uncertainty could be improved by adding more data (e.g., with longer image acquisition times) but this would increase the latency. Of clinical relevance could be to quantify epistemic uncertainty in-real time and use it to gate the treatment, similarly to what is currently done with the confidence value on the 0.35 T MRI-linac [56].

Conclusion

In this work, we have described the current status, both clinically and at a research level, of real-time intra-fractional motion management in MRIGRT. Based on recent scientific developments, we have then outlined an envisioned workflow representing the most accurate treatment from a technological point of view. In the proposed workflow, motion is visualized with multiple 2D or 3D cine MRI, target/OAR structures are obtained on an estimated 3D CT and dose-optimized MLC-tracking is applied to minimize the difference between delivered and planned dose in real-time. The proposed workflow is challenging especially due to the real-time requirements; however, we think that by leveraging AI in the different steps of the workflow, this ambition could become clinical reality. In future work, the authors aim to create a shared dataset for AI model comparison (e.g., unlabelled and labelled cine MRIs for target localization) and to define common end-to-end tests for real-time adaptive MRIGRT.

CRediT authorship contribution statement

Elia Lombardo: Conceptualization, Methodology, Writing – original draft. **Jennifer Dhont:** Conceptualization, Methodology, Writing – original draft. **Denis Page:** Methodology, Writing – review & editing. **Cristina Garibaldi:** Methodology, Writing – review & editing. **Luise A. Künzel:** Methodology, Writing – review & editing. **Coen Hurkmans:** Methodology, Writing – review & editing. **Rob H.N. Tijssen:**

A.1. Topical review: Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects

113

E. Lombardo et al.

Radiotherapy and Oncology 190 (2024) 109970

Methodology, Writing – review & editing. **Chiara Paganelli**: Methodology, Writing – review & editing. **Paul Z.Y. Liu**: Methodology, Writing – review & editing. **Paul J. Keall**: Methodology, Writing – review & editing. **Marco Riboldi**: Methodology, Writing – review & editing. **Christopher Kurz**: Methodology, Writing – review & editing. **Guillaume Landry**: Methodology, Writing – review & editing. **Davide Cusumano**: Supervision, Validation, Writing – review & editing. **Marco Fusella**: Supervision, Validation, Writing – review & editing. **Lorenzo Placidi**: Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

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

References

- [1] Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol* 2017;62:L41. <https://doi.org/10.1088/1361-6560/aa9517>.
- [2] Raaymakers BW, Lagendijk JJW, Overweg J, et al. Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol* 2009;54:N229. <https://doi.org/10.1088/0031-9155/54/12/N01>.
- [3] Boldrini L, Cusumano D, Cellini F, Azario L, Mattiucci GC, Valentini V. Online adaptive magnetic resonance guided radiotherapy for pancreatic cancer: state of the art, pearls and pitfalls. *Radiat Oncol* 2019;14:71. <https://doi.org/10.1186/s13014-019-1275-3>.
- [4] Hunt A, Hansen VN, Oelfke U, Nill S, Hafeez S. Adaptive Radiotherapy Enabled by MRI Guidance. *Clin Oncol (R Coll Radiol)* 2018;30:711–9. <https://doi.org/10.1016/j.clon.2018.08.001>.
- [5] Corradini S, Alongi F, Andrasschke N, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. *Radiat Oncol* 2019;14:92. <https://doi.org/10.1186/s13014-019-1308-y>.
- [6] van der Heide UA, Houweling AC, Groenendaal G, Beets-Tan RGH, Lambin P. Functional MRI for radiotherapy dose painting. *Magn Reson Imaging* 2012;30:1216–23. <https://doi.org/10.1016/j.mri.2012.04.010>.
- [7] Thorwarth D. Functional imaging for radiotherapy treatment planning: current status and future directions—a review. *Br J Radiol* 2015;88:20150056. <https://doi.org/10.1259/bjr.20150056>.
- [8] Tomaszewski MR, Latifi K, Boyer E, et al. Delta radiomics analysis of Magnetic Resonance guided radiotherapy imaging data can enable treatment response prediction in pancreatic cancer. *Radiat Oncol* 2021;16:237. <https://doi.org/10.1186/s13014-021-01957-5>.
- [9] Dhont J, Vandemeulebroucke J, Burghelma M, et al. The long- and short-term variability of breathing induced tumor motion in lung and liver over the course of a radiotherapy treatment. *Radiat Oncol* 2018;126:339–46. <https://doi.org/10.1016/j.radonc.2017.09.001>.
- [10] Cusumano D, Dhont J, Boldrini L, et al. Predicting tumour motion during the whole radiotherapy treatment: a systematic approach for thoracic and abdominal lesions based on real time MR. *Radiat Oncol* 2018;129:456–62. <https://doi.org/10.1016/j.radonc.2018.07.025>.
- [11] Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a. *Med Phys* 2006;33:3874–900. <https://doi.org/10.1118/1.2349696>.
- [12] Seppenwoolde Y, Shirato H, Kitamura K, et al. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *Int J Radiat Oncol Biol Phys* 2002;53:822–34. [https://doi.org/10.1016/s0360-3016\(02\)02803-1](https://doi.org/10.1016/s0360-3016(02)02803-1).
- [13] Bertholet J, Knopf A, Eiben B, et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Phys Med Biol* 2019;64(15):15TR01. <https://doi.org/10.1088/1361-6560/ab2ba8>.
- [14] Kiser KJ, Smith BD, Wang J, Fuller CD. “Après Mois, Le Déluge”: preparing for the coming data flood in the MRI-guided radiotherapy era. *Front Oncol* 2019;9. <https://doi.org/10.3389/fonc.2019.00983>.
- [15] Cusumano D, Boldrini L, Dhont J, et al. Artificial Intelligence in magnetic resonance guided radiotherapy: medical and physical considerations on state of art and future perspectives. *Phys Med* 2021;85:175–91. <https://doi.org/10.1016/j.jeump.2021.05.010>.
- [16] Cervino LI, Du J, Jiang SB. MRI-guided tumor tracking in lung cancer radiotherapy. *Phys Med Biol* 2011;56:3773–85. <https://doi.org/10.1088/0031-9155/56/13/003>.
- [17] Liu PZY, 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;47:6440–9. <https://doi.org/10.1002/mp.14536>.
- [18] Dhont J, Verellen D, Poels K, et al. Feasibility of markerless tumor tracking by sequential dual-energy fluoroscopy on a clinical tumor tracking system. *Radiat Oncol* 2015;117:487–90. <https://doi.org/10.1016/j.radonc.2015.08.021>.

