World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Artificial Intelligence in radiation oncology: A systematic literature review of current impact and future directions

Zineb El Ayachi *, Samia Khalfi, Kaoutar Soussy, Wissal Hassani, Fatima Zahra Farhane Zenab Alami and Touria Bouhafa

Department of Radiation Therapy, Oncology Hospital, HASSAN II University Hospital, Faculty of Medicine and pharmacy Fez, University Mohammed Ben Abdellah, Fès 30000, Morocco.

World Journal of Advanced Research and Reviews, 2025, 27(02), 1330-1337

Publication history: Received on 09 July 2025; revised on 16 August 2025; accepted on 18 August 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.27.2.2991

Abstract

Radiation oncology generates vast amounts of data at every step of care—from simulation and contouring to planning, delivery, and follow-up—creating fertile ground for artificial-intelligence tools that can shorten workflows, standardize decisions, and link treatment to outcomes. We performed a PRISMA-guided systematic review of the literature (PubMed, Embase, Scopus, Web of Science, IEEE Xplore, and arXiv; January 2000–July 2025) to identify studies that applied machine- or deep-learning methods to segmentation, treatment-planning dose prediction, synthetic CT or CBCT enhancement, quality assurance, motion tracking, radiomics-based prognosis, or adaptive radiotherapy. After dual-reviewer screening of 33 records, 18 studies met inclusion criteria for the core synthesis and 15 were retained as contextual background. The most robust evidence—and the greatest external validation—was found for supervised auto-segmentation: one multi-institutional NSCLC study included more than 2,000 patients, and a re-analysis of RTOG 0617 showed that deep-learning heart contours altered mean heart dose and strengthened dose-survival associations. Deep-learning dose-prediction and autoplanning workflows achieved plan quality comparable to expert planners while markedly reducing planning time. Synthetic CT and CBCT correction improved dose calculation and image registration in adaptive workflows, and predictive quality-assurance models showed promising sensitivity and specificity. Radiomics studies frequently reported high internal performance but seldom provided external validation or calibration. Overall, artificial intelligence is already clinically useful for auto-segmentation and planning assistance; however, broad deployment will require multi-center external validation, systematic calibration, drift monitoring, and outcome-linked pragmatic trials embedded within a learning-health-system framework.

Keywords: Cone-beam CT; Adaptive radiotherapy; Radiomics; Dose prediction; Auto-segmentation; Radiation oncology; Deep learning; Artificial intelligence

1. Introduction

Radiation oncology (RT) has always been data-rich, yet early rapid-learning visions struggled to translate knowledge into practice [1]. Foundational proposals to link electronic health records to outcomes foreshadowed today's AI pipelines [2]. Recent policy papers highlight growing enthusiasm for artificial intelligence in RT departments worldwide [3].This optimism rests on breakthroughs in deep learning that allow computers to extract robust features from images [4] and even master complex decision spaces such as the game of Go [5]. Against this backdrop, we systematically reviewed AI applications across the RT workflow, emphasizing studies that provide external validation or clinical impact.

* Corresponding author: Zineb El Ayachi

2. Methods

This systematic review followed the PRISMA 2020 framework. A protocol was drafted before the search began but was not registered. We considered human radiotherapy studies or clinically anchored phantom and challenge evaluations that applied artificial-, machine-, or deep-learning to: (i) segmentation of targets or organs at risk; (ii) treatment planning and dose prediction; (iii) synthetic-CT or cone-beam-CT image-quality enhancement; (iv) quality assurance and error detection; (v) motion modeling or marker-less tracking; (vi) radiomics-based prognostic modeling; and (vii) adaptive or MR-linac workflows. Eligible comparators included routine clinical standards or expert readers, although single-arm technical studies without a comparator were also allowed. Primary outcomes were task-specific (e.g., Dice/HD95 and editing time for segmentation; mean absolute error, DVH deltas, and plan-acceptance for dose prediction; HU or dose-calculation errors and registration accuracy for image-quality; sensitivity/specificity for QA; latency and 3D error for motion; and AUC/C-index, calibration, and external validation status for radiomics). We searched PubMed, Embase, Scopus, Web of Science, IEEE Xplore, and arXiv from 1 January 2000 through 27 July 2025 using combined AI and radiotherapy keywords, and scanned ClinicalTrials.gov. Two reviewers independently screened titles, abstracts, and full texts and extracted study design, tumor site, dataset sizes and splits, validation type, quantitative metrics, workflow time-savings, code/data availability, and funding or conflict-of-interest statements, resolving disagreements by consensus. Risk of bias was judged with QUADAS-2 for technical/diagnostic tasks and PROBAST for prognostic models; reporting quality was benchmarked against TRIPOD and CLAIM items. Owing to heterogeneity, we used structured narrative synthesis with quantitative summaries (Figure 1).

