Intelligent Oncology 1 (2025) 244–255

Contents lists available at ScienceDirect

Intelligent Oncology

journal homepage: www.keaipublishing.com/en/intonc/

Review

Deep learning applications in motion management for radiotherapy

Bining Yang, Ran Wei, Jianrong Dai*

National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

ARTICLE INFO

Keywords:
Motion management
Deep learning
Radiotherapy

ABSTRACT

The aim of radiotherapy (RT) is to deliver prescribed doses to tumors while sparing neighboring organs at risk. As the demand for treatment precision increases in modern RT, intrafractional motion management becomes critical for achieving high precision, particularly for tumors in the thorax and abdomen that are affected by respiration and other physiological motions. Deep learning (DL) has demonstrated strong potential in addressing the limitations of conventional methods by enabling rapid and accurate tumor motion detection, tumor location prediction, and management measures. This review provides a comprehensive overview of recent DL-based applications in intrafractional motion management, which are categorized into three areas: (1) tumor motion detection for real-time tumor localization, (2) tumor location prediction to compensate for the latency of the motion management system, and (3) management measures, including gating and real-time adaptive RT. In addition, this review discusses key challenges and their potential solutions for DL-based motion management.

1. Introduction

Radiotherapy (RT) has proven efficient for various cancers.[1–4] As the demand for treatment precision increases in modern RT, addressing the challenges posed by intrafractional motion becomes essential. Intrafractional motion can generally be divided into two types: faster, periodic motion, such as respiration and heartbeat, and slower, non-periodic motion, such as bladder filling and peristalsis. Both can affect dose delivery accuracy and thus require continuous monitoring and appropriate measures to minimize the impact.

Effectively managing motion is difficult, as it requires both high spatial precision and rapid management. It involves real-time tumor detection and the application of different measures depending on the extent of motion: (1) margin expansion[5] (using the expanded planning target volume as the target of treatment to avoid underdosing caused by tumor motion), (2) breath-hold method[6] (instructing the patient to hold their breath at a predefined phase of the respiratory cycle), (3) abdominal compression[7] (forcing the patient to breathe shallowly by applying external pressure to the abdomen), (4) gating[8,9] (activating the beam only when the tumor or its surrogate remains within a predefined threshold), (5) beam tracking[10] (dynamically adjusting the beam to follow tumor motion), and (6) real-time adaptive radiotherapy (ART)[11] (reoptimizing the treatment plan based on the current anatomy).

However, there is an inevitable time delay between tumor motion detection and the corresponding measures. During this latency period, the tumor location changes continuously. The American Association of Physicists in Medicine Task Group 76 report[12] recommends that the total system delay for real-time detection and compensation should be kept below 0.5 s. It is necessary to predict the future location of the tumor, which enables the system to anticipate motion and initiate timely corrective actions.

To deal with this issue, various intrafractional motion management methods have been developed; however, direct detection and management are impractical due to technical and anatomical limitations. The rapid advancement of deep learning (DL) has enabled its widespread application in motion detection, tumor location prediction, and management measures across multiple disease sites and imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound. Leveraging these capabilities, a variety of methods have been proposed that offer promising solutions to the challenges of fast and complex anatomical changes during treatment delivery.

As shown in Figure 1, applications in motion management have increased in recent years, with a major contribution from DL methods. This article provides a comprehensive overview of DL-based methods for intrafractional motion management, highlighting representative models and discussing existing challenges and possible solutions. Figure 2 illustrates the

* Corresponding author.
E-mail address: dai_jianrong@cicams.ac.cn (J. Dai).

https://doi.org/10.1016/j.intonc.2025.06.006
Received 17 April 2025; Received in revised form 17 June 2025; Accepted 20 June 2025
Available online 1 July 2025
2950-2616/© 2025 The Author(s). Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

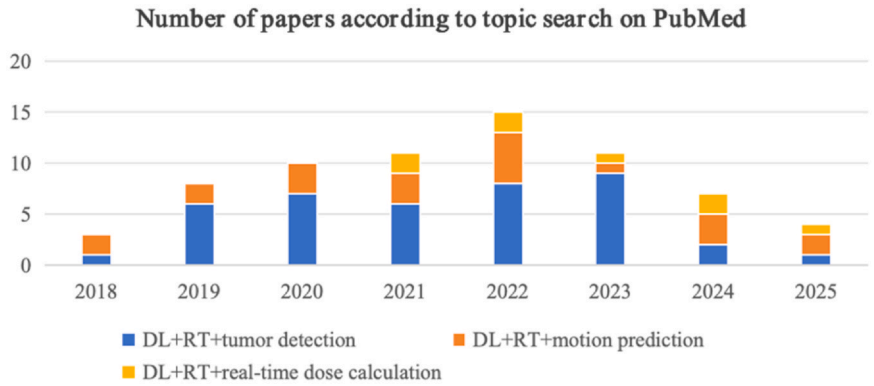


Figure 1. Publication volume from PubMed topic search. Keywords: "deep learning," "radiotherapy," "tumor detection," "motion prediction," and "real-time dose calculation".
Abbreviations: 1. DL: deep learning 2. RT: radiotherapy.

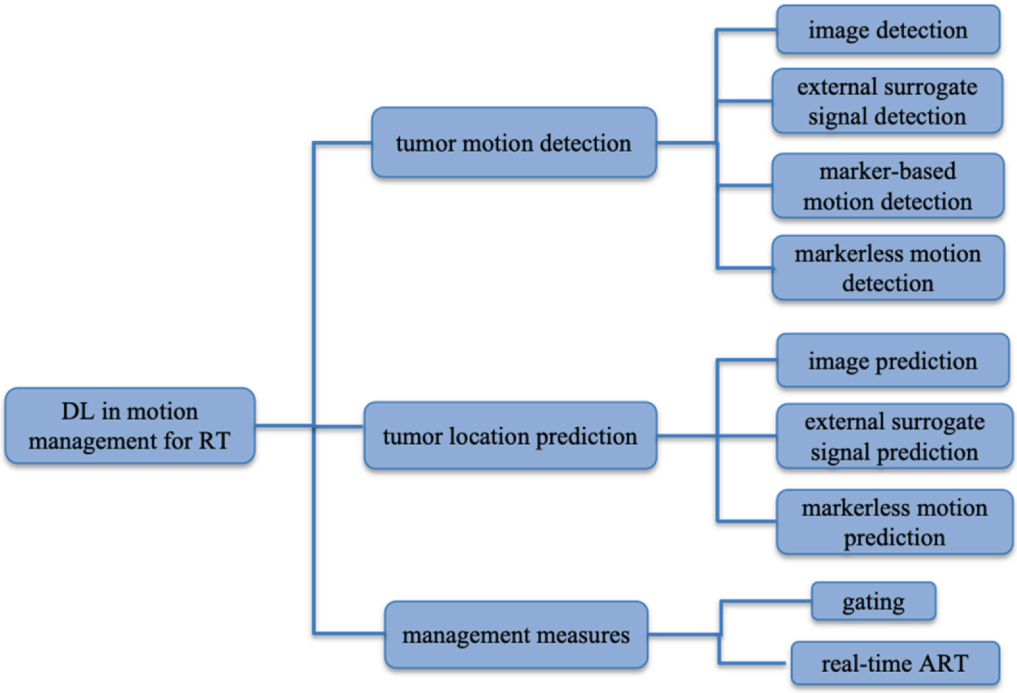


Figure 2. Classification of DL applications in motion management for RT.
Abbreviations: 1. DL: deep learning 2. RT: radiotherapy 3. ART: adaptive radiotherapy.

classification of these DL applications, which are categorized into three major domains: tumor motion detection, tumor location prediction, and corresponding management measures, each encompassing various techniques.

2. DL in tumor motion detection

The first step in intrafractional motion management is accurate and timely real-time detection of the tumor or its surrogate. To provide a clearer comparison of relevant studies, we categorized existing DL-based tumor detection methods into four subgroups: image detection, external surrogate signal detection, marker-based motion detection, and markerless motion detection.

2.1. Image detection

Continuous image acquisition during treatment is the most direct method for monitoring tumor motion. DL-based image detection has shown great performance in this task. A key method involves inferring 3D volumetric information from real-time 2D images through fast image reconstruction. For MRI reconstruction, Liu et al.[13] developed a geometry-informed framework capable of estimating volumetric data with subsecond acquisition times. Their method further integrates implicit neural representation learning combined with prior anatomical information, enabling fast and accurate 3D MRI reconstruction from orthogonal cine-magnetic resonance images.[14] For CT reconstruction, Lei et al.[15] employed generative adversarial learning to enable a more realistic 2D-to-3D transformation. The proposed TransNet network comprises three parts: an encoding module, a transformation module, and a decoding module. To facilitate the performance of the model, they devised the conventional loss function of the generative adversarial network by introducing perceptual loss.