- [19] 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. <https://doi.org/10.1088/0031-9155/56/16/015>.
- [20] Seregni M, Paganelli C, Lee D, et al. Motion prediction in MRI-guided radiotherapy based on interleaved orthogonal cine-MRI. *Phys Med Biol* 2016;61:872. <https://doi.org/10.1088/0031-9155/61/2/872>.
- [21] Terpstra ML, Maspero M, d’Agata F, et al. Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time MRI-guided radiotherapy. *Phys Med Biol* 2020;65:155015. <https://doi.org/10.1088/1361-6560/ab9358>.
- [22] Yun J, Yip E, Gabos Z, Wachowicz K, Rathee S, Fallone BG. Neural-network based autocontouring algorithm for intrafractional lung-tumor tracking using Linac-MR. *Med Phys* 2015;42:2296–310. <https://doi.org/10.1118/1.4916657>.
- [23] Green OL, Rankine LJ, Cai B, et al. First clinical implementation of real-time, real anatomy tracking and radiation beam control. *Med Phys* 2018. <https://doi.org/10.1002/mp.13002>.
- [24] Kurz C, Buizza G, Landry G, et al. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol* 2020;15:93. <https://doi.org/10.1186/s13014-020-01524-4>.
- [25] Klüter S. Technical design and concept of a 0.35 T MR-Linac. *Clin Transl. Radiat Oncol* 2019;18:98–101. <https://doi.org/10.1016/j.ctro.2019.04.007>.
- [26] 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. <https://doi.org/10.1088/1361-6560/ab2667>.
- [27] Yun J, Wachowicz K, Mackenzie M, Rathee S, Robinson D, Fallone BG. First demonstration of intrafractional tumor-tracked irradiation using 2D phantom MR images on a prototype linac-MR. *Med Phys* 2013;40:051718. <https://doi.org/10.1118/1.4802735>.
- [28] Paganelli C, Whelan B, Peroni M, et al. MRI-guidance for motion management in external beam radiotherapy: current status and future challenges. *Phys Med Biol* 2018;63:22TR03. <https://doi.org/10.1088/1361-6560/aaebcf>.
- [29] Stemkens B, Paulson ES, Tijssen RHN. Nuts and bolts of 4D-MRI for radiotherapy. *Phys Med Biol* 2018;63:21TR01. <https://doi.org/10.1088/1361-6560/aae56d>.
- [30] Keall PJ, Brighi C, Glide-Hurst C, et al. Integrated MRI-guided radiotherapy - opportunities and challenges. *Nat Rev Clin Oncol* 2022;19:458–70. <https://doi.org/10.1038/s41571-022-00631-3>.
- [31] Kim T, Lewis B, Lotey R, Barberi E, Green O. Clinical experience of MRI4D QUASAR motion phantom for latency measurements in 0.35T MR-LINAC. *J Appl Clin Med Phys* 2020;22:128–36. <https://doi.org/10.1002/acm2.13118>.
- [32] Eze C, Lombardo E, Nierler 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;17:43. <https://doi.org/10.1186/s13014-022-02011-8>.
- [33] Ehrbar S, Käser SB, Chamberlain M, et al. MR-guided beam gating: Residual motion, gating efficiency and dose reconstruction for stereotactic treatments of the liver and lung. *Radiat Oncol* 2022;174:101–8. <https://doi.org/10.1016/j.radonc.2022.07.007>.
- [34] Hu P, Li X, Liu W, et al. Dosimetry impact of gating latency in cine magnetic resonance image guided breath-hold pancreatic cancer radiotherapy. *Phys Med Biol* 2022;67. <https://doi.org/10.1088/1361-6560/ac53e0>.
- [35] Xiong Y, Rabe M, Nierler L, et al. Assessment of intrafractional prostate motion and its dosimetric impact in MRI-guided online adaptive radiotherapy with gating. *Strahlenther Onkol* 2023;199:544–53. <https://doi.org/10.1007/s00066-022-02005-1>.
- [36] Borman PTS, Tijssen RHN, Bos C, Moonen CTW, Raaymakers BW, Glitzner M. Characterization of imaging latency for real-time MRI-guided radiotherapy. *Phys Med Biol* 2018;63:155023. <https://doi.org/10.1088/1361-6560/aad2b7>.
- [37] Mickevicius NJ, Paulson ES. Simultaneous orthogonal plane imaging. *Magn Reson Med* 2017;78:1700–10. <https://doi.org/10.1002/mrm.26555>.
- [38] Grimbergen G, Pötgens GG, Eijkelenkamp H, Raaymakers BW, Intven MPW, Meijer GJ. Feasibility of delivered dose reconstruction for MR-guided SBRT of pancreatic tumors with fast, real-time 3D cine MRI. *Radiat Oncol* 2023;182. <https://doi.org/10.1016/j.radonc.2023.109506>.
- [39] Zeng G, Guo Y, Zhan J, et al. A review on deep learning MRI reconstruction without fully sampled k-space. *BMC Med Imaging* 2021;21:195. <https://doi.org/10.1186/s12880-021-00727-9>.
- [40] 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* 2022;66:055006. <https://doi.org/10.1088/1361-6560/abc937>.
- [41] Wei R, Chen J, Liang B, Chen X, Men K, Dai J. Real-time 3D MRI reconstruction from cine-MRI using unsupervised network in MRI-guided radiotherapy for liver cancer. *Med Phys* 2023;50:3584–96. <https://doi.org/10.1002/mp.16141>.
- [42] Shi X, Diwanji T, Mooney KE, et al. Evaluation of template matching for tumor motion management with cine-MR images in lung cancer patients. *Med Phys* 2014;41:052304. <https://doi.org/10.1118/1.4870978>.
- [43] Mazur TR, Fischer-Valuck BW, Wang Y, Yang D, Mutic S, Li HH. SIFT-based dense pixel tracking on 0.35 T cine-MR images acquired during image-guided radiation therapy with application to gating optimization. *Med Phys* 2016;43:279–93. <https://doi.org/10.1118/1.4938096>.
- [44] Paganelli C, Lee D, Greer PB, Baroni G, Riboldi M, Keall P. Quantification of lung tumor rotation with automated landmark extraction using orthogonal cine MRI images. *Phys Med Biol* 2015;60:7165. <https://doi.org/10.1088/0031-9155/60/18/7165>.
- [45] Bourque AE, Bedwani S, Filion É, Carrier JF. A particle filter based autocontouring algorithm for lung tumor tracking using dynamic magnetic