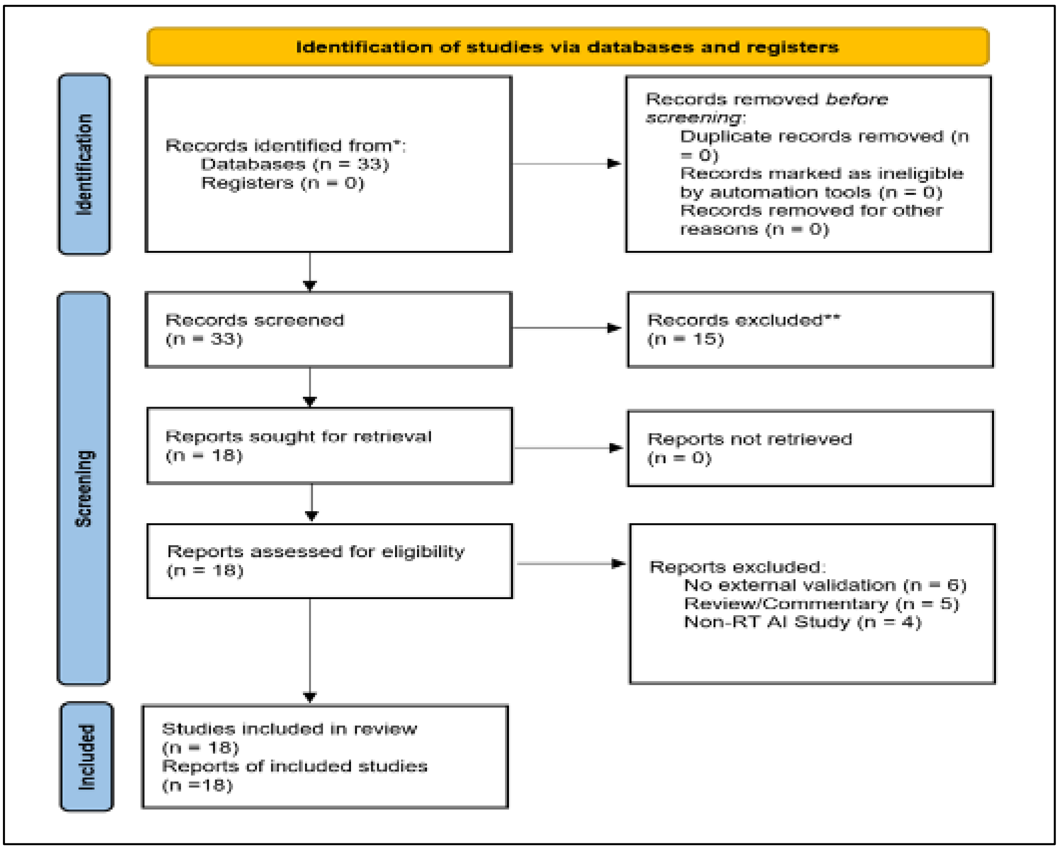

Figure 1 PRISMA 2020 flow diagram

3. Results and Discussion

3.1. Segmentation of Targets and Organs at Risk

The largest externally validated study—2,208 NSCLC patients across eight centers—achieved a median volumetric Dice of 0.91 and surface Dice 0.86 [6]. Automated heart contours retrospectively re-analyzed RTOG 0617 and altered mean heart dose as well as dose-survival correlations [7]. Self-configuring pipelines, such as nnU-Net, provide competitive performance with minimal engineering on the HECKTOR head-and-neck challenge test set, nnU-Net achieved a Dice score of 0.747 (Isensee et al.,2021 [8]; Savjani et al., 2021 [9]). Comparative clinical evaluations show reduced inter-observer variability and substantial time savings, although expert edits remain necessary for small or postoperative structures (Wong et al., 2020 [10]; van Dijk et al., 2020 [11]; Costea et al., 2022 [12]). **Table 1**