2.2. External surrogate signal detection

In addition to direct motion detection methods that rely on imaging, indirect methods based on external surrogates are widely used. Thoracic and abdominal tumors are significantly affected by respiratory

motion, and surface movements captured by optical or infrared systems can be employed as surrogates to estimate the respiratory signal, which is then correlated with internal tumor motion.

Multiple research groups have utilized surface motion as a surrogate to estimate internal tumor location. Huang et al.[16] developed a patient-specific model that predicts principal component analysis-compressed deformation vector fields (DVFs) from body surface displacement to generate real-time 4D images. Zhang et al.[17] introduced 3D-U-RAD, a U-Net-based architecture that integrates surface deformation, anatomical masks, and deformation priors to directly predict a 3D tumor mask. In addition, Shao et al.[18] proposed Surf-X-Bio, a hybrid architecture that combines surface imaging and single X-ray projection, in which deep models estimate boundary motion and use biomechanical modeling to correlate internal tumor locations.

2.3. Marker-based motion detection

Fiducial markers implanted near the tumor serve as internal surrogates for certain tumor types. Unlike external surrogates, which require internal/external correlation models, implanted markers are typically assumed to move with the tumor, allowing direct estimation of tumor motion from their trajectories.

Fiducial markers used in conjunction with kilovoltage projections are the most often employed.[19–22] For example, Liang et al.[20] trained a fully convolutional network–based detection model to localize the fiducial markers in orthogonal X-ray images. The 3D marker locations were reconstructed using the model's outputs and then used to evaluate intrafractional motion via rigid transformation. The marker-based motion detection methods are summarized in Table 1.

Megavoltage (MV) imaging is ideal for marker motion detection, as it requires no additional equipment and captures images along the treatment beam view. However, its use is limited due to low image contrast and beam modulation, whereby multileaf collimator (MLC) leaves may temporarily obscure the markers. Chrystall et al.[23] developed an MV-based detection system using a convolutional neural network classifier for fast segmentation of implanted markers in prostate stereotactic body radiation therapy, which demonstrated high accuracy in identifying markers.

2.4. Markerless motion detection

Fiducial markers are not suitable for all patients due to risks such as hemorrhage and potential migration. In contrast, markerless methods avoid surgical implantation, reducing patient discomfort and additional costs. In markerless motion detection, DL-based methods can be categorized into three types according to the output they generate, ranging from basic localization to the comprehensive estimation of DVFs.

The first category includes methods that localize tumors using bounding boxes or point coordinates on dynamic images, which mainly capture rigid motion.[24–34] While they offer real-time information, they do not account for tumor deformation or interaction with the surrounding organs at risk, limiting their ability to reflect intrafractional anatomical changes. Zhao et al.[25] combined a region proposal network with a target detection network, with the latter taking the output of the region proposal network as the input. To train this network, the planning CT images were deformed to mimic different anatomy scenarios, and the deformed images were used to generate digitally reconstructed radiographs that incorporated the tumor location. The authors also investigated the model's performance in pancreatic cancer.[26]

The second is a real-time contouring method in which DL models generate gross tumor volume (GTV) or clinical target volume masks for various tumor sites.[35–38] Compared to simple localization, contouring provides more information by outlining the target and indicating possible deformation. Zhou et al. investigated the feasibility of real-time contouring in thoracic and abdominal tumors. In lung cancer,[36] they used a ResNet and a feature pyramid network to predict the GTV

Table 1
Summary of marker-based motion detection publications.

Reference	Image modality	Dataset	Architecture	Target	Results
Chrystall et al.[23]	MV X-ray	Twenty-nine prostate cancer patients (9 patients for training, 20 patients for testing)	CNN	Implant markers in the prostate	The model achieved an AUC of 0.99, a sensitivity of 98.31%, and a specificity of 99.87%. The mean absolute geometry errors were 0.30 ± 0.27 mm (lateral) and 0.35 ± 0.31 mm (SI).
Mylonas et al.[19]	Cylindrical markers: 2D fluoroscopic images Arbitrary markers: kV projections	Cylindrical markers: kV images from a phantom and three prostate cancer patients for training and 10 patients for testing Arbitrary markers: Projections from three lung cancer patients for training and four patients for testing	Compact CNN and AlexNet	Cylindrical and arbitrarily shaped fiducial markers	Compact CNN: Sensitivity = 99%, specificity = 98.9%. AlexNet: Sensitivity = 98.6%, specificity = 99%. Cylindrical marker RMSE: 1.6 ± 0.2 pixels (x), 1.3 ± 0.4 pixels (y). Arbitrary marker RMSE: 3.0 ± 0.5 pixels (x), 2.2 ± 0.4 pixels (y).
Liang et al. [20]	Orthogonal kV X-ray	A total of 5927 real CyberKnife X-ray images and 13 liver cancer patients	fCN	Fiducial markers in the liver tumor	The model achieved a 98.6% precision rate and a 95.6% recall rate. The mean centroid error between the predicted markers and GT was 0.25 ± 0.47 pixels. Overall, 71% of fractions had rotations larger than the system limitations.
Amarsee et al.[21]	CBCT kV projections	A wax phantom and projection images from three prostate cancer patients	YOLO v2 CNN	Fiducial marker seeds in the prostate	The model had 98% detection accuracy, ± 1 mm detection variation of the seed center, and successful detection at all gantry angles.
Motley et al.[22]	CBCT kV projections	A total of 20,000 pelvis kV projection images with fiducial markers and 14 prostate patients	CNN	Fiducial markers in the pelvis	The model achieved 97.8% detection accuracy. The average deviation for interfraction marker migration was 2.0 ± 0.9 mm.

Abbreviations: 1. CBCT: cone-beam computed tomography, 2. CNN: convolutional neural network, 3. fCN: fully convolutional network, 4. RMSE: root mean square error, 5. SI: superior–inferior, 6. AUC: area under the curve, 7. GT: ground truth.

contour on X-ray images. The 3D location of the GTV was then calculated according to the centroids of the predicted contours in two orthogonal images. In pancreatic cancer,[35] a mask region-based CNN model was used to perform clinical target volume detection. ResNet and a feature pyramid network were used as the backbone to extract features, and the model was pretrained using the COCO dataset.

When direct tumor visualization is challenging, an alternative is to contour surrounding anatomical structures that maintain a stable spatial relationship with the tumor to infer its motion. Edmunds et al.[39] segmented the diaphragm for lung tumor detection. A mask region-based CNN was trained on 3,499 cone-beam CT (CBCT) images from 10 patients. The model detects real-time diaphragm motion with a mean error of 4.4 mm, without manual intervention.

The most comprehensive methods achieve tumor detection by predicting DVFs using various image modalities, including MRI,[40–46] X-ray imaging,[47–50], and ultrasound.[51] These methods capture motion at the pixel or voxel level, enabling detailed visualization of intrafractional anatomical changes. Some studies have further utilized the predicted DVFs to generate synthetic real-time images or contours, improving both visualization and clinical utility.

For instance, Shao et al.[49,52] improved the accuracy of automatic liver tumor detection. In their earlier study,[52] 3D liver boundary DVFs were optimized by intensity-driven 2D-3D deformable registration between prior 3D CT/CBCT images and sparsed projections distributed over a full scan angle. These DVFs were then used as boundary conditions in a biomechanical model to improve the solution to intraliver tumor motion. Building upon this framework, their subsequent study[49] replaced the computationally intensive registration step with a graph neural network, which directly estimated liver surface deformation from a single onboard X-ray projection, greatly improving inference speed and allowing near real-time performance (<100 ms). The markerless motion detection methods are summarized in Table 2.

3. DL in tumor location prediction

Due to the time delay between motion detection and management, the system inevitably lags behind in terms of the tumor's actual location. Accurate location prediction is therefore essential to compensate for system latency and enable timely management measures. DL-based tumor location prediction methods are divided into three subgroups: image prediction, external surrogate signal prediction, and markerless motion prediction.

3.1. Image prediction

The long short-term memory (LSTM)-based framework is a mainstream paradigm in spatiotemporal image series forecasting. Weng et al.[54] incorporated a convolutional layer within the LSTM layer (ConvLSTM) to predict future cine-magnetic resonance by image series acquired in past moments, cropping specific regions of interest as the model input instead of whole images. Jin et al.[55] proposed a pix2pixGAN with the generator replaced by ConvLSTM for a similar task. Ghasemi et al.[56] employed two network architectures—a stacked ConvLSTM with hyperparameter optimization and a newly designed parallel ConvLSTM model—to predict spatiotemporal 4DCT images.