- resonance imaging. *Med Phys* 2016;43:5161–9. <https://doi.org/10.1118/1.4961403>.
- [46] Zachiu C, Papadakis N, Ries M, Moonen C, de Senneville BD. An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs. *Phys Med Biol* 2015;60:9003. <https://doi.org/10.1088/0031-9155/60/23/9003>.
- [47] Fast MF, Eiben B, Menten MJ, et al. Tumour auto-contouring on 2d cine MRI for locally advanced lung cancer: A comparative study. *Radiother Oncol* 2017;125:485–91. <https://doi.org/10.1016/j.radonc.2017.09.013>.
- [48] Tahmasebi N, Boulanger P, Yun J, Fallone BG, Punithakumar K. Tracking tumor boundary using point correspondence for adaptive radio therapy. *Comput Methods Programs Biomed* 2018;165:187–95. <https://doi.org/10.1016/j.cmpb.2018.08.002>.
- [49] Dhont J, Vandemeulebroucke J, Cusumano D, et al. Multi-object tracking in MRI-guided radiotherapy using the tracking-learning-detection framework. *Radiother Oncol* 2019;138:25–9. <https://doi.org/10.1016/j.radonc.2019.05.008>.
- [50] Uijtewaal P, Borman PTS, 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;48:1520–32. <https://doi.org/10.1002/mp.14772>.
- [51] Uijtewaal P, Borman PTS, 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;174:149–57. <https://doi.org/10.1016/j.radonc.2022.07.004>.
- [52] Keiper TD, Tai A, Chen X, et al. Feasibility of real-time motion tracking using cine MRI during MR-guided radiation therapy for abdominal targets. *Med Phys* 2020;47:3554–66. <https://doi.org/10.1002/mp.14230>.
- [53] 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 1.5T MR-Linac. *Med Phys* 2023;50:3103–16. <https://doi.org/10.1002/mp.16342>.
- [54] Fischer-Valuck BW, Henke L, Green O, et al. Two-and-a-half-year clinical experience with the world's first magnetic resonance image guided radiation therapy system. *Adv Radiat Oncol* 2017;2:485–93. <https://doi.org/10.1016/j.adro.2017.05.006>.
- [55] Feng Y, Kawrakow I, Olsen J, et al. A comparative study of automatic image segmentation algorithms for target tracking in MR-IGRT. *J Appl Clin Med Phys* 2016;17:441–60. <https://doi.org/10.1120/jacmp.v17i2.5820>.
- [56] Palacios MA, Gerganov G, Cobussen P, et al. Accuracy of deformable image registration-based intra-fraction motion management in Magnetic Resonance-guided radiotherapy. *Phys Imaging Radiat Oncol* 2023;26. <https://doi.org/10.1016/j.phro.2023.100437>.
- [57] Terpestra ML, Maspero M, Bruijnen T, Verhoeff JJC, Legendijk JJW, van den Berg CAT. Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks. *Med Phys* 2021;48:6597–613. <https://doi.org/10.1002/mp.15217>.
- [58] Hunt B, Gill GS, Alexander DA, et al. Fast deformable image registration for real-time target tracking during radiation therapy using cine MRI and deep learning. *Int J Radiat Oncol Biol Phys* 2023;115:983–93. <https://doi.org/10.1016/j.ijrobp.2022.09.086>.
- [59] Frueh M, Kuestner T, Nachbar M, Thorwarth D, Schilling A, Gatidis S. Self-supervised learning for automated anatomical tracking in medical image data with minimal human labeling effort. *Comput Methods Programs Biomed* 2022;225:107085. <https://doi.org/10.1016/j.cmpb.2022.107085>.
- [60] Shao HC, Li T, Dohopolski MJ, et al. Real-time MRI motion estimation through an unsupervised k-space-driven deformable registration network (KS-RegNet). *Phys Med Biol* 2022;67:135012. <https://doi.org/10.1088/1361-6560/ac762c>.
- [61] Fan J, Cao X, Wang Q, Yap PT, Shen D. Adversarial learning for mono- or multi-modal registration. *Med Image Anal* 2019;58:101545. <https://doi.org/10.1016/j.media.2019.101545>.
- [62] Yun J, Yip E, Gabos Z, Wachowicz K, Rathee S, Fallone BG. Improved lung tumor autocontouring algorithm for intrafractional tumor tracking using 0.5 T linac-MR. *Biomed Phys Eng Express* 2016;2:067004. <https://doi.org/10.1088/2057-1976/2/6/067004>.
- [63] Yip E, Yun J, Gabos Z, et al. Evaluating performance of a user-trained MR lung tumor autocontouring algorithm in the context of intra- and interobserver variations. *Med Phys* 2018;45:307–13. <https://doi.org/10.1002/mp.12687>.
- [64] Friedrich F, Hörner-Rieber J, Renkamp CK, et al. Stability of conventional and machine learning-based tumor auto-segmentation techniques using undersampled dynamic radial bSSFP acquisitions on a 0.35 T hybrid MR-linac system. *Med Phys* 2021;48:587–96. <https://doi.org/10.1002/mp.14659>.
- [65] Ballhausen H, Li M, Hegemann NS, Ganswindt U, Belka C. Intra-fraction motion of the prostate is a random walk. *Phys Med Biol* 2015;60:549–63. <https://doi.org/10.1088/0031-9155/60/2/549>.
- [66] Langen KM, Jones DTL. Organ motion and its management. *Int J Radiat Oncol Biol Phys* 2001;50:265–78. [https://doi.org/10.1016/S0360-3016\(01\)01453-5](https://doi.org/10.1016/S0360-3016(01)01453-5).
- [67] Jöhl A, Ehrbar S, Guckenberger M, et al. Performance comparison of prediction filters for respiratory motion tracking in radiotherapy. *Med Phys* 2020;47:643–50. <https://doi.org/10.1002/mp.13929>.
- [68] 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;6:51262–8. <https://doi.org/10.1109/ACCESS.2018.2869780>.
- [69] 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. <https://doi.org/10.1088/1361-6560/ac60b7>.