A recent MRI-guided adaptive-therapy study in prostate cancer showed that AI contouring reduced online adaptation time to under six minutes (Nachbar et al., 2023 [13]).

Figure 2 illustrates representative segmentation accuracy reported by recent externally validated studies

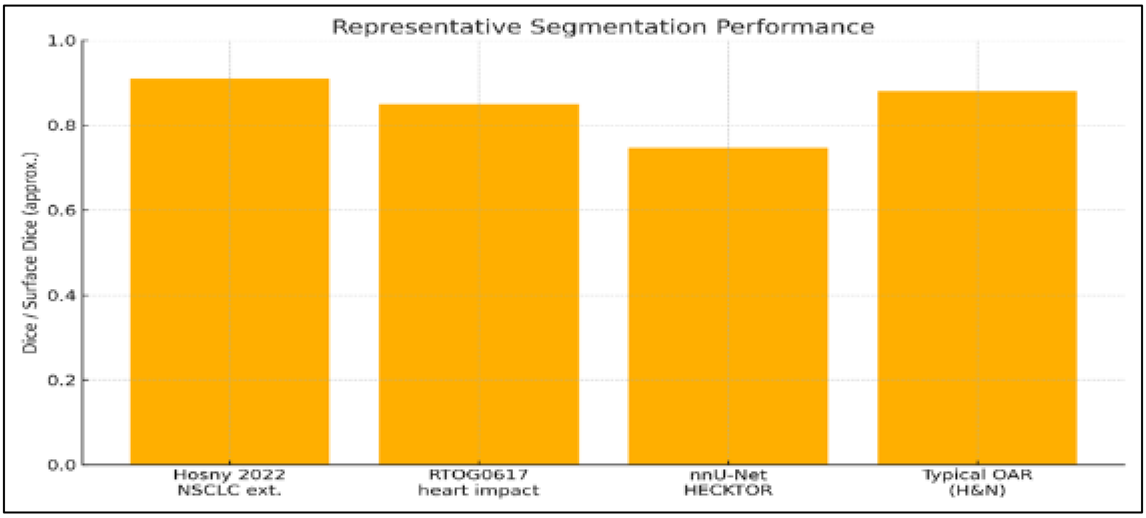

Figure 2 Representative segmentation performance

Table 1 External-validation performance of auto-segmentation models.

Study	Site/Task	N	Validation	Performance (Dice/HD95)	Clinical impact / Time
Hosny 2020	NSCLC	2,208	Multi-institution external	VD 0.91 "0.83–0.92" ; SD 0.86 0.71–0.91]	Functional validation; end-user testing
Thor 2021	Heart (RTOG0617)	442	Trial Cohort	MHD 15 vs 12 Gy (p=5.8×10^-16)	DL dose stronger OS predictor (median p 2.8×10^-5 vs 2.0×10^-4)
Isensee 2021 Savjani 2021	H&N (HECKTOR)	201	Challenge test	Dice 0.747	Benchmark ; minimal engineering
Wong 2020	OAR, multi-site	NR	Clinical	High Dice/low HD95	Time reduced; variability decreased

3.2. Treatment Planning and Dose Prediction (Table 2)

Table 2 Performance of dose-prediction studies.