3.2. External surrogate signal prediction

Most external surrogate signal prediction studies have simplified respiratory signals into 1D superior–inferior directional displacement, since it represents the dominant component of tumor motion induced by respiration.[57–65] Sun et al.[58] applied an adaptive boosting and multilayer perceptron neural network framework to enhance the prediction accuracy of respiratory signals. Lin et al.[62] were the first to introduce LSTM to address the respiratory prediction task in 500 ms. Yu et al.[59] proposed a bidirectional gated recurrent unit, another variant of a recurrent neural network, to mitigate the relatively long update times associated with LSTM. Furthermore, to overcome the weak long-term dependency modeling of recurrent neural network–based methods, Tan et al.[61] simplified the transformer encoder architecture and proposed a novel long-term information enhancement module to achieve real-time respiratory signal prediction. In contrast to these 1D prediction methods, Pohl et al.[66] uniquely predicted 3D respiratory motion, providing a more comprehensive representation of tumor trajectories.

In addition to the accurate prediction of intrafractional respiratory motion, establishing a reliable correlation between external respiratory signals and internal tumor motion is important. The superiority of DL architectures in modeling this relationship has been demonstrated in previous studies. Teo et al.[67] focused on online neural network optimization and proposed a model composed of three multilayer perceptron neural networks. Similarly, Wang et al.[68] employed an LSTM network to predict external respiratory motion signals and subsequently established correlation models between external and internal motion. The external surrogate signal prediction methods are summarized in Table 3.

3.3. Markerless location prediction

DL-based methods for markerless location prediction can be broadly categorized into two types based on their predicted outputs. The first type focuses on predicting the coordinates of the target, such as the tumor centroid or bounding box location. Huang et al.[70] proposed a CNN-CLSTM architecture that integrates attention mechanisms with temporal modeling to enable accurate predictions. The model takes 2D ultrasound image sequences as input and outputs the 2D coordinates of internal landmarks, along with a bounding box for each frame. Lombardo et al.[71] developed offline and online LSTM-based models to predict future tumor locations in the superior–inferior direction using prior trajectories extracted from 2D cine-MRI, effectively forecasting respiratory-induced tumor motion.

The second type predicts future DVFs. Romaguera et al.[72–74] conducted a series of studies across different imaging modalities, including MRI, ultrasound, and CT, to estimate 2D or 3D DVFs using different architectures. The markerless location prediction methods are summarized in Table 4.

4. DL in management measures

Depending on the extent of intrafractional motion, DL-based management can be broadly categorized into margin expansion, breath-hold methods, forced shallow breathing with abdominal compression, respiratory gating, beam tracking, and real-time ART. Among these, beam tracking and real-time ART manage motion by changing the radiation beam. The key difference lies in their complexity: Beam tracking adjusts MLCs dynamically by the shape and movement of the tumor, whereas real-time ART requires a more comprehensive plan reoptimization and dose recalculation process to account for motion-induced variations. Most DL-related studies have focused on gating and dose calculation in real-time ART, which is discussed in detail below.

4.1. Gating

Respiratory-gated RT reduces normal tissue complications for mobile tumors in the thorax and abdomen. After detecting and predicting the tumor's location using the methods stated above, optimizing the gating window is important.

While no DL-related studies have been researched, a few early works have applied conventional machine learning methods to inform gating selection. Both Lin et al.[76] and Li et al.[77] addressed markerless fluoroscopic gating using machine learning by reformulating the beam on/off decision as a binary classification problem. Lin et al. conducted a systematic comparison of five dimensionality reduction methods, such as

Table 2
Summary of markerless motion detection publications.

Reference	Image modality	Dataset	Architecture	Results
Shao et al.[40]	MRI	Cardiac dataset: Eight cases from an open-access multicoil k-space dataset Abdominal dataset: Nine liver cancer patients	KS-RegNet	KS-RefNet performed best on both the cardiac and abdominal datasets in DSC and COMEs for 5-, 9-, and 13-spoke k-space acquisitions.
Hindley et al.[47]	CBCT	Two lung cancer patients	Voxelmap network	The mean errors of 3D tumor motion continuous prediction were 0.1 ± 0.5, −0.6 ± 0.8, and 0.0 ± 0.2 mm in LR, SI, and AP directions, respectively. SSIM = 1.0 and PSNR = 65.8.
Terpstra et al.[41]	MRI	Twenty-seven lung cancer patients (17 for training, 5 for validation, and 5 for testing) and the public 4DCT dataset.	A multiresolution CNN	The model accurately estimated motion compared to the self-navigation signal by 50 spokes per dynamic. DVFs were estimated within 200 ms, including MRI acquisition. The target registration error was 1.87 ± 1.65 mm without retraining the model.
Dai et al.[48]	2D DRRs	Lung patient data from TCIA and CIRS as well as phantom data	Swin transformer	The model only required 51 ms in the inference phase for tumor localization. For patient evaluation. The RMSE of the tumor centroid was <1.5 mm.
Terpstra et al.[42]	Cine-MRI	One hundred thirty-five abdominal cancer patients	CNN	DVFs obtained by combinations of conventional- and DL-based methods were R ≤ 25, with a RMSE ≤1 mm and an SSIM > 0.8 after registration.
Shao et al.[52]	2D − 3D deformable registration	Thirty-four liver cancer patients (24 for training and 10 for testing)	U-Net	The tumor COME, mean HD, and DSC were 1.7 ± 0.4 mm, 4.5 ± 1.3 mm, and 0.78 ± 0.03, respectively.
Shao et al.[49]	CBCT kV projection	Ten liver cancer patients	GCN	The COME was <1.2 mm, the HD was <2.9 mm, and the DSC was approximately 0.9 when assessing the relative overlap of the tumor contour between the model-generated and actual situations.
Tamura et al.[43]	Cine-MRI	Five lung cancer patients	LSTM	Succeeded in predicting the amount of movement 1.5 s (six frames) ahead.
Xiao et al.[44]	Cine-MRI	Forty-three liver cancer patients for training and internal validation as well as 5 patients for external validation	Downsampling-invariant deformable registration model	The model showed registration performance superior to all baseline methods at downsampling factors up to 500. The average regions of interest detection errors were highest in the SI direction.
Wei et al.[45]	Cine-MRI	Eight liver cancer patients	CNN	The liver DSC had a mean value greater than 96.1% ± 1.3%, and the blood vessel localization error had a mean value less than 2.6 ± 2.0 mm. The time for 3D MRI reconstruction was approximately 100 ms.
Wei et al.[50]	CBCT + kV projections	Fifteen lung SBRT patients	CNN	The average tumor localization error was <1.8 mm and 1.0 mm in the SI and lateral directions. For patients whose tumors were not visible in the X-ray projections, the mean value of the feature point localization error was <1.5 mm and 1.0 mm in both directions for all patients.
Hunt et al.[46]	Cine-MRI	Twenty-one patients with abdominal or thoracic tumors	VoxelMorph and U-Net	Registration errors were 0.067, 0.040, 0.036, and 0.032 for affine, b-spline, demons, and DL in RMSE, respectively. Computation times per frame were 51, 1150, 4583, and 8 ms, respectively.
Dai et al.[51]	Ultrasound	The CAMUS dataset, with 2D ultrasound images from 450 patients, and the CLUST dataset, with 60 patients	Markov-like network	CLUST dataset: A mean detection error of 0.70 ± 0.38 mm for 2D point landmark detection and 1.71 ± 0.84 mm for 3D point landmark detection. CAMUS dataset: A mean tracking error of 0.54 ± 1.24 mm for landmarks in the left atrium.
Roggen et al.[24]	kV projection	Twelve abdominal patients for training and one patient and one phantom for testing	Mask R-CNN	The proposed network was able to detect the motion in translational shifts within a range of 1.5 mm and rotational variations of more than 1 degree.
Zhou et al.[36]	Orthogonal kV X-ray	Ten lung cancer patients	ResNet and a feature pyramid network	The median GTV volume and motion range were 7.42 (range 1.18–25.74) cm^3 and 22 (range 11–28) mm, respectively. The mean calculation time was 85 ms/image.
Zhou et al.[35]	2D DRRs	Fourteen pancreatic cancer patients	Mask R-CNN	A mean DSC of 0.98 was achieved by the model. The mean 3D error was 0.29 mm and the mean CTV contouring calculation time was 55 ms per image.
Huang et al.[53]	2D DRRs	A public dataset for training and 97 lung patients for testing	Retina U-Net	It outperformed the benchmark model on all metrics. The comparison was significant in the 37 PCT-DRR IoU >0.2 cases, but not in the undifferentiated 84 testing cases.
Zhao et al.[25]	CBCT kV projection	Ten prostate cancer patients	DNN	Differences between the positions predicted by DNN and ground truth were 1.58 ± 0.43 mm, 1.64 ± 0.43 mm, and 1.67 ± 0.36 mm in the AP, lateral, and oblique directions, respectively.
Zhao et al.[26]	2D DRRs	Two pancreatic cancer patients	RPN	The mean absolute difference between the model predictions and actual positions was less than 2.60 mm in the AP, LR, and SI directions.