- [70] 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. <https://doi.org/10.1088/1361-6560/ab13fa>.
- [71] 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. <https://doi.org/10.1016/j.radonc.2023.109555>.
- [72] Romaguera LV, Plantefève R, Romero FP, Hébert F, Carrier JF, Kadoury S. Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks. *Med Image Anal* 2020;64:101754. <https://doi.org/10.1016/j.media.2020.101754>.
- [73] Keijnemans K, Borman PTS, Uijtewaal P, Woodhead PL, Raaymakers BW, Fast MF. A hybrid 2D/4D-MRI methodology using simultaneous multislice imaging for radiotherapy guidance. *Med Phys* 2022;49:6068–81. <https://doi.org/10.1002/mp.15802>.
- [74] Sawant A, Dieterich S, Svatos M, Keall P. Failure mode and effect analysis-based quality assurance for dynamic MLC tracking systems. *Med Phys* 2010;37:6466–79. <https://doi.org/10.1118/1.3517837>.
- [75] Lamb JM, Ginn JS, O'Connell DP, et al. Dosimetric validation of a magnetic resonance image gated radiotherapy system using a motion phantom and radiochromic film. *J Appl Clin Med Phys* 2017;18:163–9. <https://doi.org/10.1002/acm2.12088>.
- [76] Fast M, van de Schoot A, van de Lindt T, Carbaat C, van der Heide U, Sonke JJ. Tumor trailing for Liver SBRT on the MR-Linac. *Int J Radiat Oncol*Biophys* 2019;103:468–78. <https://doi.org/10.1016/j.ijrobp.2018.09.011>.
- [77] Menten MJ, Fast MF, Nill S, Kamerling CP, McDonald F, Oelfke U. Lung stereotactic body radiotherapy with an MR-linac – Quantifying the impact of the magnetic field and real-time tumor tracking. *Radiother Oncol* 2016;119:461–6. <https://doi.org/10.1016/j.radonc.2016.04.019>.
- [78] Meijnters L, Hewson E, Nguyen DT, Booth J, Keall P. Dose-based optimisation for multi-leaf collimator tracking during radiation therapy. *Phys Med Biol* 2021;66:065027. <https://doi.org/10.1088/1361-6560/abe836>.
- [79] Hewson EA, Nguyen DT, Le A, Booth JT, Keall PJ, Meijnters L. Optimising multi-target multileaf collimator tracking using real-time dose for locally advanced prostate cancer patients. *Phys Med Biol* 2022;67:185003. <https://doi.org/10.1088/1361-6560/ac8967>.
- [80] Ruan D, Keall P. Dynamic multileaf collimator control for motion adaptive radiotherapy: An optimization approach. In: 2011 IEEE Power Engineering and Automation Conference. Vol 3. ; 2011:100–103. doi: 10.1109/PEAM.2011.6135024.
- [81] Schlemper J, Oksuz I, Clough JR, et al. dAUTOMAP: decomposing AUTOMAP to achieve scalability and enhance performance. Published online September 25, 2019. doi: 10.48550/arXiv.1909.10995.
- [82] Masutani EM, Bahrami N, Hsiao A. Deep learning single-frame and multiframe super-resolution for cardiac MRI. *Radiology* 2020;295:552–61. <https://doi.org/10.1148/radiol.2020192173>.
- [83] Kontaxis C, Bol GH, Legendijk JJW, Raaymakers BW. DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning. *Phys Med Biol* 2020;65:075013. <https://doi.org/10.1088/1361-6560/ab7630>.
- [84] Stark LS, Andrasschke N, Baumgartl M, et al. Dosimetric and geometric end-to-end accuracy of a magnetic resonance guided linear accelerator. *Phys Imaging Radiat Oncol* 2020;16:109–12. <https://doi.org/10.1016/j.phro.2020.09.013>.
- [85] 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. <https://doi.org/10.1111/1754-9485.12713>.
- [86] Willoughby TR, Forbes AR, Buchholz D, et al. Evaluation of an infrared camera and X-ray system using implanted fiducials in patients with lung tumors for gated radiation therapy. *Int J Radiat Oncol Biol Phys* 2006;66:568–75. <https://doi.org/10.1016/j.ijrobp.2006.05.029>.
- [87] Chen GP, Tai A, Keiper TD, Lim S, Li XA. Technical Note: Comprehensive performance tests of the first clinical real-time motion tracking and compensation system using MLC and jaws. *Med Phys* 2020;47:2814–25. <https://doi.org/10.1002/mp.14171>.
- [88] Booth J, Cailliet V, Briggs A, et al. MLC tracking for lung SABR is feasible, efficient and delivers high-precision target dose and lower normal tissue dose. *Radiother Oncol* 2021;155:131–7. <https://doi.org/10.1016/j.radonc.2020.10.036>.
- [89] Chun J, Lewis B, Ji Z, et al. Evaluation of super-resolution on 50 pancreatic cancer patients with real-time cine MRI from 0.35T MRgRT. *Biomed Phys Eng Express* 2021;7:055020. <https://doi.org/10.1088/2057-1976/ac1c51>.
- [90] Huttunga NRF, Bruijnen T, van den Berg CAT, Sbrizzi A. Gaussian Processes for real-time 3D motion and uncertainty estimation during MR-guided radiotherapy. *Med Image Anal* 2023;88:102843. <https://doi.org/10.1016/j.media.2023.102843>.
- [91] Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach Learn* 2021;110:457–506. <https://doi.org/10.1007/s10994-021-05946-3>.
- [92] Waddington DEJ, Hindley N, Koonjoo N, et al. Real-time radial reconstruction with domain transform manifold learning for MRI-guided radiotherapy. *Phys Med Biol* 2023;50:1962–74. <https://doi.org/10.1002/mp.16224>.
- [93] 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:e44–64. <https://doi.org/10.1002/mp.14625>.
- [94] Wojcieszynski AP, Rosenberg SA, Brower JV, et al. Gadaxetate for direct tumor therapy and tracking with real-time MRI-guided stereotactic body radiation therapy of the liver. *Radiother Oncol* 2016;118:416–8. <https://doi.org/10.1016/j.radonc.2015.10.024>.

