Radiotherapy and Oncology 153 (2020) 55–66

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Review Article

Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance

Liesbeth Vandewinckele [a,b,1], Michaël Claessens [c,d,1], Anna Dinkla [e,1,*], Charlotte Brouwer [f], Wouter Crijns [a,b], Dirk Verellen [c,d], Wouter van Elmpt [g]

[a] Department Oncology, Laboratory of Experimental Radiotherapy, KU Leuven; [b] Department of Radiation Oncology, UZ Leuven; [c] Faculty of Medicine and Health Sciences, University of Antwerp; [d] Department of Radiation Oncology, Iridium Cancer Network, Wilrijk (Antwerp); [e] Department of Radiation Oncology, Amsterdam University Medical Center, University of Amsterdam; [f] University of Groningen, University Medical Center Groningen, Department of Radiation Oncology; and [g] Department of Radiation Oncology (Maastro), GROW School for Oncology, Maastricht University Medical Centre+

ARTICLE INFO

Article history:
Received 3 July 2020
Received in revised form 2 September 2020
Accepted 3 September 2020
Available online 10 September 2020

Keywords:
Artificial intelligence
Radiotherapy
Commissioning
Quality assurance
Auto-contouring
Treatment planning

ABSTRACT

Artificial Intelligence (AI) is currently being introduced into different domains, including medicine. Specifically in radiation oncology, machine learning models allow automation and optimization of the workflow. A lack of knowledge and interpretation of these AI models can hold back wide-spread and full deployment into clinical practice. To facilitate the integration of AI models in the radiotherapy workflow, generally applicable recommendations on implementation and quality assurance (QA) of AI models are presented. For commonly used applications in radiotherapy such as auto-segmentation, automated treatment planning and synthetic computed tomography (sCT) the basic concepts are discussed in depth. Emphasis is put on the commissioning, implementation and case-specific and routine QA of AI models needed for a methodical introduction in clinical practice.

The radiotherapy (RT) workflow is a complex process consisting of several time-consuming steps that have an impact on treatment quality and hence patient outcome. Artificial intelligence (AI) has been proposed as a tool to increase quality, standardization and acceleration of these steps leading to a more safe and accurate radiation administration by automation and optimization of workflows [1–3]. Especially with the introduction of adaptive radiotherapy (ART), a streamlined workflow is mandatory in clinical routine. AI is characterized as a collection of algorithms that perform tasks correlated with human thinking or intelligence [4] with machine learning (ML) and deep learning (DL) as subdomains [5]. Several review papers have been published on the use of AI, ML and DL in radiotherapy [6–12]. However, not much is written on clinical implementation of these new techniques [13,14].

Recently, a survey on the clinical use of AI in radiotherapy [15] revealed that most popular AI supported applications were automatic segmentation and treatment planning, followed by synthetic CT (sCT) generation. It also revealed a demand for guidance on the implementation of AI in clinical practice. Therefore, the aim of the current paper is to provide recommendations on the use of AI in radiotherapy focussing on automatic tumor and organ-at-risk (OAR) segmentation, automated planning techniques and sCT generation (Fig. 1). General and application-specific recommendations on commissioning, implementation and quality assurance (QA) are both described in detail[2].

General recommendations

The recommendations described below follow the typical scenario for introducing new technology in clinical practice: starting with the commissioning phase of the AI-based application, followed by the clinical implementation phase and finally the daily use of the AI model together with model and case-specific QA.

* Corresponding author at: University Medical Center, Department of Radiation Oncology, Amsterdam UMC, De Boelelaan 1118, 1081 HZ, Amsterdam, The Netherlands.

E-mail addresses: liesbeth.vandewinckele@uzleuven.be (L. Vandewinckele), michael.claessens@uantwerpen.be (M. Claessens), a.dinkla@amsterdamumc.nl (A. Dinkla), c.l.brouwer@umcg.nl (C. Brouwer), wouter.crijns@uzleuven.be (W. Crijns), dirk.verellen@gza.be (D. Verellen), wouter.vanelmpt@maastro.nl (W. van Elmpt).

[1] These authors contributed equally to this work.

[2] This work commenced during the 3rd European Society for Therapeutic Radiology and Oncology (ESTRO) physics workshop on 'Implementation/commissioning/QA of artificial intelligence techniques' in Budapest (2019)

https://doi.org/10.1016/j.radonc.2020.09.008
0167-8140/

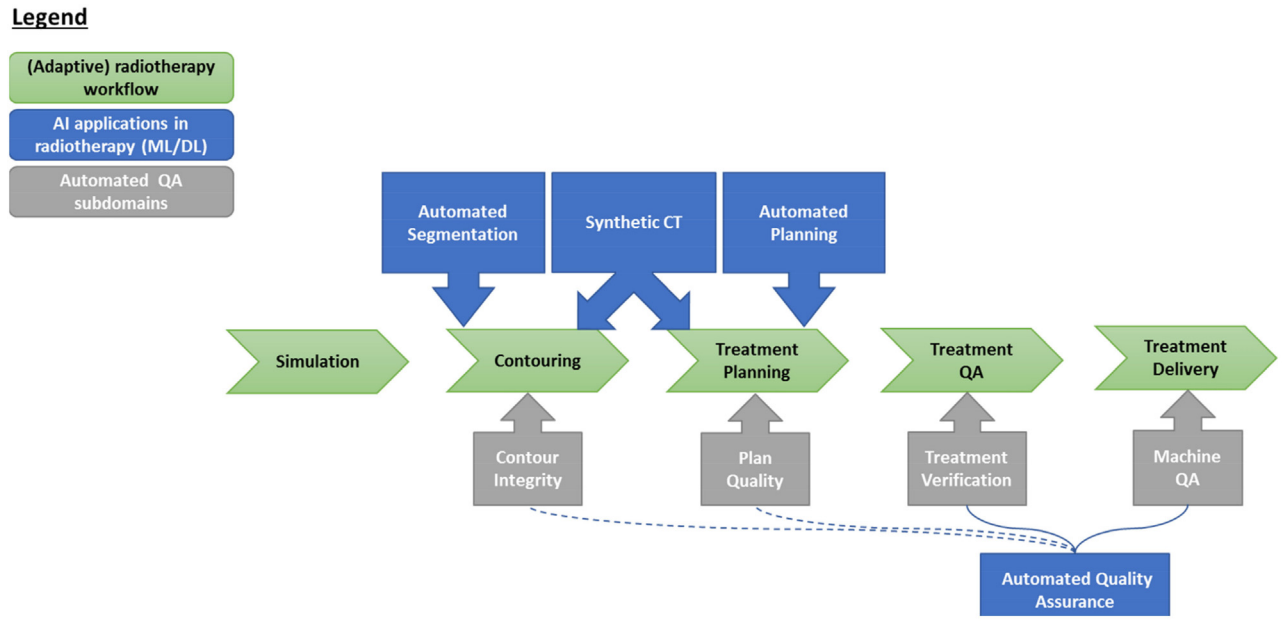

Fig. 1. Current use of selected AI applications in the radiotherapy workflow discussed in this paper. Other workflow steps that could benefit from AI such as image reconstruction, registration, etc. that are not discussed in this paper are not shown in this figure.

Commissioning

The commissioning procedure is two-fold, to train an AI-algorithm/model and to investigate the accuracy and reproducibility of the model prior to clinical use. As shown in Fig. 2, this commissioning procedure can be divided into a training/validation phase (first phase) and a test phase (second phase)[3] [16]. The phases that need to be completed depend on whether the AI model has been built in house, in collaboration with a vendor or was commercially available.

The training/validation phase is performed to tune the model to the clinical need. This phase needs to be completed only when the model is built in house or when a previously trained model allows for customization. Training of a model is generally accomplished with a large amount of data, preferably of high quality and annotated [2]. Using locally acquired data offers the advantage of preserving the department's clinical guidelines and (imaging) protocols. The data should be reviewed to ensure that it is a curated representation of the patient population and clinical practice under consideration, which can require triage to generate reliable subsets samples. In practice, well-known, not too large, high-quality datasets are generally preferred over very large datasets of lower quality due to the evolving nature of clinical protocols and guidelines. Validation of the model is accomplished using both a quantitative and a qualitative analysis on a smaller set of data: i.e. the validation set [16]. This set should represent the data on which the model will be applied clinically. Consider a range of metrics to evaluate (in addition to those used for training) tailored to the application of the model. In general, one starts with evaluating quantitative metrics by comparing the model output to the clinical data. Once satisfied, the results of the model are presented to clinical experts (e.g. physicians, physicists or radiotherapy technologists (RTTs)) for revision. This step may reveal where the model and the expert disagree on the prediction, which can be incorporated in the model to improve the results [17]. Models that are built in collaboration with the vendor or are commercially available do typically not allow or require this phase, but should then be accompanied by proof of FDA/CE marking [18] and a vendor's validation report detailing the performance of the model. This baseline set of investigations forms the starting point for further clinical commissioning. Availability of detailed information concerning model intended use and limits, description of training and validation set, used standards (if applicable), metrics and overall validation protocol is highly recommended. Lastly, the vendor should provide an annual accuracy assessment protocol, that describes the procedure for the customer to follow on a yearly basis.

The goal of the test phase is to obtain an independent evaluation of the final performance of the model, investigate the robustness of the model and define for which (type of) patients the model can be applied. This phase should be applied to all AI models used clinically. An independent dataset, which should represent the data for which the model will be used clinically and show similar variation as in the training data, is used to evaluate the model on a qualitatively and quantitatively manner. A consensus is hard to define for the amount of patients to include in the test set [15], depending also on the variation of the input data. Nevertheless, a minimum selection of ten patients is recommended as a good starting point, which can be adjusted in case a large variation in the results is present. Case-specific QA should cover the detection of outlier results. No adaptation of the model should be made to improve performance that is based on these specific patient cases. However, if model modifications are deemed necessary based on this test, a new and independent test set should again be collected after modification of the model [16]. Optionally, other relevant endpoints such as the reduction of interobserver variation and/or time-saving could be evaluated in this phase as well.