(*continued on next page*)

Table 2 (*continued*)

Reference	Image modality	Dataset	Architecture	Results
Zhao et al.[27]	CBCT kV projection	Three prostate patients	Faster R-CNN	The deviations between the target positions obtained by the DL model and the corresponding annotations ranged from 1.66 mm to 2.77 mm for the AP direction and from 1.15 mm to 2.88 mm for the lateral direction.
Terunuma et al.[37]	2D DRRs	Four lung cancer patients	CNN	The proposed method achieved accurate tumor detection of low-visibility tumors of over 0.95 based on the Jaccard index and accurate tumor detection with an error of approximately 1 mm.
Edmunds et al.[39]	CBCT kV projection	Ten lung cancer patients	Mask R-CNN	The mean diaphragm apex prediction error was 4.4 mm. The mean percentage of projection images for which a successful prediction was made was 87.3%.
Takahashi et al.[28]	2D DRRs	X-CAT phantom	fCN	The mean detection error was <0.2 mm for the X-CAT digital phantom and <1 mm for the epoxy phantom.
Hirai et al.[29]	2D Fluoroscopic images	Five lung cancer patients and five liver cancer patients.	DNN	Averaged track accuracy was 1.64 ± 0.73 mm. Accuracy was 1.37 ± 0.81 mm for liver cases and 1.9 ± 0.65 mm for lung cases. Computation time was <40 ms.
Wang et al.[30]	CBCT kV projection	Thirteen lung cancer patients	RNN	The 3D position error between model prediction and ground truth was 1.3 ± 1.4 mm, and the calculation time was 20 ms for one projection.
Terunuma et al.[38]	2D DRRs	Ten lung cancer patients	CNN	The estimated 3D 95 percentile detection error, Jaccard index, and HD for the 10 patients were 1.3–3.9 mm, 0.85–0.94, and 0.6–4.9 mm, respectively.
Grama et al.[33]	2D DRRs	3D-printed anthropomorphic phantom and six lung cancer patients	Siamese	The MAE was 0.57–0.79 mm compared to the ground truth when tested on the phantom data. As for the patient data, a correlation coefficient of 0.71–0.98 was achieved when comparing the tumor location predicted by the model with the records of the respiratory motion system.
Peng et al.[31]	2D Fluoroscopic images	X-CAT phantom	ConvLSTM and GAN	In the first comprehensive study, the mean IoU and DSC were 0.93 and 0.96. The COMEs were 0.16 cm and 0.07 cm in the SI and LR directions, respectively. In the second amplitude study, the mean IoU and DSC were 0.98 and 0.99, respectively.
Lei et al.[32]	CBCT kV projection	Ten lung cancer patients	CNN	A mean 3D position error of 2.6 ± 0.7 mm was obtained when comparing the center of mass of the tumor predicted by the model with the ground truth.
Zhang et al.[34]	4DCT	10-phase 4DCT of 58 patients + 6-phase 4DCT of five patients	Encoder–decoder model	Lung contour deformation features predicted internal targets with a localization error of 2.6 ± 1.0 mm. A localization error of 4.7 ± 0.9 mm and a real-time performance of 256.9 ± 6.0 ms were achieved by the cascade ensemble model.

Abbreviations: 1. DSC: dice similarity coefficient 2. COME: center-of-mass error 3. LR: left–right 4. SI: superior–inferior 5. AP: anterior–posterior 6. SSIM: structural similarity index measure 7. PSNR: peak signal-to-noise ratio 8. DVF: deformation vector field 9. CNN: convolutional neural network 10. HD: Hausdorff distance 11. RMSE: root mean square error 12. GCN: graph convolutional network 13. LSTM: long short-term memory 14. IoU: intersection over union 15. DNN: deep neural network 16. RNN: recurrent neural network 17. MAE: mean absolute error 18. GAN: generative adversarial network 19. ConvLSTM: convolutional layer within the LSTM layer 20. DRRs: digitally reconstructed radiographs 21. CBCT: cone-beam computed tomography.

Table 3
Summary of external surrogate signal prediction publications.

Reference	Surrogate	Dataset	Architecture	Results
Wimmert et al.[57]	Surface motions	A total of 2,502 breathing signals split into training (50%), validation (20%), and test sets (30%)	LSTM, TRANS-ENC, and TRANS-TSF	The LSTM model had the lowest normalized RMSE for all prediction horizons. However, there was a substantial spread of error across the test signals. LSTM and TRANS-ENC showed the least loss, while the performance of the simpler models' LINEAR dropped the most.
Sun et al.[58]	Surface motions	Respiratory signals acquired from 138 4DCT scans	ADMLP-NN	For predicting 500 ms ahead of prediction, the average correlation coefficients were improved from 0.83 (baseline MLP-NN) to 0.89 (ADMLP-NN). The RMSE for ADMLP-NN was reduced by 27.9% compared to MLP-NN.
Wang et al.[60]	Surface motions	Three hundred six respiratory motion traces from 103 patients	Deep Bi-LSTM	It had great performance in cases with 400 ms latency, an MAE of 0.074 mm, an RMSE of 0.097 mm, and a normalized RMSE of 0.081.
Lin et al.[62]	Surface motions	A total of 1,703 respiratory motion traces from 985 patients across three institutions	LSTM	The proposed model had superior accuracy over ANN. With a prediction window of 500 ms, it achieved an MAE of 0.037, an RMSE of 0.048, and an ME of 1.687 in internal data and an MAE of 0.112, an RMSE of 0.139, and an ME of 1.811 in external data.
Yu et al.[59]	Surface motions	Multiple external reference positions (15–20 min sampling data of 6 patients)	Bi-GRU	The prediction model guaranteed low prediction error and fast updating speed (RMSE <0.09 mm, update time <200 s). Compared to LSTM, it reduced the 30% updating time.
Tan et al.[61]	Surface motions	Three hundred four motion traces of 38 patients' respiratory motion	LSTformer	The LSTformer model had higher prediction accuracy than RNN-based models, especially for long prediction windows (RMSE reduced to 0.426 mm at 600 ms).
Wang et al.[68]	Surface motions	Seven groups of external/internal respiratory liver motion samples	LSTM	LSTM was superior to SVM for respiratory motion prediction (RMSE <0.5 mm) at a latency of 450 ms and internal liver motion prediction (RMSE <0.6 mm).
Zhou et al.[69]	Surface motions	A total of 1,079 logfiles extracted from Vero4DRT	CNN and adaptive neuro-fuzzy inference system	For CNN, ANFIS, and the regression model, the median values of MAE for 76 logfiles were 0.65, 0.66, and 1.02 mm, respectively, and the percentages of predicted target positions within 2 mm of the detected target position were 95.1%, 92.6%, and 85.6%, respectively.
Shi et al.[65]	Surface motions	Respiratory signals from 103 patients treated with CyberKnife	ECA + TCN	When the delay time was 450 ms, its MAE, RMSE, and coefficient of determination (R2) were 0.027 mm, 0.038 mm, and 97.29%, respectively.
Pohl et al.[66]	Surface motions	Nine records of the 3D positions of three external markers of individuals	Unbiased online recurrent optimization	It achieved the lowest RMSE and maximum error. These errors were 1.3 mm and 8.8 mm, and the prediction time per time step was <2.8 ms.
Sun et al.[64]	Surface motions	Two hundred two patients' respiratory signals	MLP-NN	The MLP-NN framework significantly improved prediction accuracy. The best-performing method (U1) reduced RMSE by up to 51.9% compared to traditional methods.
Yoganathan et al.[63]	Surface motions	One hundred eleven patients' respiratory signals	Dual deep RNN	The dual-RNN method using GRU achieved superior prediction accuracy compared to LSTM, with a RMSE of 0.4 $\pm$ 0.2 mm and an MAE of 0.4 $\pm$ 0.2 mm, maintaining submillimeter precision for latency periods up to 440 ms.
Teo et al.[67]	Surface motions + MV images	Forty-seven (27 in Group 1 and 20 in Group 2; unseen data)	A 3-layer perceptron neural network	The MAE was 0.59 $\pm$ 0.13 mm for Group 1 and 0.56 $\pm$ 0.18 mm for Group 2. The RMSE was 0.76 $\pm$ 0.34 mm for group 1 and 0.63 $\pm$ 0.36 mm for group 2.