A.1. Topical review: Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects

115

E. Lombardo et al.

Radiotherapy and Oncology 190 (2024) 109970

- [95] Kim T, Lewis BC, Price A, et al. Direct tumor visual feedback during free breathing in 0.35T MRgRT. *J Appl Clin Med Phys* 2020;21:241–7. <https://doi.org/10.1002/acm2.13016>.
- [96] Charters JA, Abdulkadir Y, O'Connell D, Yang Y, Lamb JM. Dosimetric evaluation of respiratory gating on a 0.35-T magnetic resonance-guided radiotherapy linac. *J Appl Clin Med Phys* 2022;23:e13666.
- [97] Ma TM, Neylon J, Casado M, et al. Dosimetric impact of interfraction prostate and seminal vesicle volume changes and rotation: A post-hoc analysis of a phase III randomized trial of MRI-guided versus CT-guided stereotactic body radiotherapy. *Radiother Oncol* 2022;167:203–10. <https://doi.org/10.1016/j.radonc.2021.12.037>.
- [98] Grimbergen G, Hackett SL, van Ommen F, et al. Gating and interfraction drift correction on a 1.5 T MR-Linac: clinical dosimetric benefits for upper abdominal tumors. *Radiother Oncol* 2023. <https://doi.org/10.1016/j.radonc.2023.109932>.
- [99] Wahlstedt I, Andratschke N, Behrens CP, et al. Gating has a negligible impact on dose delivered in MRI-guided online adaptive radiotherapy of prostate cancer. *Radiother Oncol* 2022;170:205–12. <https://doi.org/10.1016/j.radonc.2022.03.013>.
- [100] 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*. doi:10.1002/mp.16770.
- [101] Xiao F, Cai J, Zhou X, Zhou L, Song T, Li Y. TransDose: a transformer-based UNet model for fast and accurate dose calculation for MR-LINACs. *Phys Med Biol* 2022; 67. <https://doi.org/10.1088/1361-6560/ac7376>.
- [102] Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW. Towards adaptive IMRT sequencing for the MR-linac. *Phys Med Biol* 2015;60:2493–509. <https://doi.org/10.1088/0031-9155/60/6/2493>.
- [103] Rabe M, Paganelli C, Schmitz HP, et al. Continuous time-resolved estimated synthetic 4D-CTs for dose reconstruction of lung tumor treatments at a 0.35 T MR-linac. *Phys Med Biol* Published online 2023. <https://doi.org/10.1088/1361-6560/acff60>.
- [104] Seppenwoolde Y, Berbeco RI, Nishioka S, Shirato H, Heijmen B. Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: a simulation study. *Med Phys* 2007;34:2774–84. <https://doi.org/10.1118/1.2739811>.
- [105] Hiraoka M, Mizowaki T, Matsuo Y, Nakamura M, Verellen D. The gimbaled-head radiotherapy system: Rise and downfall of a dedicated system for dynamic tumor tracking with real-time monitoring and dynamic WaveArc. *Radiother Oncol* 2020; 153:311–8. <https://doi.org/10.1016/j.radonc.2020.07.002>.
- [106] Ng JA, Booth JT, O'Brien RT, et al. Quality assurance for the clinical implementation of kilovoltage intrafraction monitoring for prostate cancer VMAT. *Med Phys* 2014;41:111712. <https://doi.org/10.1118/1.4898119>.
- [107] Thorwarth D, Low DA. Technical challenges of real-time adaptive MR-guided radiotherapy. *Front Oncol* 2021;11. <https://doi.org/10.3389/fonc.2021.634507>.
- [108] Han R, Jones CK, Lee J, et al. Deformable MR-CT image registration using an unsupervised, dual-channel network for neurosurgical guidance. *Med Image Anal* 2022;75:102292. <https://doi.org/10.1016/j.media.2021.102292>.
- [109] Keall PJ, Glide-Hurst CK, Cao M, et al. ICRU REPORT 97: MRI-guided radiation therapy using MRI-linear accelerators. *J ICRU* 2022;22:1–100. <https://doi.org/10.1177/14736691221141950>.
- [110] Zhang Y, Trnkova P, Toshito T, et al. A survey of practice patterns for real-time intrafractional motion-management in particle therapy. *Physics and Imaging. Radiat Oncol* 2023;26. <https://doi.org/10.1016/j.phro.2023.100439>.
- [111] Hoogeman M, Prévost JB, Nuytens J, Pöll J, Levendag P, Heijmen B. Clinical accuracy of the respiratory tumor tracking system of the cyberknife: assessment by analysis of log files. *Int J Radiat Oncol Biol Phys* 2009;74:297–303. <https://doi.org/10.1016/j.ijrobp.2008.12.041>.

