Physics and Imaging in Radiation Oncology 32 (2024) 100677

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original research article

Pro-active risk analysis of an in-house developed deep learning based autoplanning tool for breast Volumetric Modulated Arc Therapy

Liesbeth Vandewinckele [a],*, Chahrazad Benazzouz [b], Laurence Delombaerde [a,b], Laure Pape [b], Truus Reynders [b], Aline Van der Vorst [a,b], Dylan Callens [a,b], Jan Verstraete [b], Adinda Baeten [b], Caroline Weltens [a,b], Wouter Crijns [a,b]

[a] Department of Oncology, Laboratory of Experimental Radiotherapy, KU Leuven, Belgium
[b] Department of Radiation Oncology, UZ Leuven, Belgium

ARTICLE INFO

Keywords:
Risk analysis
FMEA
In-house tool
Automated breast planning
Deep learning

ABSTRACT

Background and Purpose: With the increasing amount of in-house created deep learning models in radiotherapy, it is important to know how to minimise the risks associated with the local clinical implementation prior to clinical use. The goal of this study is to give an example of how to identify the risks and find mitigation strategies to reduce these risks in an implemented workflow containing a deep learning based planning tool for breast Volumetric Modulated Arc Therapy.

Materials and Methods: The deep learning model ran on a private Google Cloud environment for adequate computational capacity and was integrated into a workflow that could be initiated within the clinical Treatment Planning System (TPS). A proactive Failure Mode and Effect Analysis (FMEA) was conducted by a multidisciplinary team, including physicians, physicists, dosimetrists, technologists, quality managers, and the research and development team. Failure modes categorised as 'Not acceptable' and 'Tolerable' on the risk matrix were further examined to find mitigation strategies.

Results: In total, 39 failure modes were defined for the total workflow, divided over four steps. Of these, 33 were deemed 'Acceptable', five 'Tolerable', and one 'Not acceptable'. Mitigation strategies, such as a case-specific Quality Assurance report, additional scripted checks and properties, a pop-up window, and time stamp analysis, reduced the failure modes to two 'Tolerable' and none in the 'Not acceptable' region.

Conclusions: The pro-active risk analysis revealed possible risks in the implemented workflow and led to the implementation of mitigation strategies that decreased the risk scores for safer clinical use.

1. Introduction

One step in the radiotherapy treatment planning workflow is the optimisation of the treatment device parameters to obtain an acceptable balance between the dose deposited in the Organs-At-Risk (OAR) and targets, which is an operator dependent and very time-consuming process [1]. During recent years, it has been shown that automated planning tools based on deep learning (DL) can reduce these disadvantages [2,3]. Models have been created for the prediction of 3D dose distributions [4,5], fluence maps [6,7] and treatment device parameters directly [2,8]. However, the use of these in-house created DL models in clinical routine comes with (technical) challenges. One of these challenges is creating a reliable workflow such that the models can be easily executed in the clinic with minimal associated risks for the patient. Performing a risk-analysis before clinical use is important to reveal these potential risks [9].

The goal of this work was to perform a pro-active risk analysis of a clinical implementation of an in-house developed DL based automated planning tool and to mitigate the risks associated with this workflow by proposing corrective actions that lower the risk scores.

2. Materials and methods

2.1. Deployment of the automated planning tool

Recently, our research group created a DL based planning tool for breast radiotherapy [2]. The planning tool [2] was created for right breast cancer patients that fell within our clinical breast Simultaneous

* Correspondence to: Laboratory of Experimental Radiotherapy, UZ Herestraat 49 - bus 7003, 3000 Leuven, Belgium.
E-mail address: liesbeth.vandewinckele@azdelta.be (L. Vandewinckele).

https://doi.org/10.1016/j.phro.2024.100677
Received 30 April 2024; Received in revised form 1 November 2024; Accepted 13 November 2024
Available online 22 November 2024