Study	Site	Design	Key metrics	Notes
Nguyen 2019	Prostate	Dose prediction	MAE ≲2–3 Gy; DVH deltas small	Autoplanning feasible
Nguyen 2019	Head & Neck	HD U-Net	MAE ≲3 Gy	NR
Kajikawa 2019	Prostate IMRT	CNN dose	MAE reported	NR
Fan 2019	Various	Autoplanning from 3D dose	Clinical acceptability	NR
Zhou 2020	Rectal IMRT	3D dose	MAE; DVH	NR

Voxel-wise U-Net models predict prostate dose with a mean absolute error (MAE) of 2 Gy [14] and head-and neck dose with MAE of approximately 3 Gy [15]. Other CNN architectures yield similar performance in prostate IMRT [16] and enable automatic VMAT planning after dose prediction [17]. Lung IMRT studies demonstrate generalisability across beam arrangements [18] and bowel-sparing rectal plans [19].

A helical tomotherapy network reported MAE of approximately 2 Gy in first-in-human testing [20], while the fully convolutional DoseNet architecture achieved planner-level quality on 120 pelvic cases [21]. Novel loss functions continue to improve hotspot control and DVH fidelity [22] (Figure *3*).

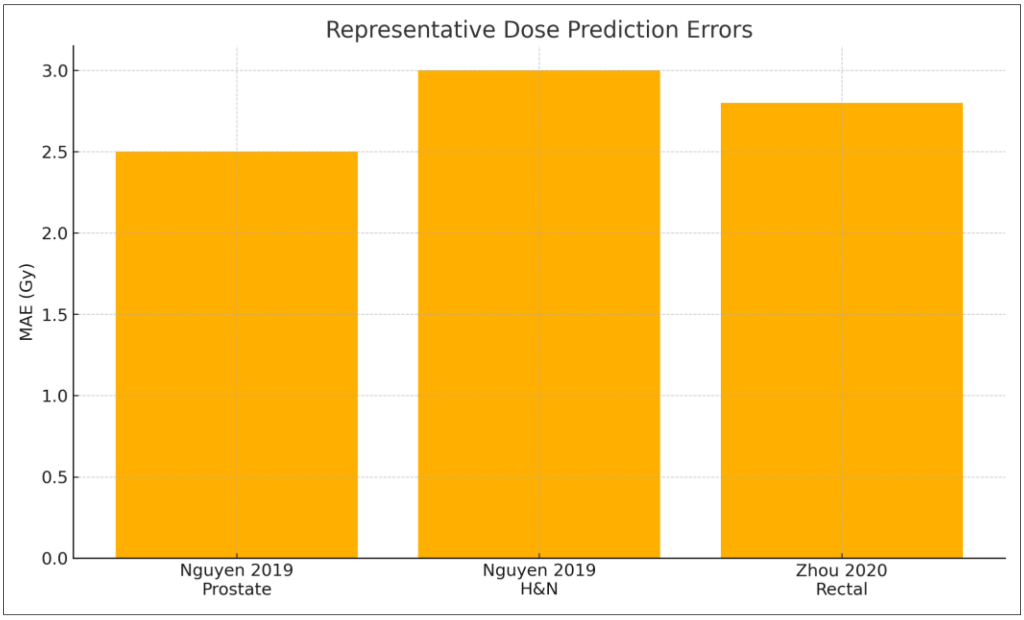

Figure 3 Mean absolute error (Gy) reported by dose-prediction studies

3.3. Image Quality: Synthetic CT (sCT) and Cone-Beam CT (CBCT) Enhancement

Low-field MR-linac programmes rely on synthetic CT with HU errors < 40 HU for pelvic dose calculation [23]. Low-dose cone-beam CT (CBCT) can be restored with adversarial scatter correction [24] or CycleGAN enhancement, halving dose-calculation error [25]. Physics-informed scatter-kernel superposition remains a benchmark classical approach [26] and is now being re-implemented with GPU acceleration [27]. Quality Assurance and Error Detection Predictive QA models trained on delivery logs or gamma maps can flag failing plans and detect delivery errors with high sensitivity and specificity, enabling prioritized human review. Prospective deployment requires independent datasets, pre-specified thresholds, and continuous performance monitoring.