Bibliography

1. *Global cancer burden growing, amidst mounting need for services* <https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services#:~:text=In%202022%2C%20there%20were%20an,cases%20and%209.7%20million%20deaths> (2024).
2. Borrás, J. M. *et al.* The optimal utilization proportion of external beam radiotherapy in European countries: an ESTRO-HERO analysis. *Radiotherapy and Oncology* **116**, 38–44. 10.1016/j.radonc.2015.04.018 (2015).
3. Baumann, M. *et al.* Radiation oncology in the era of precision medicine. *Nature Reviews Cancer* **16**, 234–249. 10.1038/nrc.2016.18 (2016).
4. Langen, K. M. & Jones, D. T. Organ motion and its management. *International Journal of Radiation Oncology* Biology* Physics* **50**, 265–278. [https://doi.org/10.1016/S0360-3016\(01\)01453-5](https://doi.org/10.1016/S0360-3016(01)01453-5) (2001).
5. Corradini, S. *et al.* MR-guidance in clinical reality: current treatment challenges and future perspectives. *Radiation Oncology* **14**, 1–12. 10.1186/s13014-019-1308-y (2019).
6. Kurz, C. *et al.* Medical physics challenges in clinical MR-guided radiotherapy. *Radiotherapy and Oncology* **15**, 93. 10.1186/s13014-020-01524-4 (2020).
7. Beyzadeoglu, M., Ozyigit, G. & Ebruli, C. *Basic radiation oncology* (Springer, 2010).
8. Hall, E. J., Giaccia, A. J., *et al.* *Radiobiology for the Radiologist* (Philadelphia, 2006).
9. Jaffray, D. A. Image-guided radiotherapy: from current concept to future perspectives. *Nature reviews Clinical oncology* **9**, 688–699. <https://doi.org/10.1038/nrclinonc.2012.194> (2012).
10. Ladbury, C. *et al.* Clinical applications of magnetic resonance-guided radiotherapy: A narrative review. *Cancers* **15**, 2916. 10.3390/cancers15112916 (2023).
11. Mutic, S. & Dempsey, J. F. *The ViewRay system: magnetic resonance-guided and controlled radiotherapy* in *Seminars in radiation oncology* **24** (2014), 196–199. 10.1016/j.semradonc.2014.02.008.
12. Keall, P. J. *et al.* ICRU REPORT 97: MRI-guided radiation therapy using MRI-linear accelerators. *Journal of the ICRU* **22**, 1–100. <https://doi.org/10.1177/147366912211419> (2022).
13. Klueter, S. Technical design and concept of a 0.35 T MR-linac. *Clinical and translational radiation oncology* **18**, 98–101. 10.1016/j.ctro.2019.04.007 (2019).

14. Keall, P. J., Barton, M., Crozier, S., *et al.* The Australian magnetic resonance imaging–linac program in *Seminars in radiation oncology* **24** (2014), 203–206. 10.1016/j.semradonc.2014.02.015.
15. Kawula, M. *et al.* Patient-specific transfer learning for auto-segmentation in adaptive 0.35 T MRgRT of prostate cancer: a bi-centric evaluation. *Medical Physics* **50**, 1573–1585. <https://doi.org/10.1002/mp.16056> (2023).
16. Li, Z. *et al.* Patient-specific daily updated deep learning auto-segmentation for MRI-guided adaptive radiotherapy. *Radiotherapy and Oncology* **177**, 222–230. <https://doi.org/10.1016/j.radonc.2022.11.004> (2022).
17. Sahin, B. *et al.* First 500 fractions delivered with a magnetic resonance-guided radiotherapy system: initial experience. *Cureus* **11**. 10.7759/cureus.6457 (2019).
18. Green, O. L. *et al.* First clinical implementation of real-time, real anatomy tracking and radiation beam control. *Medical physics* **45**, 3728–3740. <https://doi.org/10.1002/mp.13002> (2018).
19. Grimbergen, G. *et al.* Gating and intrafraction drift correction on a 1.5 T MR-Linac: Clinical dosimetric benefits for upper abdominal tumors. *Radiotherapy and Oncology* **189**, 109932. <https://doi.org/10.1016/j.radonc.2023.109932> (2023).
20. Lombardo, E. *et al.* Real-time motion management in MRI-guided radiotherapy: Current status and AI-enabled prospects. *Radiotherapy and Oncology*, 109970. 10.1016/j.radonc.2023.109970 (2023).
21. Paganelli, C. *et al.* MRI-guidance for motion management in external beam radiotherapy: current status and future challenges. *Physics in Medicine & Biology* **63**, 22TR03. 10.1088/1361-6560/aaebcf (2018).
22. Grimbergen, G. *et al.* Feasibility of delivered dose reconstruction for MR-guided SBRT of pancreatic tumors with fast, real-time 3D cine MRI. *Radiotherapy and Oncology* **182**, 109506. <https://doi.org/10.1016/j.radonc.2023.109506> (2023).
23. Waddington, D. E. *et al.* Real-time radial reconstruction with domain transform manifold learning for MRI-guided radiotherapy. *Medical physics* **50**, 1962–1974. <https://doi.org/10.1002/mp.16224> (2023).
24. Chun, J. *et al.* Evaluation of super-resolution on 50 pancreatic cancer patients with real-time cine MRI from 0.35 T MRgRT. *Biomedical Physics & Engineering Express* **7**, 055020. 10.1088/2057-1976/ac1c51 (2021).
25. Feng, Y. *et al.* A comparative study of automatic image segmentation algorithms for target tracking in MR-IGRT. *Journal of applied clinical medical physics* **17**, 441–460. <https://doi.org/10.1120/jacmp.v17i2.5820> (2016).
26. Palacios, M. A. *et al.* Accuracy of deformable image registration-based intra-fraction motion management in Magnetic Resonance-guided radiotherapy. *Physics and Imaging in Radiation Oncology* **26**, 100437. <https://doi.org/10.1016/j.phro.2023.100437> (2023).
27. Jassar, H. *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. *Medical Physics* **50**, 3103–3116. <https://doi.org/10.1002/mp.16342> (2023).