Implementation

Prior to clinical implementation, it is recommended to involve a dedicated multidisciplinary team of relevant expertise (e.g. physician, physicist, RTT, IT specialist) to ensure safe and clinically relevant use. This team should have a basic knowledge of AI/ML/DL in general and an understanding of the particular model including the

[3] Note that the used terminology about validation/test is typically used by ML experts. Clinical experts sometimes turn this around and use test instead of validation and vice versa.

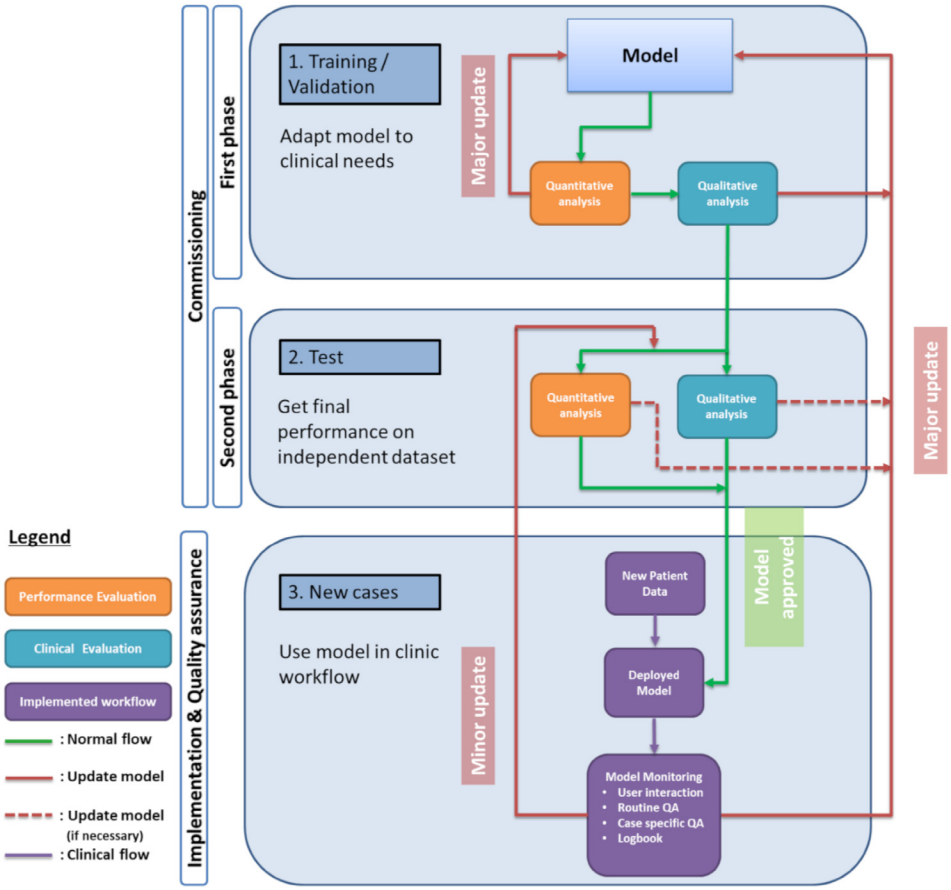

Fig. 2. Workflow for the commissioning, implementation and QA of a new AI model in the clinic. Commissioning starts already during (in-house) development of a model. When the model is built, or when using a previously developed model, commissioning starts in the second phase with an independent test set. For major model updates, one should go back to the first phase. For minor updates, only the second phase is repeated.

patient cohorts for which it is applicable to evaluate the strengths and possible limitations. Furthermore, this team should guide the users and provide training and education for correct use and interpretation of the model output. It is recommended to perform an independent end-to-end test [19] prior to clinical implementation executed by clinical personnel that will actually use the application. Lastly, it should be stressed that it is important that all users know and understand what the exact intended use and scope of testing of the application is.

In Fig. 2, a potential workflow to implement an AI model in the clinic is presented. A feedback system is important to maintain safety and quality as clinical practice may change over time. For minor adaptations that do not have a large influence on the output, such as changes in post-processing, only the test phase of the commissioning stage needs to be performed and will reveal whether additional recommissioning is necessary or not. A change in the clinical workflow or a systematic reduction of the performance [20], may require retraining with a new and updated training set. Afterwards, it is important to (re)commission the model with a new validation and test set and repeat the end-to-end test.

During the first weeks of clinical use, it is recommended to hold regular meetings between the users of the model and the implementation team. This way, minor issues in the implemented workflow can be addressed in a timely manner. It is important to keep in mind that while AI brings consistency, systematic errors can remain present if undetected during the implementation phase.

As for any new tool, it is strongly recommended to perform a risk analysis before any model is implemented (as per Euratom Directive 2013/59) [21] by analyzing the risks, the implementation is generally improved by early detections of possible risks or malfunction and made more robust against failure. A well-known method is a Failure Mode and Effect Analysis (FMEA) [20,22–24] or Risk Analysis Matrices [25], which includes a brainstorm on the potential risks with people from all disciplines dealing with implementation and use of the model. For AI tools, risk classification will be influenced by the user control and interpretability of the model, to what extent humans are involved in the (QA) workflow and the presence of safety barriers. Moreover, alternative models or independent tools for the intended application should be considered in case of failure or misbehavior.

Quality assurance

After successful implementation of any AI-based application, regular QA is highly recommended. We distinguish between ‘case-specific QA’ and ‘routine model QA’.

Case-specific QA

Case-specific QA refers to patient specific or per machine QA. The performance of the model is estimated during commissioning for known situations. However, this does not guarantee the desired behaviour in new, unknown situations. Depending on the interpretability of the method, the user can choose to perform a more comprehensive evaluation. Many of the current AI models appear as black boxes to the end-users. More transparent methods are generally easier to interpret and more straightforward to verify [26]. However, there is typically a trade-off between transparency of the model and (potential) accuracy, since more complex models

with more parameters tend to have higher ability to fit to the data, possibly at the cost of generalizability [27].

Depending on the application of the model, the quality of the output of the model can be checked in different ways. Currently, supervision of the output is seen as one of the most important tools. In some applications, like contouring and planning, the output of the model can be adjusted by the users to their preference. It is recommended to log the corrections made by the clinical staff, to keep track of poorly performing cases and to log errors made either by the model or the user itself. In this way, potential risks in the model or implementation workflow can be found to update and improve the model (Continuous Quality Improvement) [28,29].

Whether a new patient (data) is eligible for an AI-based application or not, is in the first place the decision of the user. Involvement of all users and a shared knowledge of the intended use and limits is therefore critical. To help in this decision, an (automatic) similarity check comparing the new data set to the training data could tell if there is reason to doubt performance. Alternatively, there are possibilities to perform plausibility or sanity checks on the model outcome: e.g. an independent, secondary algorithm can be used to benchmark the performance of the clinical (AI) model and point divergent behaviour; or automatic case-specific QA tools can be utilized to facilitate the detection of outliers. For this plausibility checking, it is foreseen that AI could also play a role for current non-AI based steps in the workflow. Finally, uncertainty estimates of the AI output can be used as a valuable tool to flag outcomes that require additional verification [30,31]. It is important to note that these methods are under investigation and that supervision is currently the main tool.

Routine model QA

Setting up a Quality Management Program (QMP) [32] is recommended mainly to monitor if the model has not been unintentionally modified and to verify that the model is still valid after a (minor) software update. A reference data set should be selected at the time of commissioning for this purpose, reflecting clinical practice. However, a QMP also helps to detect changes in the workflow (e.g. changes in imaging device, protocol, immobilization) and review their impact. The reference dataset should be re-predicted on a regular basis and compared to the initial predictions during commissioning (end-to-end performance) to verify consistency of the model.

Additionally, an informatics platform or log file should be created for structured data collection of cases in which the performance of the model was suboptimal, as described in Section Case-specific QA. This allows identification of the limitations of the model, trending and may also facilitate future model revisions to improve the performance for such patient cases i.e. post-market surveillance [33].

Automatic segmentation

Accurate segmentation of OARs and target volumes is the starting point of radiation treatment planning. Manual segmentation is a time-consuming task with high intra and interobserver variability both within [34] and across [35] radiotherapy centers. Recently, convolutional neural network (CNN)-based auto-segmentation models have been shown to improve consistency and efficiency of this process [36,37]. These models typically classify every voxel in an image as belonging to an OAR or target based on features of the position and intensity of the voxel and surrounding voxels [38–40]. These DL models are now outperforming traditional auto-contouring methods and reaching the accuracy range of manual delineations [41].

Commissioning

The number of patients needed to train a segmentation model depends on the variability within the data, the completeness of the data according to missing labels and the used AI model. A model for tumor segmentation generally needs more patient data than for OARs, because tumor shape and location are more variable with respect to normal anatomy. Currently, state-of-the-art CNN-based contouring models typically consist of more than a 100 patients [40,42]. However, also models of 50–100 patients have been shown to segment OARs with reasonable accuracy [36,38]. Variability in the training data should reflect the variability of the clinical data for which the model will be used. For example, if the model will be used for different imaging acquisition protocols or devices, the training set should include all of these data types. Pre-processing, such as resampling or cropping, can be used to tackle varying pixel spacing or Field of view (FOV). Post-processing, such as connected component selection, hole filling, or smoothing, can be performed to obtain more clinically relevant contours [38,41,43]. Validation and test sets typically contain around 20 patients [43]. A minimum of 10 patients is recommended, but has to be increased in case large variation in the result is present [44,45].