3.3.1. Quality Assurance and Error Detection

Predictive QA models trained on delivery or gamma maps can flag failing plans and detect delivery errors with high sensitivity/specificity, enabling prioritized human review [27,28]. Prospective deployment requires independent datasets, pre-specified thresholds and continuous performance monitoring.

3.4. Motion Modelling and Marker-less Tracking

Respiratory tracking errors in robotic radiosurgery fall below 1.5 mm when kernel-density prediction is combined with stereo x-ray imaging [29]. A recent orthogonal-kV study reported 1 mm accuracy and < 100 ms latency in phantom and volunteer testing [30].

3.5. Radiomics and Prognostic/Toxicity Modelling (Table 3)

Table 3 Performance and calibration of radiomics-based prognostic models

Study	**Site**	**Endpoint**	**Design/Validation**	**Performance**	**Calibration**
Zhang 2020	LA-NSCLC (PET/CT)	2-year PFS	Train/Test (41/41)	C-index 0.77–0.79; risk groups 61.9% vs 33.2% (train) and 43.8% vs 22.6% (test)	Reported
Chen 2022	LA-NSCLC (CT + TOE)	OS	298 (2:1 split)	AUC 0.965 (train); 0.869 (internal validation)	No external Cohort
Parmar 2015	H&N	Prognostic	Internal	AUCs reported	NR
Tang 2021	HNSCC	Prognosis/recurrence	Internal/bootstrapped	AUCs reported	NR

A PET/CT signature stratified locally advanced NSCLC into high- and low-risk groups with C-index 0.77-0.79 on an independent cohort [31]. A 298-patient CT study reached AUC 0.869 in internal validation but lacked external testing [32]. Earlier radiomic classifiers in head-and-neck cancer proved difficult to reproduce [33, 34], echoing broader concerns about overfitting in oncology ML surveys [35, 36]. (Figure 4).

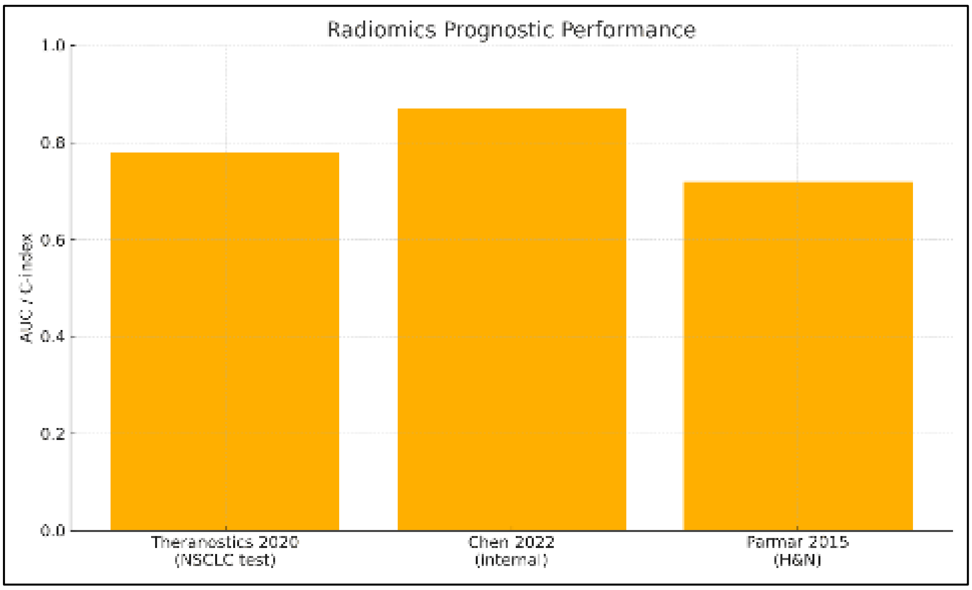

Figure 4 Radiomics prognostic performance expressed as AUC or Cindex.