Abbreviations: 1. LSTM: long short-term memory, 2. MLP-NN: multilayer perceptron neural networks, 3. ADMLP-NN: adaptive boosting and multilayer perceptron neural network, 4. RMSE: root mean square error, 5. ANN: artificial neural network, 6. MAE: mean absolute error, 7. ME: mean error, 8. Bi-GRU: bidirectional gated recurrent unit, 9. RNN: recurrent neural network, 10. SVM: support vector machine, 11. ANFIS: adaptive neuro-fuzzy inference system 12. ECA: efficient channel attention, 13. TCN: temporal convolutional network 14. CNN: convolutional neural network.

Table 4
Summary of markerless location prediction publications.

Reference	Image modality	Dataset	Architecture	Results
Romaguera et al.[72]	Cine-MRI	Twenty-five healthy subjects	Conditional-based transformer	For a prediction window of 450 ms, geometrical errors of image prediction for ConvGRU, ConvLSTM, and transformer network were 1.60 ± 1.09 mm, 1.37 ± 0.92 mm, and 1.25 ± 0.74 mm, respectively. The target registration errors of volumetric prediction were 1.75 ± 1.19 mm, 1.66 ± 1.21 mm, and 1.56 ± 1.13 mm, respectively.
Romaguera et al.[74]	Cine-MRI and ultrasound	MRI: 25 volunteers and 11 patients with hepatocellular carcinoma Ultrasound: 20 volunteers	Encoder–decoder network	The model used image surrogates for volumetric prediction and yielded mean errors of 1.67 ± 1.68 mm and 2.17 ± 0.82 mm for unseen MRI and ultrasound patient datasets, respectively.
Romaguera et al.[73]	MRI, US, and CT	MRI: liver scans from 12 subjects Ultrasound: 63 free-breathing sequences CT: 10 thoracic patients	Multiscale recurrent encoder–decoder model	The model was able to predict vessel positions in the next temporal image with a median accuracy of 0.45 (0.55) mm, 0.45(0.74) mm, and 0.28 (0.58) mm in MRI, ultrasound, and CT datasets, respectively.
Lombardo et al.[71]	Cine-MRI	Seventy patients from Group 1 for training and 18 patients from Group 1 and three patients from Group 2 for testing	LSTM	LSTM outperformed the linear regression model. For the 500 ms forecasted interval, RMSEs of 1.20 and 1.00 mm were obtained for LSTM, while the LR model yielded RMSEs of 1.42 mm and 1.22 mm for the Group 1 and Group 2 testing sets, respectively.
Huang et al.[70]	Ultrasound	CLUST	fCN with convolutional LSTM	The mean error was 0.97 ± 0.52 mm and the maximum error was 1.94 mm for 85-point landmarks. The speed per frame per landmark ranged from 66 to 101 frames per second.
Pohl et al.[75]	4DCBCT and 4DCT	Four lung cancer patients	RNN	RNN was superior to linear predictor and least mean squares. For RNN, linear prediction, and least mean squares, the maximum prediction errors were 1.51 mm, 1.8 mm, and 1.59 mm, and the RMSEs were 0.444 mm, 0.449 mm, and 0.49 mm, respectively.

Abbreviations: 1. LSTM: long short-term memory 2. ConvLSTM: convolutional layer within LSTM layer 3. RMSE: root mean square error 4. fCN: fully convolutional network 5. RNN: recurrent neural network 6. LR: left-right 7. CT: computed tomography 8. MRI: magnetic resonance imaging 9. ConvGRU: convolutional gate recurrent unit.

principal component analysis and manifold learning, combined with artificial neural networks and support vector machine classifiers, ultimately recommending principal component analysis + artificial neural networks for real-time applications. In contrast, Li et al. focused on generalized linear discriminant analysis, a supervised dimensionality reduction method that demonstrated superior robustness and computational efficiency, particularly in low-duty cycles. Notably, it generates 1D features, enabling faster inference while maintaining classification accuracy.

Additionally, Johno et al.[78] proposed a prediction-based respiratory gating framework to compensate for system latency in beam on/off control. This method employs a nearest-neighbor regression algorithm to perform multistep-ahead prediction of respiratory signals, allowing the system to anticipate and issue gating commands in advance. The simulation results demonstrated significant reductions in gating errors under regular breathing patterns. Although its structure is simple, this method shows an effective application of nonparametric machine learning in respiratory motion management.

4.2. Real-time ART

Real-time ART is one of the most advanced methods of intrafractional motion management. However, due to challenges in imaging, precision, and efficiency, it has not yet been adopted in clinical practice.

The ART workflow involves three key steps, each with its own technical challenges. First, imaging systems must continuously monitor changes in patient anatomy. As described above, DVFs can be predicted to reflect 3D organ motion and generate corresponding images and contours that form the foundation of real-time adaptation. Second, current and prior information is used to optimize the treatment plan based on the available adaptive degrees of freedom, which require efficient algorithms for real-time dose calculations. Third, the adapted plan must be delivered accurately.

Researchers have recently been working to overcome the challenges of real-time dose calculations. Traditional methods are limited by computational efficiency and are not feasible for real-time ART. All used Monte Carlo simulations as ground truth to evaluate both accuracy and speed.

Neishabouri et al.[79] introduced an LSTM model for dose prediction in heterogeneous tissues and demonstrated high accuracy (γ-pass >98%) and sub-20 ms inference times per beam, but with limited generalizability and clinical integration. Building on this study, their proposed CC-LSTM is a compact convolutional LSTM architecture capable of predicting monoenergetic dose distributions at millisecond speed per beamlet and subminute full-plan inference. The model reduced the number of parameters by 98%, and its lightweight design is more suitable for real-time ART deployment.[80] Unlike other studies focusing on proton therapy, Maier et al.[81] addressed real-time photon-based estimation for patient-specific dose distributions in diagnostic CT imaging. Their model achieved Monte Carlo–level accuracy within 1.5 s per scan without requiring acquisition parameters, enabling on-the-fly CT dosimetry. They identified that this method could be used in a similar way for real-time ART. Table 5 summarizes studies focusing on accelerating dose prediction without compromising accuracy.

5. Challenges and potential solutions

Despite the growing number of DL-driven methods for real-time motion management in RT, some challenges and possible solutions still need to be addressed.

5.1. Data limitations

Data limitations are among the most fundamental challenges in applying DL to intrafractional motion management. As models become

Table 5
Summary of real-time dose calculation publications.

Reference	Dataset	Architecture	Results
Maier et al.[81]	Whole-body CT scans from 45 patients for training and eight CT from the Visceral dataset for testing	U-Net	The method achieved dose distributions comparable to MC simulations, with γ-passing rates of ≥91%. It enabled real-time inference (~1.5 s per scan) while maintaining high accuracy across various scan protocols.
Neishabouri et al.[80]	Ninety low-grade-glioma patients	ConvLSTM	The model had a γ-pass rate of 97.8% and a peak dose deviation of 3.9%. It enabled 3.6–5.6 ms per pencil beam and a 1 s inference for per iso-energy surface.
Neishabouri et al.[79]	Water box phantoms and six lung cancer patients	LSTM	The LSTM model achieved up to a 98.6% γ-index pass rate on phantoms and 98.5% on patient data, closely matching MC accuracy. The prediction time was 1.5 ms per pencil beam.
Pastor-Serrano et al.[82]	Thirty patients (lung, prostate, and head and neck)	Dose transformer algorithm	It had a γ-pass rate of 99.37% (1%, 3 mm) for single beamlets and ≥94.7% for full treatment plans, while reducing computation time to 5 ms per beamlet and under 15 s for entire plans.
Xiao et al.[83]	Thirty-three prostate cancer patients	LSTM	The model predicted prompt gamma emission distributions with a γ-pass rate of up to 99.1% and a range deviation as low as 0.11 mm, closely matching MC results. It achieved prediction times of under 130 ms per pencil beam.
Radonic et al.[84]	Thirty-five prostate cancer patients (25 patients for training and 10 for testing).	LSTM	The model had γ-pass rates of up to 99.9% and range errors less than 0.5 mm compared to MC, enabling real-time inference at ~65 ms per pencil beam.
Wang et al.[85]	Eighty-five prostate cancer patients	Recurrent U-Net	Compared to the baseline model, the proposed model had a smaller MSE (3.03% vs. 2.05%), a higher γ-passing rate (97.42% vs. 93.69%), and a prediction time of < 0.5 s for each fold.
Zhang et al.[86]	One hundred and three prostate cancer patients and 83 lung cancer patients	3D U-Net	The beam mask method improved the γ-passing rate, and the sliding window method further improved it (for prostate cancer, targets: 96.93% ± 0.53% vs. 98.88% ± 0.49% vs. 99.97% ± 0.07%, body: 86.88% ± 0.74% vs. 93.21% ± 0.56% vs. 95.17% ± 0.59%).