To specifically analyze the performance of the auto-segmentation model, quantitative and qualitative validation are appropriate (Fig. 2) [36,37,42,46] and should be combined. Quantitative analysis to determine agreement between clinical contours and auto-segmentations is accomplished using similarity metrics (see Table 1). For an overview of metrics related to image segmen-

Table 1
Commonly used quantitative metrics for the commissioning process of auto-segmentation. The results are commonly be compared to the values obtained by inter and intraobserver variability, and in the case of a commercial model to the values specified by the vendor's validation report.

Technique	Evaluation	Advantages/Disadvantage
Overlap metrics	Overlap of the contour volume/ surface (e.g. Dice Similarity Coefficient (DSC) or Surface DSC) [50,51]	Practical during training/validation Easy to calculate Depends on volume; every voxel is equally important; no specific focus on the border of the organ. Not well correlated to subjective acceptance of contours Surface DSC computationally more demanding
Distance metrics	Distance between segmented and true label volume (e.g. Hausdorff (max) Distance (HD) or Mean Surface Distance (MSD) on registered images) Added Path Length: The amount of adjusted path length of the contour [52]	Focus on the boundary of the contour Eliminates the impact of outliers Not dependent on absolute volume Difficult to interpret for small contours
Volume	Comparison of volume (systematic under/over segmentation)	Easy to calculate and interpret No relation to location
Dose	Dosimetric impact of delineation uncertainty [48]	Clinical impact of difference between delineationsLabor intensive

tation, we refer to [47]. However, more clinically meaningful metrics for radiotherapy exist as well, such as the resulting dose differences due to discrepancies or imperfections in contours [42,48] or an efficiency metric such as the time needed to make contours clinically acceptable [43]. In terms of clinical acceptability, a qualitative analysis should be performed based on head-to-head comparison of manual contours and auto-segmentations for each patient by one or more radiation oncologists (RO) or RTTs [41,46]. Head-to-head comparisons could also be performed in a blinded fashion with a modified Turing Test approach [42,49]. Scoring of the auto-contoured OARs or targets as a 'pass' or 'fail'; or estimated adjustments needed to reach clinical grade quality would give a subjective estimate of the quality of the auto-segmentation model [43]. The required quality of segmentations depends on the goal: whether it is time saving to reach clinical acceptance or completely eliminating manual intervention. Due to persisting intra-and interobserver variability, achieving an accuracy comparable to this variability is generally considered as sufficiently accurate. When time saving is the rationale, knowing how much manual editing is required (or not) is an important result [42].

Implementation and QA

Case-specific QA

Every automatically generated contour should be reviewed, corrected if necessary, and approved by clinical staff. Besides manual verification, methods exist to facilitate case-specific QA by highlighting outliers up front [53]. For instance, a statistical model can be used to detect outliers by evaluating structure shape, volume and centroid of automatically generated contours [54]. Alternatively, implementing a secondary independent segmentation method may help to reveal segmentation difficulties when differences are present between the two segmentations [55]. Finally, AI-based QA methods can estimate/classify uncertain or potentially incorrect segmentations and present them to experts for revision [7,30,56]. Although these methods are not intended as a complete QA, they facilitate the identification of outliers. This may also provide guidance on which contours actually need editing and which are sufficiently accurate [42].

Routine model QA

Regular tests as described in Section Routine model QA should be performed. An extra test set should be selected if the imaging acquisition protocol changes; (e.g. everything that could influence the quality of the generation of the patient 3D imaging: CT/magnetic resonance imaging (MRI) protocol or scanner, patient positioning, FOV, input sequences or fixation aids). The test set should contain patient images acquired with the new workflow, and the model output for this new test set should be reviewed by clinical staff. Recommissioning of the model might be warranted.

Automated treatment planning

Radiotherapy treatment planning contains an optimization problem having many degrees of freedom. It typically requires advanced skills, is labor intensive [57] and associated with large user variability [58]. Developments in AI have led to applications in the field of treatment planning to decrease human intervention and workload, improve plan quality and consistency [59,60]. In addition, it could enable comparisons of treatment techniques with minimal bias, study eligibility and shared/informed decision making for personalized treatment planning (e.g. patient selection) [61,62].

This section focuses on all data-driven automated treatment planning, often referred to as knowledge-based planning (KBP). KBP is considered as any method that uses previous knowledge, i.e. previously optimized plans to predict the treatment plan or dose for a new patient. The dose can be predicted in terms of dose-volume histograms (DVHs) for a contoured structure, or complete dose distributions. DVHs are typically predicted using a ML model based on patient geometry features [63,64]. Predictions of dose distributions are done using ML [65] or DL techniques [66–69], subsequently used to drive an automated optimization procedure resulting in clinically deliverable solutions. Other automated planning techniques using DL (e.g. fluence map prediction, multicriteria optimization (MCO), beam orientation optimization are also being described [70–72]. In this section, other ways to automate treatment planning, such as scripting or protocol-based iterative planning are not discussed.

Commissioning

Automated treatment planning with AI should result in technically deliverable and clinically acceptable plans. The starting point of any ML-based automated planning model is a patient cohort planned according to a pre-defined protocol or class solution with consistency in tumor site and treatment setup, but also with enough variability in patient geometry. Hence, the dataset should represent the whole population for a specific treatment site and prescription. The class solution could optionally be categorized to 'easy', 'intermediate' or 'hard' to optimize plans. Triaging into subcategories makes the model more specific, while combining the categories makes the model more robust to clinical practice.

The type of model defines the amount of required training data. Classical ML models need less training data than DL models since they have less parameters to optimize: ML models are trained with 20–100 patients in current practice [59,65,73–75]. Including more patients could enlarge the possible variability in the training data and therefore the range of patients for whom predictions will be satisfactory [74], although several studies have also reported satisfactory results with smaller training sets [73,75]. DL models on the other hand, typically require more than 100 patients for training [67], but efforts are ongoing to develop strategies to bypass this requirement (around 80 patients) [66]. Pre-processing steps like dose normalization help to decrease the dependency of the prescribed dose and fractionation scheme [76]. Irrespective of the method, it is important that (clinical) plans are reviewed and curated to ensure protocol compliant delineations and plan quality. Lastly, validation and test sets typically consist of minimally 10 patients for both types of models [77]. In case of large variation within the data and/or results, it is advisable to evaluate more patients [62,78].

To analyze the performance, both a quantitative and qualitative analysis should be performed. The former could be based on clinical guidelines and consists of the calculation of DVH parameters/-clinical constraints, conformity index, homogeneity index, plan quality index [60], amount of monitor units (MU), etc. (Table 2) and compared to plan acceptance criteria and/or clinical plans [59,73–75]. In terms of clinical acceptability, reviewers should be able to directly compare DVHs, clinical dose goals and dose distributions via correlated scrolling in addition to dose difference maps and should score/rank the ML plans according to criteria as overall approval, target coverage, OAR sparing, high dose conformity, dose gradient, etc. Depending on the optimizer (whether all delivery constraints are taken into account: dose rate, leaf speed, gantry speed, interdigitation, couch position etc) it might be necessary to perform pre-treatment QA of the predicted plans to assess deliverability.

Table 2
Quantitative metrics for the evaluation of automated treatment planning models. The results should be compared to international and institutional planning protocols.

Technique	Evaluation	Advantages/Disadvantages
DVH parameters	Local/institutional constraints to review the clinically acceptable DVHs	Clinically relevant Generally automatically displayed Gives no information on spatial dose distribution Institution dependent
Isodose lines	Qualitative observation of isodose lines or color wash of the dose distribution	Show overall dose distribution May reveal differences in shape Not quantitative
Conformity index	Degree to which the high dose region conforms to the target volume [79]	Clinically relevant Easy to calculate It does not always take dose outside the target into account (depending on definition)
Homogeneity index	The uniformity of the absorbed dose distribution in the target [79]	Easy to calculate No information about location of hot/cold spots No information about OAR
Number of monitor units	Based on the amount of radiation required from the linac (including scatter, absorption etc.)	Metric for complexity/modulation Influenced by target volume and location Not always comparable between energies/linacs/institutes
Plan quality index	Combination of components that describe healthy tissue conformity, target coverage and sparing of critical organs [60]	Possibility to differentiate between "good" and "bad" plans Weighting of individual parameters may differ between institutes/physicians
(Blind) rating of plans by RO	Rate or order plans from best to worst, or choose preferred plan	Clinically relevant Subjective
Dose gradient	Local dosimetric differences as a function of the dose gradient [80]	Highlight differences (algorithm and delivery related) in sensible areas Not easily available in commercial offerings

Implementation and QA

Case-specific QA

Every automatically generated plan should be reviewed and approved by the clinical team before clinical use, considering automatically generated treatment plans as if they are designed by an RTT in training. Additional planning or optimization steps might be required to obtain better target coverage or sparing of OARs. There is potential to automate this QA step [81], or the other way around: use the AI-based method as an independent benchmark of the (manual or automated) clinical plan [82]. Algorithms that predict the obtainable DVHs only, are regularly used as a benchmark to assess whether better target coverage or OAR sparing is possible [82–85]. Note that it is important to keep planning and validation (QA) algorithms independent of each other. Therefore, one should not use the same algorithm for planning as for QA.

Routine model QA

It is important to note that an ML model is trained on treatment plan data for a specific treatment technique according to fixed protocols and objectives. Changing to a different treatment technique could entail that a new model should be trained and validated (recommissioning). To verify the robustness of the model to this new workflow, a new test set needs to be obtained within the new clinical workflow (e.g. change in treatment device, fractionation scheme, technique, beam set-up). The output of the model for this test set should be reviewed by the clinical staff. One needs to check if the accuracy of the model output is maintained and plan optimization leads to equally good plans. If not, manual (re)planning many patients within the new workflow is required to be able to recommission a new model, which may hamper continuity and increase workload. Finally, a TPS update or protocol change may also be a reason to perform recommissioning.