28. Friedrich, F. *et al.* Stability of conventional and machine learning-based tumor auto-segmentation techniques using undersampled dynamic radial bSSFP acquisitions on a 0.35 T hybrid MR-linac system. *Medical Physics* **48**, 587–596. <https://doi.org/10.1002/mp.14659> (2021).
29. Hunt, B. *et al.* Fast Deformable image registration for real-time target tracking during radiation therapy using cine MRI and deep learning. *International Journal of Radiation Oncology* Biology* Physics* **115**, 983–993. <https://doi.org/10.1016/j.ijrobp.2022.09.086> (2023).
30. Kim, T., Lewis, B., Lotey, R., Barberi, E. & Green, O. Clinical experience of MRI4D QUASAR motion phantom for latency measurements in 0.35 T MR-LINAC. *Journal of applied clinical medical physics* **22**, 128–136. 10.1002/acm2.13118 (2021).
31. Glitzner, M., Woodhead, P., Borman, P., Lagendijk, J. & Raaymakers, B. MLC-tracking performance on the Elekta unity MRI-linac. *Physics in Medicine & Biology* **64**, 15NT02. 10.1088/1361-6560/ab2667 (2019).
32. Liu, P. Z. *et al.* First experimental investigation of simultaneously tracking two independently moving targets on an MRI-linac using real-time MRI and MLC tracking. *Medical Physics* **47**, 6440–6449. <https://doi.org/10.1002/mp.14536> (2020).
33. 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* **6**, 51262–51268. 10.1109/ACCESS.2018.2869780 (2018).
34. Krauss, A., Nill, S. & Oelfke, U. The comparative performance of four respiratory motion predictors for real-time tumour tracking. *Physics in Medicine & Biology* **56**, 5303. 10.1088/0031-9155/56/16/015 (2011).
35. Uijtewaal, P. *et al.* Dosimetric evaluation of MRI-guided multi-leaf collimator tracking and trailing for lung stereotactic body radiation therapy. *Medical Physics* **48**, 1520–1532. 10.1002/mp.14772 (2021).
36. Eze, C. *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. *Radiation Oncology* **17**, 43. <https://doi.org/10.1186/s13014-022-02011-8> (2022).
37. Ehrbar, S. *et al.* MR-guided beam gating: Residual motion, gating efficiency and dose reconstruction for stereotactic treatments of the liver and lung. *Radiotherapy and Oncology* **174**, 101–108. <https://doi.org/10.1016/j.radonc.2022.07.007> (2022).
38. Uijtewaal, P. *et al.* First experimental demonstration of VMAT combined with MLC tracking for single and multi fraction lung SBRT on an MR-linac. *Radiotherapy and Oncology* **174**, 149–157. 10.1016/j.radonc.2022.07.004 (2022).
39. Ruan, D. & Keall, P. *Dynamic multileaf collimator control for motion adaptive radiotherapy: An optimization approach* in *2011 IEEE Power Engineering and Automation Conference* **3** (2011), 100–103. 10.1109/PEAM.2011.6135024.
40. Keall, P. J. *et al.* The first clinical implementation of electromagnetic transponder-guided MLC tracking. *Medical physics* **41**, 020702. 10.1118/1.4862509 (2014).

41. Booth, J. T. *et al.* The first patient treatment of electromagnetic-guided real time adaptive radiotherapy using MLC tracking for lung SABR. *Radiotherapy and Oncology* **121**, 19–25. 10.1016/j.radonc.2016.08.025 (2016).
42. Shen, C. *et al.* An introduction to deep learning in medical physics: advantages, potential, and challenges. *Physics in Medicine & Biology* **65**, 05TR01. 10.1088/1361-6560/ab6f51 (2020).
43. Alpaydin, E. *Introduction to machine learning* <https://mitpress.mit.edu/9780262012119/introduction-to-machine-learning/> (MIT press, 2020).
44. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444. <https://doi.org/10.1038/nature14539> (2015).
45. Pandya, S. *et al.* A study of the recent trends of immunology: key challenges, domains, applications, datasets, and future directions. *Sensors* **21**, 7786. <https://doi.org/10.3390/s21237786> (2021).
46. Cusumano, D. *et al.* Artificial Intelligence in magnetic Resonance guided Radiotherapy: Medical and physical considerations on state of art and future perspectives. *Physica medica* **85**, 175–191. 10.1016/j.ejmp.2021.05.010 (2021).
47. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Computation* **9** (Nov. 1997).
48. Lombardo, E. *et al.* Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. *Physics in Medicine & Biology* **67**, 095006. <https://doi.org/10.1088/1361-6560/ac60b7> (2022).
49. Shi, X. *et al.* Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* **28**. <https://doi.org/10.48550/arXiv.1506.04214> (2015).
50. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*. <https://arxiv.org/abs/1409.0473> (2014).
51. Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30**. <https://arxiv.org/abs/1706.03762> (2017).
52. Feng, S., Garg, D., Bunnaprast, E. & Lee, S. CS 25: *Overview of Transformers* <https://web.stanford.edu/class/cs25/> (Stanford University, 2024).
53. Islam, S. *et al.* A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications*, 122666. <https://arxiv.org/abs/2306.07303> (2023).
54. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. <https://arxiv.org/abs/2010.11929> (2020).
55. Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows in *Proceedings of the IEEE/CVF international conference on computer vision* (2021), 10012–10022. <https://arxiv.org/abs/2103.14030>.
56. Chen, J. *et al.* Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis* **82**, 102615. <https://doi.org/10.1016/j.media.2022.102615> (2022).

57. Jöhl, A. *et al.* Performance comparison of prediction filters for respiratory motion tracking in radiotherapy. *Medical physics* **47**, 643–650. <https://doi.org/10.1002/mp.13929> (2020).
58. Lin, H. *et al.* Towards real-time respiratory motion prediction based on long short-term memory neural networks. *Physics in Medicine & Biology* **64**, 085010. [10.1088/1361-6560/ab13fa](https://doi.org/10.1088/1361-6560/ab13fa) (2019).
59. Romaguera, L. V. *et al.* Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks. *Medical image analysis* **64**, 101754. <https://doi.org/10.1016/j.media.2020.101754> (2020).
60. Zou, K. H. *et al.* Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology* **11**, 178–189. [https://doi.org/10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8) (2004).
61. Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence* **15**, 850–863. [10.1109/34.232073](https://doi.org/10.1109/34.232073) (1993).
62. Beck, M. *et al.* xLSTM: Extended Long Short-Term Memory. *arXiv preprint arXiv:2405.04517*. <https://arxiv.org/abs/2405.04517> (2024).
63. Paganelli, C. *et al.* Feasibility study on 3D image reconstruction from 2D orthogonal cine-MRI for MRI-guided radiotherapy. *Journal of medical imaging and radiation oncology* **62**, 389–400. [10.1111/1754-9485.12713](https://doi.org/10.1111/1754-9485.12713) (2018).
64. Rabe, M. *et al.* Porcine lung phantom-based validation of estimated 4D-MRI using orthogonal cine imaging for low-field MR-Linacs. *Physics in Medicine & Biology* **66**, 055006. [10.1088/1361-6560/abc937](https://doi.org/10.1088/1361-6560/abc937) (2021).
65. Wei, R. *et al.* Real-time 3D MRI reconstruction from cine-MRI using unsupervised network in MRI-guided radiotherapy for liver cancer. *Medical Physics* **50**, 3584–3596. [10.1002/mp.16141](https://doi.org/10.1002/mp.16141) (2023).
66. Romaguera, L. V., Mezheritsky, T., Mansour, R., Carrier, J.-F. & Kadoury, S. Probabilistic 4D predictive model from in-room surrogates using conditional generative networks for image-guided radiotherapy. *Medical image analysis* **74**, 102250. <https://doi.org/10.1016/j.media.2021.102250> (2021).