Abbreviations: 1. LSTM: long short-term memory, 2. ConvLSTM: convolutional layer within LSTM layer, 3. MC: Monte Carlo 4. MSE: mean square error 5. CT: computed tomography.

increasingly complex and depend on large-scale, diverse datasets, significant variability across institutions, including differences in imaging protocols, reconstruction algorithms, registration techniques, and annotation standards, leads to high data heterogeneity. This inconsistency reduces model reliability, limits reproducibility, and impairs generalizability across clinical environments.

In addition, data privacy and security are also concerns. Most DL models rely on patient imaging and treatment data that contain sensitive information. However, legal and ethical restrictions often prevent cross-institutional data sharing, making it difficult for researchers to access the diverse, high-quality data necessary to build robust and clinically applicable models.

To address these challenges, several possible solutions can be considered. On the one hand, data standardization is essential for improving model reliability and generalizability across institutions. Recommended measures include spatial resampling to unify voxel size, intensity normalization across imaging modalities, and the use of standardized annotation protocols, such as consistent label definitions and naming conventions, to reduce intercenter variability.

On the other hand, national medical institutions can lead the development of anonymized, standardized multicenter databases by coordinating unified deidentification and annotation processes. This would support compliant data sharing and collaborative research. In parallel, an alternative solution could involve publishing pretrained models and enabling local fine-tuning at each institution. This would avoid direct data transfer, protect patient privacy, and allow models to adapt to local clinical settings. However, successful implementation requires transparency in terms of code and architecture, as well as cross-institutional collaboration.

Thus, building a technically feasible, privacy-conscious, and collaboratively managed framework for data standardization and governance is essential for the safe and sustainable adoption of DL models in clinical RT.

5.2. Computational efficiency

Another key challenge in motion management is the high demand for computational efficiency. Since the beam delivery system requires time to respond to tumor location updates, motion prediction is essential for latency compensation. The irregularity of the breathing cycle presents challenges for predicting more than 0.5 s ahead. Therefore, the American Association of Physicists in Medicine Task Group 76 report[12] defines that the total time delay of a real-time detection/compensation system should be below 0.5 s. To succeed, the system needs to do four things within an extremely short time: identify the tumor location, anticipate the tumor motion to allow for time delays in the beam-positioning system, reposition the beam, and adapt the dosimetry.

Since the first three steps can typically be completed shortly, the most time-consuming component lies in real-time dose calculation and plan optimization. DL offers a promising solution by leveraging high-performance parallel computing architectures and replacing traditional iterative algorithms with one-step regression models, significantly reducing computational complexity. Therefore, developing efficient and accurate DL-based algorithms for dose calculation and optimization is key to minimizing system latency.

However, most mainstream DL architectures remain complex and have limited computational efficiency. Future efforts should focus on designing lightweight network architectures tailored for dose calculation and plan optimization. Further improvements can be achieved by employing multi-GPU systems to accelerate processing.

5.3. Implementation barriers to real-time ART

As previously noted, real-time ART involves three key steps: continuous imaging, real-time plan adaptation, and accurate dose delivery. These processes must be performed under strict time constraints while ensuring high accuracy and robustness, which is technically demanding. While cine-MRI enables nonionizing dynamic imaging, it is still limited in spatial coverage and temporal resolution. Moreover, the real-time ART process increasingly relies on DL models. Unlike single-step tasks, such as auto-segmentation, real-time ART simultaneously integrates multiple DL modules, leaving little room for human correction. Any undetected error during DL inference could directly affect patient safety.

Workflow integration is also a major challenge. Real-time ART requires tight coordination between imaging, planning, and delivery systems, placing high demands on infrastructure and staff. Additionally, a lack of prospective clinical trials limits its regulatory approval and broader adoption.

Future research should focus on developing lightweight and reliable DL models that are compatible with existing clinical platforms. Addressing these challenges is key to making real-time ART a safe and practical solution in routine RT.

6. Discussion

Motion management techniques in RT have evolved from margin expansion to gating, tumor tracking, and, more recently, real-time ART, with each stage addressing the limitations of the previous one.

Margin expansion, based on clinical experience or 4DCT-derived motion envelopes, could not account for real-time intrafractional motion. Around the 1990s, gating techniques were introduced, enabling beam on/off by the relative relationship between the tumor and its reference position. While this improved motion compensation, its applicability was limited, particularly in nonperiodic motion (e.g., bladder filling), and treatment efficiency was reduced under free breathing.

Then, tumor detection techniques emerged in the early 2000s, progressing from invasive marker-based and indirectly surrogate-based methods to markerless methods. Each imaging modality introduced specific challenges: X-ray projection suffers from poor soft-tissue contrast and radiation exposure; MV imaging offers limited contrast and field of view; MR-Linac systems, available clinically since 2018, provide high soft-tissue contrast without radiation but are typically limited to 2D cine acquisitions, with reduced visibility outside the imaging plane.After this, beam tracking also evolved, from rigid robotic tracking (e.g., CyberKnife), which lacks dose conformality, to MLC-based tracking. However, MLC tracking lacks real-time dose feedback, hindering accurate assessments.

Most recently, real-time ART has emerged as a promising solution, integrating real-time imaging and adaptive feedback to update treatment plans dynamically. While offering the potential for highly personalized precision therapy, ART poses significant demands on imaging speed, computing capacity, and system integration.

While DL has shown promise across various imaging modalities, systematically comparing model performance among MRI, ultrasound, and X-ray remains challenging. Each modality differs substantially in acquisition principles and image characteristics, thereby directly affecting DL model design and performance.

For instance, ultrasound and cine-MRI are both time-resolved slice-based techniques, but ultrasound images are typically noisy, low in contrast, and subject to probe-induced variability. This requires DL models with high robustness and strong feature extraction capabilities. Cine-MRI, by contrast, offers high soft-tissue contrast and consistent acquisition but lacks out-of-plane motion information, demanding interpolation or reconstruction from sparse slices. X-ray projection imaging suffers from overlapping structures and poor soft tissue contrast, thereby requiring DL models to infer motion from ambiguous intensity cues.

Due to these modality-specific challenges, it is difficult to design a unified model architecture that performs equally well across all modalities. Moreover, conducting a fair cross-modality comparison using the same patient cohort is rarely feasible due to clinical, logistical, and ethical constraints. As a result, most existing studies were modality-specific, and direct performance comparisons were nontrivial. Future efforts involving multimodal datasets, standardized protocols, and harmonized evaluation criteria will be critical for enabling more systematic and clinically meaningful comparisons across imaging modalities.

Thus, with the advancement of DL technologies, many limitations of traditional methods, such as low computational efficiency and limited accuracy, have significantly improved. As DL continues to integrate with RT, these technologies are steadily evolving toward greater clinical feasibility.

7. Conclusions

The use of DL in intrafractional motion management during RT has demonstrated significant success in recent years. DL methods have been widely applied in motion detection, tumor location prediction, and corresponding management measures. Although DL methods show higher precision and efficiency compared with conventional methods, achieving the goal of accurate real-time intrafractional motion management still needs to overcome the limitations of data standardization, deficiencies, and computational efficiency.

CRediT authorship contribution statement

Yang Bining: Writing – original draft, Resources, Methodology. **Wei Ran:** Software. **Jianrong Dai:** Supervision.

Funding

This study is funded by Beijing Natural Science Foundation(Grant No.: 7242112; 1242036); the National Natural Science Foundation of China(Grant No.: 12205375; 12205376); Beijing Hope Run Special Fund of Cancer Foundation of China(Grant No.: LC2022B28).