Synthetic CT

Synthetic CT is often used to allow for (improved) accuracy of dose calculations on Cone-Beam CT (CBCT) or MRI images [86]. CT images are the standard for radiotherapy dose calculations, as they provide robust information about the electron density of tissues [87]. Obtaining accurate CT numbers (Hounsfield units, HU) for non-CT data opens the door for MR-only workflows [88–90]. Synthetic CTs can be generated using voxel, atlas based or hybrid approaches [91–93]. Thus far, the most practical and common approach is deformable image registration to map planning CT HU to the (daily) treatment image. Recently, it has been shown that CNNs (DL) provide promising methods for synthetic CT generation based on CBCT or MRI images [86,94–99].

Commissioning

The starting point of a robust DL sCT model is typically a dataset consisting of images of a single FOV, acquired using the same image acquisition protocol on the same image devices. However, recent studies have shown that some of these requirements can also be circumvented [100]. Generally, there should be a good correspondence between the MRI or CBCT images and the CT scan. Models for sCT generation were traditionally trained in a supervised manner with 25–40 patients, but unpaired training has shown to also eliminate or loosen the requirement for perfect alignment [101,102]. A patch-based approach to training the data can be beneficial for smaller datasets or if 3D training is desired, but appropriate weighting of the patches may be of concern [103,104]. To establish a paired training set, one should carefully check the (voxel-wise) alignment of patients and possibly improve this alignment by further image processing such as deformable registration [96]. Selected patients should also preferably be positioned with the same immobilization devices present on both imaging modalities. If not, post-processing of the images may be required. Additionally, training data should be reviewed for image artefacts, investigating whether these may lead to inaccurate dose calculation. An end-to-end test helps to determine if there are other requirements on the sCT (for instance, some systems expect square pixels of the (synthetic) CT). After training, one should establish strict inclusion and exclusion criteria for its use, which implies for example patients with dental implants causing signifi-

Table 3
Quantitative metrics for the evaluation of synthetic CT models.

Technique	Evaluation	Advantages/Disadvantages
Mean error (ME)/Mean absolute error (MAE)	Mean (absolute) difference between HU values of synthetic CT and ground truth CT: Paired/voxel-wise comparison within a specified volume (e.g. body contour or other structure/region of interest)	Is usually reported Relatively easy to calculate Can be calculated within different structures/regions of interest Does not show the spread in the voxel-wise differences Difficult to compare between studies Might not be clinically relevant Gives information of absolute error and not of the relative error
Peak signal to noise ratio (PSNR)	Ratio between maximum value of a signal and the power of distorting noise that affects the quality [111,112]	Easy to calculate Gives some information about the relative error No information about position of the error
Structural similarity metric (SSIM)	SSIM is used for measuring the similarity between two images and designed to improve on metrics like MSE and PSNR [113]	Carries information about inter-dependencies between pixels More difficult to calculate
Dice coefficient of bony structures	Overlap volume of the bony structures	Relevant for dose calculation Relevant for positioning Dice is dependent on the volume of the structures Dependent on thresholding of bones Gives no information on actual HU values
Dose difference in DVH	Calculate the DVH of the same plan on both synthetic CT and ground truth CT, using either the same set of delineations or 2 different structure sets. In the latter case newly delineated or obtained by (deformable) registration.	Easy to calculate if both DVHs are available Care should be taken in transferring/warping structure sets A difference in DVH or DVH parameter can be caused by an 'error' in the sCT or in the contour, or by a difference in the anatomy
Dose difference using gamma index	Calculate the gamma value in every point in the image (volume)	Used to give an overall representation of error usability of the synthetic CT Analysis can be tweaked to preference (dose difference, DTA value and threshold) Difficult to compare between studies
Matching accuracy	Compare matching values at the linac of (CB)CT and sCT	Vital information for accurate patient positioning True matching value is unknown: no gold standard available No standard on the procedure, e.g. handling of residual misregistration [114]

cant streaking artefacts, or postoperative patients with metal screws that lead to signal voids on MRI images. In literature, the validation and test sets typically consist of at least 10 patients, no real consensus seems to exist yet [95,104–109]. This minimum number also depends on how the sCT is situated in the workflow.

sCTs should be evaluated in terms of image similarity, geometric fidelity and dosimetric accuracy [110]. Several metrics exist to quantify these aspects (Table 3). Almost all studies calculate mean absolute error (MAE). Care should be taken to compare these values between studies, since there are many factors influencing this result [110]. In addition, it is common to report the standard deviation of the mean (across patients), but not the spread in HU differences on a patient level, which is usually much larger. To interpret differences in dose, it is important that CT and sCT are in the same frame of reference and that body contours and contours are the same (or at least well known). Finally one should check that matching accuracy for sCT-based positioning at the linac is at least as accurate as in the routine workflow.

Implementation and QA

Users of sCT models should have knowledge of imaging modalities to be able to detect artefacts and their associated causes. If post-processing of the sCT generated by the (DL-based) model is performed or required, it is important to know the details and its impact (on dose calculation). Knowledge of dose calculation and electron density tables helps to establish the required accuracy of the synthetic CT. In some cases, simple models may provide sufficiently accurate dose calculations [115].

Case-specific QA

Every sCT generated should be reviewed visually to ensure no artefacts are present. A sanity check could be designed or a simple bulk assignment performed. Use of a secondary independent sCT generation algorithm (e.g. a segmentation based or atlas based method) might further improve the level of quality of the sCT. If available, CBCT can be used as a QA tool of an MRI-based synthetic CT that was used for planning [116]. Differences between the dose calculation based on CBCT and sCT can reveal prediction difficulties. Local uncertainty regions can also be detected by uncertainty maps generated as the second output of an AI sCT generation algorithm, though these approaches are still under investigation [31,117,118]. If possible, it is recommended to compare all newly incoming patient data to the data used for training using similarity metrics. This could help to establish a level of confidence based on a correspondence between data and resulting quality of the sCT. Subsequently, this could serve as a method to flag potentially failing sCTs.

Routine model QA

Changes in workflow (e.g. changing MRI or CBCT device or imaging protocol) may require a new test set acquired with this new workflow. The sCT of the patients in this new test set should be generated and compared to the ground truth CT for this patient and reviewed by the clinical team involved. Recommissioning of the model will be required if the model output is no longer satisfactory. If the sCT generation model is based on MRI input, a regular quality check of the MRI geometric fidelity should be part of the QA program. The electron density calibration from the training CT data should be used. To obtain a new dataset of (paired) images after a significant protocol change, means that additional imaging has to be performed for a group of (new) patients; this might become a large hurdle to change and improve imaging protocols. In this scenario, having an MRI-only workflow can prohibit the introduction of new imaging possibilities. A temporarily extended

MRI protocol could be designed to have the commissioned sequences in the imaging session, but this may not always be possible due to increased exam time or ethical guidelines.

Treatment and machine QA

QA supervises the radiotherapy workflow to assess the reliability of treatment delivery by systematically monitoring the patient safety [119]. Efforts are made to provide necessary checks, secondary independent measurements, and evaluations against practice standards [81]. Different national and international institutions have contributed to efficient and safe quality control protocols [24,32,120–123]. These guidelines often recommend an extensive list of measurements on treatment units with limited time and personnel available. To assist with these procedures, AI-based QA has already shown to achieve similar or better performance than standard QA procedures [20]. Ultimately, these efforts might guide QA to a 'measureless' framework where verifications and measurement opportunities move beyond the scope of human attention [20]. Nevertheless, the increasing usage of AI models for automation of QA creates awareness for the need of QA procedures to assure the safety of these AI processes [81]. Because commissioning and implementation depends on the specific architecture and solution chosen, no specific recommendations are made in this section but more a general description of commonly used methods is described.

Patient-specific treatment verification

The most commonly used metric to assess agreement between two dose distributions is the gamma index, which combines both dose difference and distance-to-agreement in a single metric [124]. Treatment planning and delivery techniques are subject to a wide range of uncertainties that may contribute to decreased gamma pass rates (GPR), such as phantom/patient setup, detector resolution and calibration, beam output and profile, beam modeling, and especially plan complexity [125]. Predictive algorithms could be conducive in reducing the time required to measure plans that are at low risk of failure, and perhaps channeling resources into producing a better plan for difficult cases. ML models based on hand-crafted features like plan complexity metrics (PQM) and/or machine parameters have been demonstrated to predict GPR with high accuracy [125–129]. In addition, CNN approaches based on fluence maps can achieve similar prediction capabilities as ML methods [130–134]. These tools can be used as feedback into the treatment planning process. For instance, if the prediction shows the plan is unlikely to pass QA, the dosimetrist or physicist may choose to reduce the plan complexity in the optimization process. In this way, failing plans could be potentially eliminated and a possible treatment delay can be avoided [125].

Machine QA

QA of linear accelerators is periodically performed to monitor longitudinal stability [135]. Measured data contain nonlinearity in a multidimensional space, making it difficult to interpret [136]. Due to developments seen in the delivery and monitoring systems, opportunities arise to complement with approaches such as Probabilistic Safety Assessment (PSA) [137] or risk analysis [21] to focus where AI can amplify detection levels and prediction accuracy of potential failure or deviation from intent. Either through machine internal sensors and logs (measuring speeds, positions, rates, etc.) or external devices (measuring dose or surrogates, positions, etc.), AI has the potential to foresee stray behaviors with high selectivity allowing efficient triage for problem solving as well as pre-emptive actions. This will improve machine uptime, reliability and congruence between planned and delivered treatment [138]. One source of investigation includes the use of AI models to predict deviations in multileaf collimators (MLCs) positions to perform maintenance accordingly [139,140]. AI can be used in an ART workflow to flag when a plan is likely to fail a QA due to overmodulation or violation of delivery constraints [141]. Lastly, time-related monitoring of the beam quality has been investigated to better identify tolerance boundaries based on time-related data collection [142]. Ultimately, this should lead to a better understanding of the underlying function and relationship between measurements and help to take preventive actions [135].