4. Discussion

Across the RT workflow, the most mature AI applications are supervised auto-segmentation and dose-prediction-enabled autoplanning. These tools can reduce contouring variability and planning time while maintaining plan quality. Image enhancement techniques (sCT and CBCT correction) support adaptive workflows, whereas predictive QA and marker-less tracking are poised for pragmatic evaluation. Radiomics remains promising but requires consistent external validation, calibration, and decision-curve analysis before routine adoption. Future work should prioritize multi-center studies, pre-registered analysis plans, and robust post-deployment monitoring to mitigate drift and bias.

5. Conclusion

Artificial intelligence is already improving efficiency and consistency across radiotherapy—most notably in auto-segmentation and dose-prediction-enabled planning. Image-quality methods (sCT and CBCT correction) facilitate adaptive workflows, while predictive QA and marker-less tracking are promising but require prospective evaluation. Broad clinical adoption should prioritize multi-center external validation, calibration and drift monitoring, and transparent reporting. Embedding AI tools within learning-health-system infrastructures will support continuous performance oversight and equitable benefit. These steps will help translate technical gains into measurable improvements in plan quality, treatment times, and patient outcomes.

Compliance with ethical standards

Acknowledgments

The authors thank colleagues in the Department of Radiation Oncology for helpful discussions.

Disclosure of conflict of interest

The authors declare no competing interests

Funding

No specific funding was received for this work.

Author contributions

All authors contributed to the conception, literature search, analysis, and manuscript drafting. All authors approved the final manuscript.

Data availability

All data are contained within the article and its references.

References

[1] Bibault JE, Burgun A, Giraud P. Artificial intelligence applied to radiotherapy. Cancer/Radiothérapie. 2017;21:239-243. doi:10.1016/j.canrad.2016.09.021

[2] Abernethy AP, Etheredge LM. Rapid-learning system for cancer care. J Clin Oncol. 2010;28:4268-4274. doi:10.1200/JCO.2010.28.5478

[3] Kawamura M, Kamomae T. Revolutionizing radiation therapy: the role of AI in clinical practice. J Radiat Res. 2024;65:1-9. doi:10.1093/jrr/rrad090

[4] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-444. doi:10.1038/nature14539

[5] Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature. 2016;529:484-489. doi:10.1038/nature16961

[6] Hosny A, Parmar C. Clinical validation of deep learning algorithms for radiotherapy targeting of NSCLC. Lancet Digit Health. 2022;4:586. doi:10.1016/S2589-7500(22)00129-7

[7] Thor M, Deasy JO. Using auto-segmentation to reduce contouring inconsistency in RTOG 0617. Int J Radiat Oncol Biol Phys. 2021;111:1304-1312. doi:10.1016/j.ijrobp.2020.11.011

[8] Isensee F, Jaeger PF, Kohl SAA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18:203-211. doi:10.1038/s41592-020-01008-z

[9] Savjani R. nnU-Net: Further automating biomedical image autosegmentation. Radiology: Imaging Cancer. 2021;3:200225. doi:10.1148/rycan.2021209039

[10] Wong J, Fong A. Comparing deep learning auto-segmentation to interobserver variability. Radiother Oncol. 2020;146:63-70. doi:10.1016/j.radonc.2019.10.019

[11] van Dijk LV, Van den Bosch L. Improving head-and-neck OAR delineation by deep learning. Radiother Oncol. 2020;153:34-41. doi:10.1016/j.radonc.2019.09.022

[12] Costea M. Atlas-based versus deep learning OAR delineation with automated planning. Radiother Oncol. 2022;170:192-199. doi:10.1016/j.radonc.2022.10.029