Acknowledgments

When I started my Bachelor in Physics I already knew I wanted to do a PhD in medical physics. While there are many people I could thank for all the support and inspiration throughout my studies, I will here only list those, having a direct connection to this thesis.

First of all, I consider myself extremely lucky to have had not one, not two, but three amazing supervisors which guided me throughout these years. I would like to start by thanking Prof. Marco Riboldi who taught me all I know about statistical tests and for being an infinite source of tranquillity and mindfulness even during the most difficult meetings. I do not think I have ever seen him angry or impatient and I will always try to follow his example in my career. Next, I would like to thank PD Dr. Christopher Kurz for the cornucopia of ideas he always threw at me during our regular meetings, for his acute observational spirit when judging results and for nice conversations about pizza in Naples. Last, but not least among my supervisors, I owe extreme gratitude to Prof. Guillaume Landry. Besides being an inspiration professionally for many things including but not limited to caring about the well-being of students, running meetings, writing papers (especially figures), managing collaborations and following good scientific practice, I by now consider him as a friend with whom I can talk about where to buy good fish in Munich, the latest article in *The Economist* and which metal album to listen to during workout. Thank you all again and stay as you are!

I am happy I did my PhD in Munich as it is not only a great city but offers a great environment for academic research in the field of medical physics. My work has been funded by the German Research Foundation (DFG) within the Research Training Group GRK 2274 (*Advanced Medical Physics for Image-Guided Cancer Therapy*) and by the Bavarian Cancer Research Center (BZKF). I am grateful for having been a member of the GRK which gave me the the chance to participate in interesting seminars, summer schools, retreats, and courses and to meet and get to network with the other doctoral candidates within the training group. I would therefore like to thank the coordination and management team of the GRK and in particular PD Dr. Dr. Christian Thieke, Prof. Katia Parodi, Prof. Claus Belka and Prof. Franz Pfeiffer for creating this environment.

Thanks to DFG funding, I also got the chance to carry out measurements and stay for two months in Sydney, another great city both scenically/culturally and in terms of medical physics research. I am very grateful to Prof. Paul Keall for being truly inspirational and his group for the extremely warm welcome, which made me feel like I was part of Image X since day one. Special thanks go to Dr. Paul Liu, Dr. David Waddington and James Grover for their amazing support with the experiments, which, as expected, worked out only on the last available day!

Special thanks also go to Dr. Davide Cusumano, Dr. Luca Boldrini and Dr. Lorenzo Placidi at Mater Olbia and Gemelli in Rome for the enriching and fun collaborations and for giving me visibility within the ESTRO community. Grazie!

In parallel to my PhD, I was lucky to be given the possibility to carry out the training to become a qualified medical physics expert (MPE). This training gave me a different, more clinical perspective on the things I did in research and I am very grateful to Dr. Michael Reiner for giving me this possibility and for his supervision. Special thanks for the supervision during the treatment planning training go to Dr. Sylvia Garny for being the best multi-tasker, to Dr. Christoph Losert for guiding me into Total-Body-Irradiation and to Roel Shpani for being my favorite Albana. Special thanks for the supervision during my quality assurance training go to Robert Kießling for being there during challenging Frühchecks, Tobias Winderl for all the fun we had setting up QATrack+ and Winfried Hoischen for the equally valuable insights in Linac-QA and recommendations in northern Italy. I would like to thank Amra Mekic-Krejovic, Dr. Katrin Straub, Dr. Vasiliki Anagnostatou, Zakaria Bahammou, Dr. Philipp Freisleder, Dr. Helmut Weingandt, Dr. Jan Hofmaier, Dr. Vanessa Filipa da Silva Mendes, Dr. Mohammad Safari, Lukas Nierer and Catrin Rodenberg for the interesting and funny chats in the coffee room about almost everything. On the same note I would like to thank Dr. Christian Winkelhofer, Dr. Anita Ladenburger and Dr. Christian Heinz for the nice conversations at Wirtshaus Franz. I also would like to thank Prof. Claus Belka, all physicians and other co-workers at the department for the stimulating environment and especially Prof. Stefanie Corradini for being the AI ground-truth and Dr. Sebastian Marschner for the help provided in countless projects.

Of course, I also had fantastic colleagues in the research group with which I share many nice memories of conferences, beer gardens and doctor hat fabrication. Here I would like to thank Jackie Xiong (don't copy this), Maria Kawula, Lili Huang, Ivy Chan, Nikos Ntelopoulos, Dr. Adrian Thummerer, Dr. Yiling Wang, Chengtao Wei, Lei Chan, Fan Xiao, Marvin Ribeiro, Domagoj Radonic, Lukas Schimdt, Rabea Klaar, Tom Blöcker and the students I co-supervised Claudia Tejero, Laura Velezmoro and Nicolas Mühlshlegel. Special thanks go to Dr. Moritz Rabe for equal contributions to my image registration and beer knowledge and to Sven Marquart and Dr. Henning Schmitz for the fun conversations in the U3 office.

Last but not least, I wish to dedicate this thesis to my love and friend Alessia, who gave everything when studying for her Staatsexamen and inspired me to do the same. We really like the following quote which is valid for life as well as for science:

*"You have to try the impossible to reach the possible."
 "Man muss das Unmögliche versuchen, um das Mögliche zu erreichen."*

Hermann Hesse