Discussion

This paper provides an overview of recommendations to help implementation of AI software in clinical routine. A summary of the manuscript can be found in Table 4. The paper focuses on commonly used specific RT applications [15]. Applications such as computer-aided diagnosis and detection, image registration, image reconstruction, outcome prediction, etc., were not discussed but AI is also appearing in these parts of the radiotherapy workflow, see e.g. [6,7,9,10,12,143–145]. For future perspectives of AI in RT, we refer the reader to [11,146,147]. This paper was built upon the experience with the integration of AI software in the RT workflow and should as such be interpreted as a consensus of radiotherapy centers represented at the 3rd ESTRO Physics Workshop. However, it is possible that centers or users might have already specific protocols in place to safely introduce AI software in their clinic. The recommendations outlined here are a good starting point for clinics starting introduction of AI. As AI is a rapidly evolving field where models, applications and training methods are finding their way in radiotherapy, it is evident that such recommendations need to be regularly updated as well.

A concern raised by automation is the possible disappearance of domain knowledge among physicians, physicists or RTTs. In part, skills are dependent on experience and automation reduces the possibility of gaining experience in creating manual plans or segmentations in the clinical workflow. Manually performing these tasks on a regular basis helps to preserve domain knowledge. However, we also expect a shift from being active in the patient's individual care path to "offline" preparation of models which requires similar experience. Since clinical workflows, fractionation schemes, devices, etc. change over time, models will also have to change over time. New data will be required to update models to new clinical practices and specific knowledge will still be needed to create this new data. Furthermore, we expect a shift in the knowledge and experience from pure treatment planning (workflow) to an understanding of the working principle of the models and interpretation of the output of the models. Medical physicists involved in AI should familiarize themselves with all relevant aspects, and (future) curricula of the Qualified Medical Physics expert and radiation oncologists should incorporate big data and artificial intelligence [148].

Because regulation and quality labels are still being developed for AI, these topics were not included in this paper. But as the EU white paper on AI frames it appropriately: "the use of AI brings both opportunities and risks." [149]. Users of AI solutions are strongly recommended to become familiar with the recently published Medical Device Regulation (MDR) [150] as it will influence the clinical application and liabilities related to in-house created models or software tools. The 2013/59/EURATOM directive [151] also sets the requirement to perform risk analysis audits on the clinical management plans, which also includes any clinical software/algorithmic solution such as AI. Data sharing agreements between clinics would be beneficial and could facilitate the use

Table 4
A summary of principal recommendations to guide the implementation and use of AI models in the clinical workflow.

	Automated segmentation	Automated treatment planning	Synthetic CT
Objective(s)	Increase accuracy/consistency Time saving Decrease inter- and intra-observer variability		Using MRI or CBCT for dose calculation (and patient positioning)
Commissioning	*training/validation phase:* tune model to clinical needs *test phase:* independent evaluation (accuracy/robustness)		
Data characterization	• Large amount of high quality annotated, retrospective data [2] • Variability of the clinical practice should be represented Consistency in treatment site and/or treatment technique		Consistency in treatment site, image acquisition protocol and image device
Model analysis	*Quantitative:* calculating similarity to clinical ground truths Overlap, distance and dose measures (cf. Table 1)	Dosimetric measures (cf. Table 2)	Measures to evaluate image similarity, geometric fidelity and dosimetric accuracy (cf. Table 3)
	Qualitative: present output to clinical experts for revision		
Implementation	• Multidisciplinary team with relevant expertise and AI knowledge needed • End-to-end test • Feedback system • Risk analysis (recommended)		
Quality assurance	To monitor the accuracy/consistency of model output over time and robustness to adapted workflows, i.e. continuous quality improvement [28]		
Case-specific QA	• Manual verification • Involvement of all users/education • Statistical or AI models to identify outliers • Independent, secondary automated algorithm or plausibility check		
Routine model QA	• Quality Management Program (QMP) with reference dataset [32] • Informatics platform/log file to collect suboptimal cases, i.e. post-market surveillance [33]		

of AI models. However, setting up agreements for exchanging data, scripts and models between different centers will become even more complicated. The General Data Protection Regulation (GDPR) [150] also sets the framework for privacy compliance which is recommended to use as a benchmark for practices and future AI implementation. Data anonymization and informed consent of patients are the two ways to be able to create and use curated databases in the context of AI training, validation, testing and clinical usage. It is fair to state that AI pushes our perception of privacy to the extreme. Finding the proper balance between privacy protection of individuals and progress in research to help improve individualized health care is an ongoing discussion. Evidently, AI is based on Big Data, collaboration and sharing, all of which are regulated by both MDR and GDPR. Regardless of regulation issues, users should be aware of accountability, safety issues and robustness when developing AI tools in a clinical setting. There are strong arguments in favor of providing some kind of quality label or uncertainty assessment, risk analysis and sanity checks when introducing models that have been created in-house or by other groups into clinical carepaths.

AI models and especially data driven models such as ML and DL are advancing in the clinical radiotherapy workflow. Since many of the current AI models appear as black boxes to the users, the commissioning, implementation and QA procedures are essential. In this article, we aimed to provide clinical recommendations to support clinical teams during implementation of AI models in the radiotherapy workflow for contouring, planning and synthetic CT. Recommendations were not provided for the implementation of automated QA models since this is still in its infancy. However, the general principles are transferable to other applications.

Conflicts of interest

Michaël Claessens and Dirk Verellen have an on-going research collaboration with RaySearch Laboratories, SunNuclear Corporation and Sordina IORT Technologies.

Acknowledgements

We would like to thank Elizabeth Adams, Sigrun Saur Almberg, Jeremy Booth, Andrew Green, Diana Grishchuk, Matteo Maspero, David Menichelli and Carlos Rodrigues for their thoughtful review on this paper. Furthermore, this research is funded through the AI Research Program of the Flemish Government, Belgium (Liesbeth Vandewinckele) and by a grant of the Flemish League Against Cancer, Belgium, ref: 000019356 (Michaël Claessens).

References

[1] Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. Br J Radiol 2019;92:1–12. https://doi.org/10.1259/bjr.20190001.
[2] El Naqa I, Ruan D, Valdes G, Dekker A, McNutt T, Ge Y, et al. Machine learning and modeling: Data, validation, communication challenges. Med Phys 2018;45:e834–40. https://doi.org/10.1002/mp.12811.
[3] Feng M, Valdes G, Dixit N, Solberg TD. Machine learning in radiation oncology: Opportunities, requirements, and needs. Front Oncol 2018;8:1–7. https://doi.org/10.3389/fonc.2018.00110.
[4] Wang P. On Defining Artificial Intelligence. J Artif Gen Intell 2019;10:1–37. https://doi.org/10.2478/jagi-2019-0002.
[5] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44. https://doi.org/10.1038/nature14539.
[6] Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. Comput Biol Med 2018;98:126–46. https://doi.org/10.1016/j.compbiomed.2018.05.018.
[7] Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. Med Phys 2019;46:e1–e36. https://doi.org/10.1002/mp.13264.
[8] Shen C, Nguyen D, Zhou Z, Jiang SB, Dong B, Jia X. An introduction to deep learning in medical physics: Advantages, potential, and challenges. Phys Med Biol 2020:65. https://doi.org/10.1088/1361-6560/ab6f51.
[9] Boldrini L, Bibault JE, Masciocchi C, Shen Y, Bittner MI. Deep learning: A review for the radiation oncologist. Front Oncol 2019:9. https://doi.org/10.3389/fonc.2019.00977.
[10] Rattan R, Kataria T, Banerjee S, Goyal S, Gupta D, Bisht S, et al. artificial intelligence in oncology, its scope and future prospects with specific reference to radiation oncology 1. BJR Open 2019;1:20180031. https://doi.org/10. 1259/ bjro. 20180031.
[11] Bibault JE, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: State of the art and future prospects. Cancer Lett 2016;382:110–7. https://doi.org/10.1016/j.canlet.2016.05.033.

[12] Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, et al. Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation?. Radiother Oncol 2018;129:421–6. https://doi.org/10.1016/j.radonc.2018.05.030.

[13] Xing L, Krupinski EA, Cai J. Artificial intelligence will soon change the landscape of medical physics research and practice. Med Phys 2018;45:1791–3. https://doi.org/10.1002/mp.12831.

[14] Tang X, Wang B, Rong Y. Artificial intelligence will reduce the need for clinical medical physicists. J Appl Clin Med Phys 2018;19:6–9. https://doi.org/10.1002/acm2.12244.

[15] Brouwer CL, Dinkla AM, Vandewinckele L, Crijns W, Claessens M, Verellen D, et al. Artificial Intelligence in Radiation Oncology: Current use and needs to support clinical implementation. PhiRO (in Submission) 2020.

[16] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med 2015;162:55–63. https://doi.org/10.7326/M14-0697.

[17] Gennatas ED, Friedman JH, Ungar LH, Pirracchio R, Eaton E, Reichmann LG, et al. Expert-augmented machine learning. Proc Natl Acad Sci U S A 2020;117:4571–7. https://doi.org/10.1073/pnas.1906831117.

[18] US Food and Drug Administration. Artificial Intelligence and Machine Learning in Software as a Medical Device 2019:1–20.

[19] Halvorsen PH, Cirino E, Das IJ, Garrett JA, Yang J, Yin FF, et al. AAPM-RSS Medical Physics Practice Guideline 9.a. for SRS-SBRT. J Appl Clin Med Phys 2017;18:10–21. https://doi.org/10.1002/acm2.12146.