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. Dagar G, Gupta A, Shankar A, et al. The future of cancer treatment: combining radiotherapy with immunotherapy. *Front Mol Biosci.* 2024;11:1409300.
2. Dona Lemus OM, Cao M, Cai B, Cummings M, Zheng D. Adaptive Radiotherapy: Next-Generation Radiotherapy. *Cancers (Basel).* 2024;16(6):1206.
3. Fadavi P, Nafissi N, Mahdavi SR, Jafarnejadi B, Javadinia SA. Outcome of hypofractionated breast irradiation and intraoperative electron boost in early breast cancer: A randomized non-inferiority clinical trial. *Cancer Rep (Hoboken).* 2021;4(5):e1376.
4. Kawamoto T, Shikama N, Nakamura N, Mizowaki T. The role of external-beam radiotherapy for differentiated thyroid cancer. *Jpn J Clin Oncol.* 2025;55(2):95–99.
5. Shih HA, Jiang SB, Aljarrah KM, Doppke KP, Choi NC. Internal target volume determined with expansion margins beyond composite gross tumor volume in three-dimensional conformal radiotherapy for lung cancer. *Int J Radiat Oncol Biol Phys.* 2004;60(2):613–622.
6. Li Z, Jian C, Li Y, Pan Z, Yang G, Sun X. Clinical benefits of deep inspiration breath-hold in postoperative radiotherapy for right-sided breast cancer: a meta-analysis. *BMC Cancer.* 2024;24(1):1238.
7. Daly M, McWilliam A, Radhakrishna G, Choudhury A, Eccles CL. Radiotherapy respiratory motion management in hepatobiliary and pancreatic malignancies: a systematic review of patient factors influencing effectiveness of motion reduction with abdominal compression. *Acta Oncol.* 2022;61(7):833–841.
8. Modiri A, Sabouri P, Gu X, Timmerman R, Sawant A. Inversed-Planned Respiratory Phase Gating in Lung Conformal Radiation Therapy. *Int J Radiat Oncol Biol Phys.* 2017;99(2):317–324.
9. Ohta A, Kaidu M, Tanabe S, et al. Respiratory gating and multifield technique radiotherapy for esophageal cancer. *Jpn J Radiol.* 2017;35(3):95–100.
10. 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(12):6440–6449.
11. Thorwarth D, Low DA. Technical Challenges of Real-Time Adaptive MR-Guided Radiotherapy. *Front Oncol.* 2021;11:634507.
12. Kissick MW, Mackie TR. 76 Report on 'The management of respiratory motion in radiation oncology' [Med. Phys. 33, 3874-3900 (2006)]. Task Group *Med Phys.* 2009;36(12):5721–5722.
13. Liu L, Shen L, Johansson A, et al. Real time volumetric MRI for 3D motion tracking via geometry-informed deep learning. *Med Phys.* 2022;49(9):6110–6119.
14. Liu L, Shen L, Johansson A, et al. Volumetric MRI with sparse sampling for MR-guided 3D motion tracking via sparse prior-augmented implicit neural representation learning. *Med Phys.* 2024;51(4):2526–2537.
15. Lei Y, Tian Z, Wang T, et al. Deep learning-based real-time volumetric imaging for lung stereotactic body radiation therapy: a proof of concept study. *Phys Med Biol.* 2020;65(23):235003.