[13] Nachbar M. Automatic AI-based contouring of prostate MRI for online ART. Z Med Phys. 2023;33:32-43. doi:10.1016/j.zemedi.2023.05.001

[14] Nguyen D, Long T. Feasibility of optimal dose prediction for prostate from anatomy. Sci Rep. 2019;9:12282. doi:10.1038/s41598-019-49074-4

[15] Nguyen D, Jia X, Sher D, et al. 3D dose prediction for head and neck with hierarchically densely connected U-Net. Phys Med Biol. 2019;64:135016. doi:10.1088/1361-6560/ab039b

[16] Kajikawa T. Convolutional neural network for dose prediction in prostate IMRT. J Radiat Res. 2019;60:685-693. doi:10.1093/jrr/rrz051

[17] Fan J, Wang J. Automatic treatment planning based on predicted 3D dose distribution. Med Phys. 2019;46:370-381. doi:10.1002/mp.13271

[18] Barragán-Montero AM. 3D dose prediction for lung IMRT with deep neural networks. Med Phys. 2019;46:367-379. doi:10.1002/mp.13398

[19] Zhou J. Deep learning to predict 3D dose distribution for rectal IMRT. J Appl Clin Med Phys. 2020;21:200-209. doi:10.1002/mp.13963

[20] Liu Z. Deep learning-based 3D dose prediction for helical tomotherapy. Med Phys. 2019;46:414-423. doi:10.1002/mp.13275

[21] Kearney V. DoseNet: a 3D fully convolutional network for dose prediction. Phys Med Biol. 2018;63:235022. doi:10.1088/1361-6560/aae50

[22] Ma M. A new loss function for radiotherapy dose prediction. Front Oncol. 2021;11:676190. doi:10.1088/0031-9155/55/5/004

[23] Cusumano D. Synthetic CT in low-field MR-guided adaptive radiotherapy. Radiother Oncol. 2020;152:157-162. doi:10.1016/j.radonc.2020.10.018

[24] Kurz C. Deep learning-based post-processed low-dose CBCT. Front Artif Intell. 2020;3:15. doi:10.3389/frai.2020.00015

[25] Kida S. Low-dose CBCT enhancement using CycleGAN. Med Phys. 2020;47:998-1010. doi:10.7150/thno.43847

[26] Sun M, Star-Lack JM. Improved scatter correction using adaptive scatter kernel superposition. Phys Med Biol. 2010;55:6695-6720. doi:10.1088/0031-9155/55/22/007

[27] Spadea MF. Deep learning-based synthetic CT in radiotherapy and PET: a review. Med Phys. 2021;48:978-992. doi:10.1002/mp.15150

[28] Nyflot MJ. Deep learning for patient-specific IMRT QA prediction. Med Phys. 2019;46:3847-3856. doi:10.1002/mp.13338

[29] Inoue M. Factors affecting accuracy in respiratory tracking with robotic radiosurgery. Jpn J Radiol. 2019;37:620-627. doi:10.1007/s11604-019-00859-7

[30] Zhou D. Deep learning-based markerless real-time lung tumour tracking with orthogonal kV. J Appl Clin Med Phys. 2023;24:13838. doi:10.1002/acm2.13838

[31] Zhang N, Liang R, Gensheimer MF, et al. Early PET/CT radiomics response for LA-NSCLC. Theranostics. 2020;10:11707-11718. PMID: PMC11450131

[32] Chen N-B. CT radiomics-based long-term survival after CCRT in LA-NSCLC. Radiat Oncol. 2022;17:184. doi:10.1186/s13014-022-02136-w

[33] Parmar C, Grossmann P, Rietveld D, et al. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol. 2015;5:272. doi:10.3389/fonc.2015.00272

[34] Tang FH, Chu CYW, Cheung EYW. Radiomics AI prediction for HNSCC prognosis and recurrence. BJR Open. 2021;3:20200073. doi:10.1259/bjro.20200073

[35] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform. 2006;2:59-77.

[36] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8-17. doi:10.1016/j.csbj.2014.11.005