[20] Kalet AM, Luk SMH, Phillips MH. Radiation therapy quality assurance tasks and tools: the many roles of machine learning. Med Phys 2020;47:e168–77. https://doi.org/10.1002/mp.13445.

[21] EU - Directorate General for Energy. General guidelines on risk management in external beam radiotherapy. 2015.

[22] Huq MS, Fraass BA, Dunscombe PB, Gibbons JP, Ibbott GS, Medin PM, et al. A method for evaluating quality assurance needs in radiation therapy. Int J Radiat Oncol Biol Phys 2008;71:170–3. https://doi.org/10.1016/j.ijrobp.2007.06.081.

[23] Scorsetti M, Signori C, Lattuada P, Urso G, Bignardi M, Navarria P, et al. Applying failure mode effects and criticality analysis in radiotherapy: Lessons learned and perspectives of enhancement. Radiother Oncol 2010;94:367–74. https://doi.org/10.1016/j.radonc.2009.12.040.

[24] Huq MS, Fraass BA, Dunscombe PB, Gibbons JP, Ibbott GS, Mundt AJ, et al. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management. Med Phys Med Phys Med Phys 2016;43:4078–13874. https://doi.org/10.1118/1.2349696.

[25] IAEA. Application of the risk matrix method to radiotherapy. IAEA 2016:2016–21. https://doi.org/10.13140/2.1.2586.1122.

[26] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

[27] Jia X, Ren L, Cai J. Clinical implementation of AI technologies will require interpretable AI models. Med Phys 2020;47:1–4. https://doi.org/10.1002/mp.13891.

[28] Berwick DM. Continuous Improvement as an Ideal in Health Care. N Engl J Med 1989;320:53–6. https://doi.org/10.1056/NEJM198901053200110.

[29] Chan JW, Kearney V, Haaf S, Wu S, Bogdanov M, Reddick M, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. Med Phys 2019;46:2204–13. https://doi.org/10.1002/mp.13495.

[30] Mehrtash A, Wells WM, Tempany CM, Abolmaesumi P, Kapur T. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. ArXiv Prepr ArXiv191113273 2019.

[31] Bragman FJS, Tanno R, Wenqi ZE, David L, Ourselin S, Alexander DC, et al. Quality control in radiotherapy-treatment planning using multi-task learning and uncertainty estimation 2018.

[32] Smilowitz JB, Das IJ, Feygelman V, Fraass BA, Kry SF, Marshall IR, et al. AAPM Medical Physics Practice Guideline 5.a.: Commissioning and QA of treatment planning dose calculations - megavoltage photon and electron beams. J Appl Clin Med Phys 2015;16:14–34.

[33] Maisel WH. Medical device regulation: an introduction for the practicing physician. Ann Intern Med 2004;140:296–302. https://doi.org/10.7326/0003-4819-140-4-200402170-00012.

[34] Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32. https://doi.org/10.1186/1748-717X-7-32.

[35] van der Veen J, Gulyban A, Nuyts S. Interobserver variability in delineation of target volumes in head and neck cancer. Radiother Oncol 2019;137:9–15. https://doi.org/10.1016/j.radonc.2019.04.006.

[36] van der Veen J, Willems S, Deschuymer S, Robben D, Crijns W, Maes F, et al. Benefits of deep learning for delineation of organs at risk in head and neck cancer. Radiother Oncol 2019;138:68–74. https://doi.org/10.1016/j.radonc.2019.05.010.

[37] Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol 2018;126:312–7. https://doi.org/10.1016/j.radonc.2017.11.012.

[38] Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Med Phys 2017;44:547–57. https://doi.org/10.1002/mp.12045.

[39] Men K, Chen X, Zhang Y, Zhang T, Dai J, Yi J, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. Front Oncol 2017;7:1–9. https://doi.org/10.3389/fonc.2017.00315.

[40] Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. Med Phys 2017;44:6377–89. https://doi.org/10.1002/mp.12602.

[41] Savenije MHF, Maspero M, Sikkes GG, Van Der Voort Van Zyp JRN, Alexis AN, Bol GH, et al. Clinical implementation of MRI-based organs-at-risk auto-segmentation with convolutional networks for prostate radiotherapy. Radiat Oncol 2020;15:1–12. https://doi.org/10.1186/s13014-020-01528-0.

[42] van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkers Roel JHM, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. Radiother Oncol 2020;142:115–23. https://doi.org/10.1016/j.radonc.2019.09.022.

[43] Willems S, Crijns W, La Greca Saint-Esteven A, Van Der Veen J, Robben D, Depuydt T, et al. Clinical Implementation of DeepVoxNet for Auto-Delineation of Organs at Risk in Head and Neck Cancer Patients in Radiotherapy, 2018, p. 223–32. https://doi.org/10.1007/978-3-030-01201-4_24.

[44] Valentini V, Boldrini L, Damiani A, Muren LP. Recommendations on how to establish evidence from auto-segmentation software in radiotherapy. Radiother Oncol 2014;112:317–20. https://doi.org/10.1016/j.radonc.2014.09.014.

[45] Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. Semin Radiat Oncol 2019. https://doi.org/10.1016/j.semradonc.2019.02.001.

[46] Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132: Report. Med Phys 2017;44:e43–76. https://doi.org/10.1002/mp.12256.

[47] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Med Imaging 2015:15. https://doi.org/10.1186/s12880-015-0068-x.

[48] van Rooij W, Dahele M, Ribeiro Brandao H, Delaney AR, Slotman BJ, Verbakel WF. Deep learning-based delineation of head and neck organs at risk: geometric and dosimetric evaluation. Int J Radiat Oncol Biol Phys 2019;104:677–84. https://doi.org/10.1016/j.ijrobp.2019.02.040.

[49] Gooding MJ, Smith AJ, Tariq M, Aljabar P, Peressutti D, van der Stoep J, et al. Comparative evaluation of autocontouring in clinical practice: A practical method using the Turing test. Med Phys 2018;45:5105–15. https://doi.org/10.1002/mp.13200.

[50] Nikolov S, Blackwell S, Mendes R, De Fauw J, Meyer C, Hughes C, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. ArXiv Prepr ArXiv180904430 2018.

[51] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302. https://doi.org/10.2307/1932409.

[52] Vaassen F, Hazelaar C, Vaniqui A, Gooding M, van der Heyden B, Canters R, et al. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. Phys Imaging Radiat Oncol 2020;13:1–6. https://doi.org/10.1016/j.phro.2019.12.001.

[53] Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. Semin Radiat Oncol 2019;29:185–97. https://doi.org/10.1016/j.semradonc.2019.02.001.

[54] Chen HC, Tan J, Dolly S, Kavanaugh J, Anastasio MA, Low DA, et al. Automated contouring error detection based on supervised geometric attribute distribution models for radiation therapy: a general strategy. Med Phys 2015;42:1048–59. https://doi.org/10.1118/1.4906197.

[55] Court LE, Kisling K, Mccarroll R, Zhang L, Yang J, Simonds H, et al. Radiation planning assistant – A streamlined , fully automated radiotherapy treatment planning system 2020:1–9. https://doi.org/10.3791/57411.

[56] McIntosh C, Svistoun I, Purdie TG. Groupwise conditional random forests for automatic shape classification and contour quality assessment in radiotherapy planning. IEEE Trans Med Imaging 2013;32:1043–57. https://doi.org/10.1109/TMI.2013.2251421.

[57] Craft DL, Hong TS, Shih HA, Bortfeld TR. Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys 2012;82:83–90. https://doi.org/10.1016/j.ijrobp.2010.12.007.

[58] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. Pract Radiat Oncol 2012;2:296–305. https://doi.org/10.1016/j.prro.2011.11.012.

[59] Chang ATY, Hung AWM, Cheung FWK, Lee MCH, Chan OSH, Philips H, et al. Comparison of planning quality and efficiency between conventional and knowledge-based algorithms in nasopharyngeal cancer patients using intensity modulated radiation therapy. Int J Radiat Oncol Biol Phys 2016;95:981–90. https://doi.org/10.1016/j.ijrobp.2016.02.017.

[60] Leung LHT, Kan MWK, Cheng ACK, Wong WKH, Yau CC. A new dose-volume-based plan quality index for IMRT plan comparison. Radiother Oncol 2007;85:407–17. https://doi.org/10.1016/j.radonc.2007.10.018.