16. Huang Y, Dong Z, Wu H, Li C, Liu H, Zhang Y. Deep learning-based synthetization of real-time in-treatment 4D images using surface motion and pretreatment images: A proof-of-concept study. *Med Phys.* 2022;49(11):7016–7024.
17. Zhang G, Jiang Z, Wang Y, et al. Optical surface imaging-driven tumor tracking with deformable image registration-enhanced deep learning model for surface-guided radiotherapy. *Biomed Signal Process Control.* 2025;106:107694.
18. Shao HC, Li Y, Wang J, Jiang S, Zhang Y. Real-time liver tumor localization via combined surface imaging and a single x-ray projection. *Phys Med Biol.* 2023;68(6):065002.
19. Mylonas A, Keall PJ, Booth JT, et al. A deep learning framework for automatic detection of arbitrarily shaped fiducial markers in intrafraction fluoroscopic images. *Med Phys.* 2019;46(5):2286–2297.
20. Liang Z, Zhou Q, Yang J, et al. Artificial intelligence-based framework in evaluating intrafraction motion for liver cancer robotic stereotactic body radiation therapy with fiducial tracking. *Med Phys.* 2020;47(11):5482–5489.
21. Amarsee K, Ramachandran P, Fielding A, et al. Automatic Detection and Tracking of Marker Seeds Implanted in Prostate Cancer Patients using a Deep Learning Algorithm. *J Med Phys.* 2021;46(2):80–87.
22. Motley R, Ramachandran P, Fielding A. A feasibility study on the development and use of a deep learning model to automate real-time monitoring of tumor position and assessment of interfraction fiducial marker migration in prostate radiotherapy patients. *Biomed Phys Eng Express.* 2022;8(3):10.1088/2057-1976/ac34da.
23. Chrystall D, Mylonas A, Hewson E, et al. Deep learning enables MV-based real-time image guided radiation therapy for prostate cancer patients. *Phys Med Biol.* 2023;68(9):10.1088/1361-6560/acc77c.
24. Roggen T, Bobic M, Givehchi N, Scheib SG. Deep Learning model for markerless tracking in spinal SBRT. *Phys Med.* 2020;74:66–73.
25. Zhao W, Han B, Yang Y, et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiother Oncol.* 2019;140:167–174.
26. Zhao W, Shen L, Han B, et al. Markerless Pancreatic Tumor Target Localization Enabled By Deep Learning. *Int J Radiat Oncol Biol Phys.* 2019;105(2):432–439.
27. Zhao W, Shen L, Wu Y, Han B, Yang Y, Xing L. *Automatic marker-free target positioning and tracking for image-guided radiotherapy and interventions. Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling.* 10951. SPIE; 2019:58–65.
28. Takahashi W, Oshikawa S, Mori S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. *Br J Radiol.* 2020;93(1109):20190420.
29. Hirai R, Sakata Y, Tanizawa A, Mori S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Phys Med.* 2019;59:22–29.
30. Wang C, Hunt M, Zhang L, et al. Technical Note: 3D localization of lung tumors on cone beam CT projections via a convolutional recurrent neural network. *Med Phys.* 2020;47(3):1161–1166.
31. Peng T, Jiang Z, Chang Y, Ren L. Real-time Markerless Tracking of Lung Tumors based on 2-D Fluoroscopy Imaging using Convolutional LSTM. *IEEE Trans Radiat Plasma Med Sci.* 2022;6(2):189–199.
32. Lei Y, Tian Z, Qiu R, et al. *Deep-learning-based markerless tumor localization using 2D KV/MV image. Medical Imaging 2022: Image Processing.* 12032. SPIE; 2022:914–919.
33. Grama D, Dahele M, van Rooij W, Slotman B, Gupta DK, Verbakel WFAR. Deep learning-based markerless lung tumor tracking in stereotactic radiotherapy using Siamese networks. *Med Phys.* 2023;50(11):6881–6893.
34. Zhang J, Wang Y, Bai X, Chen M. Extracting lung contour deformation features with deep learning for internal target motion tracking: a preliminary study. *Phys Med Biol.* 2023;68(19):10.1088/1361-6560/acf10e.
35. Zhou D, Nakamura M, Mukumoto N, Yoshimura M, Mizowaki T. Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning. *Med Phys.* 2022;49(3):1382–1390.
36. Zhou D, Nakamura M, Mukumoto N, Matsuo Y, Mizowaki T. Feasibility study of deep learning-based markerless real-time lung tumor tracking with orthogonal X-ray projection images. *J Appl Clin Med Phys.* 2023;24(4):e13894.
37. Terunuma T, Tokui A, Sakae T. Novel real-time tumor-contouring method using deep learning to prevent mistracking in X-ray fluoroscopy. *Radiol Phys Technol.* 2018;11(1):43–53.
38. Terunuma T, Sakae T, Hu Y, et al. Explainability and controllability of patient-specific deep learning with attention-based augmentation for markerless image-guided radiotherapy. *Med Phys.* 2023;50(1):480–494.
39. Edmunds D, Sharp G, Winey B. Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach. *Biomed Phys Eng Express.* 2019;5(3):035005.
40. 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(13):10.1088/1361-6560/ac762c.
41. Terpstra ML, Maspero M, Bruijnen T, Verhoeff JJC, Lagendijk JJW, van den Berg CAT. Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks. *Med Phys.* 2021;48(11):6597–6613.
42. 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(15):155015.
43. Tamura Y, Demachi K, Igaki H, Okamoto H, Nakano M. A Real-Time Four-Dimensional Reconstruction Algorithm of Cine-Magnetic Resonance Imaging (Cine-MRI) Using Deep Learning. *Cureus.* 2022;14(3):e22826.
44. Xiao H, Han X, Zhi S, et al. Ultra-fast multi-parametric 4D-MRI image reconstruction for real-time applications using a downsampling-invariant deformable registration (D2R) model. *Radiother Oncol.* 2023;189:109948.
45. 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(6):3584–3596.
46. 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(4):983–993.
47. Hindley N, Shieh CC, Keall P. A patient-specific deep learning framework for 3D motion estimation and volumetric imaging during lung cancer radiotherapy. *Phys Med Biol.* 2023;68(14):10.1088/1361-6560/ace1d0.
48. Dai J, Dong G, Zhang C, et al. Volumetric tumor tracking from a single cone-beam X-ray projection image enabled by deep learning. *Med Image Anal.* 2024;91:102998.
49. Shao HC, Wang J, Bai T, et al. Real-time liver tumor localization via a single x-ray projection using deep graph neural network-assisted biomechanical modeling. *Phys Med Biol.* 2022;67(11):10.1088/1361-6560/ac6b7b.
50. Wei R, Zhou F, Liu B, et al. Real-time tumor localization with single x-ray projection at arbitrary gantry angles using a convolutional neural network (CNN). *Phys Med Biol.* 2020;65(6):065012.
51. Dai X, Lei Y, Roper J, et al. Deep learning-based motion tracking using ultrasound images. *Med Phys.* 2021;48(12):7747–7756.
52. Shao HC, Huang X, Folkert MR, Wang J, Zhang Y. Automatic liver tumor localization using deep learning-based liver boundary motion estimation and biomechanical modeling (DL-Bio). *Med Phys.* 2021;48(12):7790–7805.
53. Huang L, Kurz C, Freislederer P, et al. Simultaneous object detection and segmentation for patient-specific markerless lung tumor tracking in simulated radiographs with deep learning. *Med Phys.* 2024;51(3):1957–1973.
54. Weng J, Bhupathiraju SHV, Samant T, Dresner A, Wu J, Samant SS. Convolutional LSTM model for cine image prediction of abdominal motion. *Phys Med Biol.* 2024;69(8):10.1088/1361-6560/ad3722.
55. Jin G, Liu Y, Wei R, et al. Prediction of real-time cine-MR images during MRI-guided radiotherapy of liver cancer using a GAN-ConvLSTM network. *Med Phys.* 2025;52(5):3161–3172.
56. Ghasemi Z, Samadi Miandoab P. Feasibility Study of Convolutional Long ShortTerm Memory Network for Pulmonary Movement Prediction in CT Images. *J Biomed Phys Eng.* 2024;14(1):55–66 Published 2024 Feb 1.
57. Wimmert L, Nielsen M, Madesta F, Gauer T, Hofmann C, Werner R. Benchmarking machine learning-based real-time respiratory signal predictors in 4D SBRT. *Med Phys.* 2024;51(5):3173–3183.
58. Sun WZ, Jiang MY, Ren L, Dang J, You T, Yin FF. Respiratory signal prediction based on adaptive boosting and multi-layer perceptron neural network. *Phys Med Biol.* 2017;62(17):6822–6835.
59. Yu S, Wang J, Liu J, Sun R, Kuang S, Sun L. Rapid prediction of respiratory motion based on bidirectional gated recurrent unit network. *IEEE Access.* 2020;8:49424–49435.
60. 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.
61. Tan M, Peng H, Liang X, Xie Y, Xia Z, Xiong J. LSTformer: Long Short-Term Transformer for Real Time Respiratory Prediction. *IEEE J Biomed Health Inform.* 2022;26(10):5247–5257.
62. 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(8):085010.
63. Yoganathan SA, Paloor S, Torfeh T, Aouadi S, Hammoud R, Al-Hammadi N. Predicting respiratory motion using a novel patient specific dual deep recurrent neural networks. *Biomed Phys Eng Express.* 2022;8(6):10.1088/2057-1976/ac938f.
64. Sun W, Wei Q, Ren L, Dang J, Yin FF. Adaptive respiratory signal prediction using dual multi-layer perceptron neural networks. *Phys Med Biol.* 2020;65(18):185005.
65. Shi L, Liu Y, Zhao J, et al. Lightweight attention temporal convolutional network based on multi-scale feature fusion for respiratory prediction in tumor radiotherapy. *Biomed Signal Process Control.* 2025;100:106923.
66. Pohl M, Uesaka M, Takahashi H, Demachi K, Bhusal Chhatkuli R. Prediction of the position of external markers using a recurrent neural network trained with unbiased online recurrent optimization for safe lung cancer radiotherapy. *Comput Methods Programs Biomed.* 2022;222:106908.
67. Teo TP, Ahmed SB, Kawalec P, et al. Feasibility of predicting tumor motion using online data acquired during treatment and a generalized neural network optimized with offline patient tumor trajectories. *Med Phys.* 2018;45(2):830–845.
68. Wang G, Li Z, Li G, et al. Real-time liver tracking algorithm based on LSTM and SVR networks for use in surface-guided radiation therapy. *Radiat Oncol.* 2021;16(1):13.
69. Zhou D, Nakamura M, Mukumoto N, et al. Development of AI-driven prediction models to realize real-time tumor tracking during radiotherapy. *Radiat Oncol.* 2022;17(1):42.
70. Huang P, Yu G, Lu H, et al. Attention-aware fully convolutional neural network with convolutional long short-term memory network for ultrasound-based motion tracking. *Med Phys.* 2019;46(5):2275–2285.
71. 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(9):10.1088/1361-6560/ac60b7.
72. Vazquez Romaguera L, Alley S, Carrier JF, Kadoury S. Conditional-Based Transformer Network With Learnable Queries for 4D Deformation Forecasting and Tracking. *IEEE Trans Med Imaging.* 2023;42(6):1603–1618.
73. 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.
74. Romaguera LV, Mezheritsky T, Mansour R, Carrier JF, Kadoury S. Probabilistic 4D predictive model from in-room surrogates using conditional generative networks for image-guided radiotherapy. *Med Image Anal.* 2021;74:102250.

75. Pohl M, Uesaka M, Demachi K, Bhusal Chhatkuli R. Prediction of the motion of chest internal points using a recurrent neural network trained with real-time recurrent learning for latency compensation in lung cancer radiotherapy. *Comput Med Imaging Graph.* 2021;91:101941.
76. Lin T, Li R, Tang X, Dy JG, Jiang SB. Markerless gating for lung cancer radiotherapy based on machine learning techniques. *Phys Med Biol.* 2009;54(6):1555–1563.
77. Li R, Lewis JH, Jiang SB. *Markerless fluoroscopic gating for lung cancer radiotherapy using generalized linear discriminant analysis. 2009 International Conference on Machine Learning and Applications*. IEEE; 2009:468–472.
78. Johno H, Saito M, Onishi H. Prediction-Based Compensation for Gate On/Off Latency during Respiratory-Gated Radiotherapy. *Comput Math Methods Med.* 2018;2018:5919467.
79. Neishabouri A, Wahl N, Mairani A, Köthe U, Bangert M. Long short-term memory networks for proton dose calculation in highly heterogeneous tissues. *Med Phys.* 2021;48(4):1893–1908.
80. Neishabouri A, Bauer J, Abdollahi A, Debus J, Mairani A. Real-time adaptive proton therapy: An AI-based spatio-temporal mono-energetic dose calculation model (CC-LSTM). *Comput Biol Med.* 2025;188:109777.
81. Maier J, Klein L, Eulig E, Sawall S, Kachelrieß M. Real-time estimation of patient-specific dose distributions for medical CT using the deep dose estimation. *Med Phys.* 2022;49(4):2259–2269.
82. Pastor-Serrano O, Perkó Z. Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy. *Phys Med Biol.* 2022;67(10):10.1088/1361-6560/ac692e.
83. Xiao F, Radonic D, Kriechbaum M, et al. Prompt gamma emission prediction using a long short-term memory network. *Phys Med Biol.* 2024;69(23):10.1088/1361-6560/ad8e2a.
84. Radonic D, Xiao F, Wahl N, et al. Proton dose calculation with LSTM networks in presence of a magnetic field. *Phys Med Biol.* 2024;69(21):10.1088/1361-6560/ad7f1e.
85. Wang W, Chang Y, Liu Y, et al. Feasibility study of fast intensity-modulated proton therapy dose prediction method using deep neural networks for prostate cancer. *Med Phys.* 2022;49(8):5451–5463.
86. Zhang L, Holmes JM, Liu Z, et al. Beam mask and sliding window-facilitated deep learning-based accurate and efficient dose prediction for pencil beam scanning proton therapy. *Med Phys.* 2024;51(2):1484–1498.