[61] Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning-a review of recent innovations. Br J Radiol 2018:91. https://doi.org/10.1259/bjr.20180270.
[62] Wang C, Zhu X, Hong JC, Zheng D. Artificial intelligence in radiotherapy treatment planning: present and future. Technol Cancer Res Treat 2019;18:1–11. https://doi.org/10.1177/1533033819873922.
[63] Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL. Predicting dose-volume histograms for organs-at-risk in IMRT planning. Med Phys 2012;39:7446–61. https://doi.org/10.1118/1.4761864.
[64] Yuan L, Ge Y, Lee WR, Yin FF, Kirkpatrick JP, Wu QJ. Quantitative analysis of the factors which affect the interpatient organ-At-risk dose sparing variation in IMRT plans. Med Phys 2012;39:6868–78. https://doi.org/10.1118/1.4757927.
[65] McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. Phys Med Biol 2017;62:5926–44. https://doi.org/10.1088/1361-6560/aa71f8.
[66] Nguyen D, Jia X, Sher D, Lin MH, Iqbal Z, Liu H, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. Phys Med Biol 2019:64. https://doi.org/10.1088/1361-6560/ab039b.
[67] Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Med Phys 2019;46:370–81. https://doi.org/10.1002/mp.13271.
[68] Kontaxis C, Bol GH, Lagendijk JJW, DeepDose RBW. Towards a fast dose calculation engine for radiation therapy using deep learning. Phys Med Biol 2020:65. https://doi.org/10.1088/1361-6560/ab7630.
[69] Mahmood R, Babier A, McNiven A, Diamant A, Chan TCY. Automated treatment planning in radiation therapy using generative adversarial networks. ArXiv Prepr ArXiv180706489 2018.
[70] Lee H, Kim H, Kwak J, Kim YS, Lee SW, Cho S, et al. Fluence-map generation for prostate intensity-modulated radiotherapy planning using a deep-neural-network. Sci Rep 2019;9:1–11. https://doi.org/10.1038/s41598-019-52262-x.
[71] Nguyen D, Barkousaraie AS, Shen C, Jia X, Jiang S. Generating Pareto optimal dose distributions for radiation therapy treatment planning. In: Int. Conf. Med. Image Comput. Comput. Interv.. Springer International Publishing; 2019. p. 59–67.
[72] Sadeghnejad Barkousaraie A, Ogunmolu O, Jiang S, Nguyen D. A fast deep learning approach for beam orientation optimization for prostate cancer treated with intensity-modulated radiation therapy. Med Phys 2020;47:880–97. https://doi.org/10.1002/mp.13986.
[73] Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, et al. On the pre-clinical validation of a commercial model-based optimisation engine: Application to volumetric modulated arc therapy for patients with lung or prostate cancer. Radiother Oncol 2014;113:385–91. https://doi.org/10.1016/j.radonc.2014.11.009.
[74] Tol JP, Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Evaluation of a knowledge-based planning solution for head and neck cancer. Int J Radiat Oncol Biol Phys 2015;91:612–20. https://doi.org/10.1016/j.ijrobp.2014.11.014.
[75] Fogliata A, Wang PM, Belosi F, Clivio A, Nicolini G, Vanetti E, et al. Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. Radiat Oncol 2014;9:236. https://doi.org/10.1186/s13014-014-0236-0.
[76] Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z, et al. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. Sci Rep 2019;9:1–10. https://doi.org/10.1038/s41598-018-37741-x.
[77] Krayenbuehl J, Zamburlini M, Ghandour S, Pachoud M, Tanadini-Lang S, Tol J, et al. Correction to: Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer. Radiat Oncol 2018;13:170. https://doi.org/10.1186/s13014-018-1113-z.
[78] Ge Y, Wu QJ. Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches. Med Phys 2019. https://doi.org/10.1002/mp.13526.
[79] ICRU Report 83 Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). J ICRU 1999;62.
[80] Moran JM, Radawski J, Fraass BA. A dose gradient analysis tool for IMRT QA. J Appl Clin Med Phys 2005;6:62–73. https://doi.org/10.1120/jacmp.v6i2.2006.
[81] McNutt TR, Moore KL, Wu B, Wright JL. Use of big data for quality assurance in radiation therapy. Semin Radiat Oncol 2019;29:326–32. https://doi.org/10.1016/j.semradonc.2019.05.006.
[82] Janssen TM, Kusters M, Wang Y, Wortel G, Monshouwer R, Damen E, et al. Independent knowledge-based treatment planning QA to audit Pinnacle autoplanning. Radiother Oncol 2019;133:198–204. https://doi.org/10.1016/j.radonc.2018.10.035.
[83] Tol JP, Dahele M, Delaney AR, Slotman BJ, Verbakel WFAR. Can knowledge-based DVH predictions be used for automated, individualized quality assurance of radiotherapy treatment plans?. Radiat Oncol 2015;10:234. https://doi.org/10.1186/s13014-015-0542-1.
[84] Petit SF, van Elmpt W. Accurate prediction of target dose-escalation and organ-at-risk dose levels for non-small cell lung cancer patients. Radiother Oncol 2015;117:453–8. https://doi.org/10.1016/j.radonc.2015.07.040.
[85] Wang Y, Zolnay A, Incrocci L, Joosten H, McNutt T, Heijmen B, et al. A quality control model that uses PTV-rectal distances to predict the lowest achievable rectum dose, improves IMRT planning for patients with prostate cancer. Radiother Oncol 2013;107:352–7. https://doi.org/10.1016/j.radonc.2013.05.032.
[86] Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: Benefits and challenges. Phys Med Biol 2018:63. https://doi.org/10.1088/1361-6560/aaaca4.
[87] Pereira GC, Traughber M, Muzic RF. The role of imaging in radiation therapy planning: Past, present, and future. Biomed Res Int 2014:2014. https://doi.org/10.1155/2014/231090.
[88] Nyholm T, Jonsson J. Counterpoint: opportunities and challenges of a magnetic resonance imaging-only radiotherapy work flow. Semin Radiat Oncol 2014;24:175–80. https://doi.org/10.1016/j.semradonc.2014.02.005.
[89] Veiga C, McClelland J, Moinuddin S, Lourenço A, Ricketts K, Annkah J, et al. Toward adaptive radiotherapy for head and neck patients: feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. Med Phys 2014;41:1–12. https://doi.org/10.1118/1.4864240.
[90] Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. Med Phys 2020;47:1115–25. https://doi.org/10.1002/mp.13978.
[91] Guerreiro F, Burgos N, Dunlop A, Wong K, Petkar I, Nutting C, et al. Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning. Phys Medica 2017;35:7–17. https://doi.org/10.1016/j.ejmp.2017.02.017.
[92] Farjam R, Tyagi N, Veeraraghavan H, Apte A, Zakian K, Hunt MA, et al. Multiatlas approach with local registration goodness weighting for MRI-based electron density mapping of head and neck anatomy. Med Phys 2017;44:3706–17. https://doi.org/10.1002/mp.12303.
[93] Hsu SH, Cao Y, Huang K, Feng M, Balter JM. Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy. Phys Med Biol 2013;58:8419–35. https://doi.org/10.1088/0031-9155/58/23/8419.
[94] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44:1408–19. https://doi.org/10.1002/mp.12155.
[95] Dinkla AM, Wolterink JM, Maspero M, Savenije MHF, Verhoeff JJC, Seravalli E, et al. MR-only brain radiation therapy: dosimetric evaluation of synthetic cts generated by a dilated convolutional neural network. Int J Radiat Oncol Biol Phys 2018;102:801–12. https://doi.org/10.1016/j.ijrobp.2018.05.058.
[96] Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy 2018:1–17. https://doi.org/10.1088/1361-6560/aada6d.
[97] Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJHM, et al. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol 2020:65. https://doi.org/10.1088/1361-6560/ab7d54.
[98] Kurz C, Maspero M, Savenije MHF, Landry G, Kamp F, Pinto M, et al. CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. Phys Med Biol 2019:64. https://doi.org/10.1088/1361-6560/ab4d8c.
[99] Liu F, Jang H, Kijowski R, Bradshaw T, Mcmillan AB. Deep learning Mr imaging-based attenuation correction for PeT/Mr imaging 1 TECHNICAL DEVELOPMENTS: Deep Learning MR Imaging-based Attenuation Correction for PET/MR Imaging Liu et al. Radiol n Radiol 2018;286. https://doi.org/10.1148/radiol.2017170700.
[100] Maspero M, Houweling AC, Savenije MHF, van Heijst TCF, Verhoeff JJC, Kotte ANTJ, et al. A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer. Phys Imaging Radiat Oncol 2020;14:24–31. https://doi.org/10.1016/j.phro.2020.04.002.
[101] Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Išgum I. Deep MR to CT synthesis using unpaired data BT - simulation and synthesis in medical imaging. In: Tsaftaris SA, Gooya A, Frangi AF, Prince JL, editors. Int. Work. Simul. Synth. Med. imaging, Cham: Springer International Publishing; 2017, p. 14–23.
[102] Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, et al. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. Phys Med Biol 2019:64. https://doi.org/10.1088/1361-6560/ab22f9.
[103] Dinkla AM, Florkow MC, Maspero M, Savenije MHF, Zijlstra F, Doornaert PAH, et al. Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network. Med Phys 2019;46:4095–104. https://doi.org/10.1002/mp.13663.
[104] Lei Y, Harms J, Wang T, Liu Y, Shu HK, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med Phys 2019;46:3565–81. https://doi.org/10.1002/mp.13617.
[105] Olberg S, Zhang H, Kennedy WR, Chun J, Rodriguez V, Zoberi I, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. Med Phys 2019;46:4135–47. https://doi.org/10.1002/mp.13716.
[106] Liu Y, Lei Y, Wang Y, Shafai-Erfani G, Wang T, Tian S, et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. Phys Med Biol 2019:64. https://doi.org/10.1088/1361-6560/ab41af.
[107] Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. Radiother Oncol 2019;136:56–63. https://doi.org/10.1016/j.radonc.2019.03.026.

[108] Liu Y, Lei Y, Wang Y, Wang T, Ren L, Lin L, et al. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method. Phys Med Biol 2019;64:13. https://doi.org/10.1088/1361-6560/ab25bc.
[109] Spadea MF, Pileggi G, Zaffino P, Salome P, Catana C, Izquierdo-Garcia D, et al. Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images—application in brain proton therapy. Int J Radiat Oncol Biol Phys 2019;105:495–503. https://doi.org/10.1016/j.ijrobp.2019.06.2535.
[110] Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. Radiat Oncol 2017;12:1–15. https://doi.org/10.1186/s13014-016-0747-y.
[111] Li W, Li Y, Qin W, Liang X, Xu J, Xiong J, et al. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. Quant Imaging Med Surg 2020;10:1223–36. https://doi.org/10.21037/qims-19-885.
[112] Qi M, Li Y, Wu A, Jia Q, Li B, Sun W, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. Med Phys 2020;47:1880–94. https://doi.org/10.1002/mp.14075.
[113] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. IEEE Trans Image Process 2004;13:600–12. https://doi.org/10.1109/TIP.2003.819861.
[114] Maspero M, Tyyger MD, Tijssen RHN, Seevinck PR, Intven MPW, van den Berg CAT. Feasibility of magnetic resonance imaging-only rectum radiotherapy with a commercial synthetic computed tomography generation solution. Phys Imaging Radiat Oncol 2018;7:58–64. https://doi.org/10.1016/j.phro.2018.09.002.
[115] Handrack J, Bangert M, Möhler C, Bostel T, Greilich S. Towards a generalised development of synthetic CT images and assessment of their dosimetric accuracy. Acta Oncol (Madr) 2020;59:180–7. https://doi.org/10.1080/0284186X.2019.1684558.
[116] Palmér E, Persson E, Ambolt P, Gustafsson C, Gunnlaugsson A, Olsson LE. Cone beam CT for QA of synthetic CT in MRI only for prostate patients. J Appl Clin Med Phys 2018;19:44–52. https://doi.org/10.1002/acm2.12429.
[117] Bragman FJS, Tanno R, Eaton-Rosen Z, Li W, Hawkes DJ, Ourselin S, et al. Uncertainty in Multitask Learning: Joint Representations for Probabilistic MR-only Radiotherapy Planning. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2018; 11073 LNCS:3–11. https://doi.org/10.1007/978-3-030-00937-3_1.
[118] Tanno R, Worrall D, Kaden E, Ghosh A, Grussu F, Bizzi A, et al. Uncertainty quantification in deep learning for safer neuroimage enhancement. ArXiv Prepr ArXiv190713418 2019.
[119] Mehta SR, Suhag V, Semwal M, Sharma N. Radiotherapy: Basic concepts and recent advances. Med J Armed Forces India 2010;66:158–62. https://doi.org/10.1016/S0377-1237(10)80132-7.
[120] American Association of Physicists in Medicine. AAPM's TG-46 Comprehensive QA for radiation oncology. Med Phys 1994;21:581–618.
[121] Ezzell G a., Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. Med Phys 2009;36:5359–73. https://doi.org/10.1118/1.3238104.
[122] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys 2018;45:e53–83. https://doi.org/10.1002/mp.12810.
[123] Palmans H, Andreo P, Huq MS, Seuntjens J, Christaki KE, Meghzifene A. Dosimetry of small static fields used in external photon beam radiotherapy: Summary of TRS-483, the IAEA–AAPM international Code of Practice for reference and relative dose determination. Med Phys 2018;45:e1123–45. https://doi.org/10.1002/mp.13208.
[124] Andreo P, Burns DT, Hohlfeld K, Huq MS, Kanai T, Laitano F, Smyth SV. Absorbed dose determination in external beam radiotheraphy: an international code of practice for dosimetry based on standards of absorbed dose to water. 2006;2006:183.
[125] Lam D, Zhang X, Li H, Deshan Y, Schott B, Zhao T, et al. Predicting gamma passing rates for portal dosimetry-based IMRT QA using machine learning. Med Phys 2019;46:4666–75. https://doi.org/10.1002/mp.13752.
[126] Valdes G, Scheuermann R, Hung CY, Olszanski A, Bellerive M, Solberg TD. A mathematical framework for virtual IMRT QA using machine learning. Med Phys 2016;43:4323–34. https://doi.org/10.1118/1.4953835.
[127] Valdes G, Chan MF, Lim SB, Scheuermann R, Deasy JO, Solberg TD. IMRT QA using machine learning: A multi-institutional validation. J Appl Clin Med Phys 2017;18:279–84. https://doi.org/10.1002/acm2.12161.
[128] Li J, Wang L, Zhang X, Liu L, Li J, Chan MF, et al. Machine learning for patient-specific quality assurance of VMAT: prediction and classification accuracy. Int J Radiat Oncol Biol Phys 2019;105:893–902. https://doi.org/10.1016/j.ijrobp.2019.07.049.
[129] Granville DA, Sutherland JG, Belec JG, La Russa DJ. Predicting VMAT patient-specific QA results using a support vector classifier trained on treatment plan characteristics and linac QC metrics. Phys Med Biol 2019:64. https://doi.org/10.1088/1361-6560/ab142e.
[130] Interian Y, Rideout V, Kearney VP, Gennatas E, Morin O, Cheung J, et al. Deep nets vs expert designed features in medical physics: an IMRT QA case study. Med Phys 2018;45:2672–80. https://doi.org/10.1002/mp.12890.
[131] Tomori S, Kadoya N, Takayama Y, Kajikawa T, Shima K, Narazaki K, et al. A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance. Med Phys 2018;45:4055–65. https://doi.org/10.1002/mp.13112.
[132] Mahdavi SR, Tavakol A, Sanei M, Molana SH, Arbabi F, Rostami A, et al. Use of artificial neural network for pretreatment verification of intensity modulation radiation therapy fields. Br J Radiol 2019;92:1V. https://doi.org/10.1259/bjr.20190355.
[133] Kimura Y, Kadoya N, Tomori S, Oku Y, Jingu K. Error detection using a convolutional neural network with dose difference maps in patient-specific quality assurance for volumetric modulated arc therapy. Phys Medica 2020;73:57–64. https://doi.org/10.1016/j.ejmp.2020.03.022.
[134] Nyflot MJ, Thammasorn P, Wootton LS, Ford EC, Chaovalitwongse WA. Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks. Med Phys 2019;46:456–64. https://doi.org/10.1002/mp.13338.
[135] Li Q, Chan MF. Predictive time-series modeling using artificial neural networks for Linac beam symmetry: an empirical study. Ann N Y Acad Sci 2017;1387:84–94. https://doi.org/10.1111/nyas.13215.
[136] El Naqa I, Irrer J, Ritter TA, DeMarco J, Al-Hallaq H, Booth J, et al. Machine learning for automated quality assurance in radiotherapy: A proof of principle using EPID data description. Med Phys 2019;46:1914–21. https://doi.org/10.1002/mp.13433.
[137] Juan José Vilaragut, Rubén Ferro Fernández, Manuel Rodriguez Martí, Pedro Ortiz López, María Luisa Ramírez, Arturo Pérez Mulas, Marta Barrientos Montero, Fernando Somoano, José Miguel Delgado Rodriguez, Susana B. Papadópulos, Pedro Paulo Pereira Jr, Ramón JA. Probabilistic safety assessment (PSA) of the radiotherapy treatment process with an Electron Linear Accelerator (LINAC) for medical uses 2008:1–9.
[138] Pillai M, Adapa K, Das SK, Mazur L, Dooley J, Marks LB, et al. Using artificial intelligence to improve the quality and safety of radiation therapy. J Am Coll Radiol 2019;16:1267–72. https://doi.org/10.1016/j.jacr.2019.06.001.
[139] Wu B, Zhang P, Tsirakis B, Kanchaveli D, LoSasso T. Utilizing historical MLC performance data from trajectory logs and service reports to establish a proactive maintenance model for minimizing treatment disruptions. Med Phys 2019;46:475–83. https://doi.org/10.1002/mp.13363.
[140] Carlson JNK, Park JM, Park SY, Park JI, Choi Y, Ye SJ. A machine learning approach to the accurate prediction of multi-leaf collimator positional errors. Phys Med Biol 2016;61:2514–31. https://doi.org/10.1088/0031-9155/61/6/2514.
[141] Botti A, Cagni E, Orlandi M, Sghedoni R, Lambertini D, Barani A, et al. EP-2114 predicting inaccuracy of overmodulated RapidArc plans using Machine Learning model. Radiother Oncol 2019;133:S1170–1. https://doi.org/10.1016/s0167-8140(19)32534-4.
[142] Bin CY, Farrokhkish M, Norrlinger B, Heaton R, Jaffray D, Islam M. An artificial neural network to model response of a radiotherapy beam monitoring system. Med Phys 2020;47:1983–94. https://doi.org/10.1002/mp.14033.
[143] Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. Mach Vis Appl 2020;31:1–18. https://doi.org/10.1007/s00138-020-01060-x.
[144] Wang G. A perspective on deep imaging. IEEE Access 2016;4:8914–24. https://doi.org/10.1109/ACCESS.2016.2624938.
[145] Wang G, Ye JC, Mueller K, Fessler JA. Image reconstruction is a new frontier of machine learning. IEEE Trans Med Imaging 2018;37:1289–96. https://doi.org/10.1109/TMI.2018.2833635.
[146] Moore KL. Automated radiotherapy treatment planning. Semin Radiat Oncol 2019;29:209–18. https://doi.org/10.1016/j.semradonc.2019.02.003.
[147] Mayo CS, Moran JM, Bosch W, Xiao Y, McNutt T, Popple R, et al. American Association of Physicists in Medicine Task Group 263: Standardizing nomenclatures in radiation oncology. vol. 100. 2018. https://doi.org/10.1016/j.ijrobp.2017.12.013.
[148] Kortesniemi M, Tsapaki V, Trianni A, Russo P, Maas A, Källman HE, et al. The European Federation of Organisations for Medical Physics (EFOMP) White Paper: Big data and deep learning in medical imaging and in relation to medical physics profession. Phys Medica 2018;56:90–3. https://doi.org/10.1016/j.ejmp.2018.11.005.
[149] European Commission. Shaping Europe's digital future: Commission presents strategies for data and Artificial Intelligence. Https://EcEuropaEu/Commission/Presscorner/Detail/En/Ip_20_273, 2020:1–4.
[150] Regulation (EU) 2017/ 745 of the European Parliament and of the Council - of 5 April 2017 - on medical devices, amending Directive 2001/ 83/ EC, Regulation (EC) No 178/ 2002 and Regulation (EC) No 1223/ 2009 and repealing C. n.d.
[151] European Parliament. Council Directive 2013/59/Euratom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom a. Off J Eur Commun L13 2014:1–73.