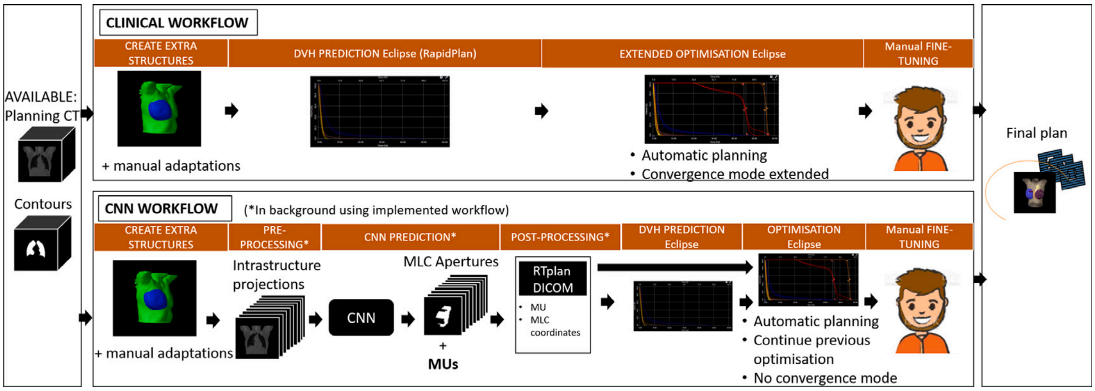


Fig. 1. Automated planning using the clinical and proposed CNN workflow. The clinical workflow refers to the workflow that is used in our clinic to obtain a breast VMAT plan. It is based on a DVH prediction, followed by an extended automatic optimisation. The CNN workflow refers to the workflow using the CNN based model that predicts MLC apertures and MU values to initialise a residual optimisation.

Integrated Boost (SIB) Volumetric Modulated Arc Therapy (VMAT) class solution, which was used for patients treated with adjuvant radiotherapy after breast conserving surgery. The dose prescriptions were 21 × 2.66 Gy (boost Planning Target Volume (PTV)) and 21 × 2.17 Gy (breast PTV). The Organs-At-Risk (OAR) consisted of the contralateral breast, heart, contralateral lung, ipsilateral lung and liver. The patient plans were created for a Halcyon^{TM} device (Varian, a Siemens Healthineers Company) and consisted of three partial arcs with each 116 control points (60 to 190 degrees and vice versa). All plans were created with RapidPlan^{TM} using an optimisation with extended convergence. A virtual bolus structure was used during planning to open the MLC leaves to account for breast swelling [10]. The created planning tool consisted of a Convolutional Neural Network (CNN) that created a relation between the anatomy of the patients and the corresponding Multileaf Collimator (MLC) apertures and Monitor Units (MU) per control point in the radiotherapy treatment plan. The created CNN was trained on 101 patient plans and validated and tested on 23 and 24 patient plans respectively [2].

For the purpose of this work, a workflow was suggested by our research and development team such that the automated planning tool could be easily used in clinical routine. The workflow is illustrated in Fig. 1. Once physicians finalised the contours, the dosimetrist began plan optimisation. First, extra planning structures were created using an automated tool in the MIM Maestro^{TM} software, further called the automated auxiliary structure tool. Then, the patient's planning CT and contours were pre-processed into projections created in the Beam's Eye View (BEV) of each control point. These projections were fed into the CNN model that predicted the MLC aperture and MU values for each control point. These MLC coordinates and MU values were used to create an RTplan in DICOM format, which was imported into the Treatment Planning System (TPS) to initialise further automated optimisation, guided by Dose Volume Histogram (DVH) predictions from RapidPlan^{TM} without extended convergence. The dosimetrist could fine-tune the plan to ensure all clinical dose constraints were met by manually adjusting optimisation objectives and starting a new optimisation iteration.

Different possible deployments exist to obtain this suggested CNN workflow. However, each deployment comes with its own associated risks, which makes it important to describe the details of the deployment of the CNN workflow before describing the performed risk analysis.

The CNN workflow was deployed as outlined in Fig. 2. Initially, the dosimetrist exported the CT, structure set and plan (only containing the isocentre position) in DICOM format from the clinical TPS, Eclipse^{TM}, to a virtual pc at UZ Leuven. New patients were added to a queue and processed sequentially. First, meta information of the patient was removed, but kept in memory, by extracting all image data from the DICOM object (pseudonymisation). This image data was then transferred to UZ Leuven's private Google Cloud environment. Pre-processing and CNN predictions were executed on Graphical Processing Units (GPU) on Google Cloud to have enough calculation capacity. The MLC aperture and MU predictions were sent back to the virtual pc for post-processing, where a DICOM plan was created using the temporarily stored meta information. The dosimetrist could then import this plan into the TPS for a residual optimisation.

A Graphical User Interface (GUI), developed using the Eclipse Scripting Application Programming Interface (ESAPI), facilitated the data export and import processes, notifying the user when the predicted plan was ready for import.

2.2. Risk analysis

For the purpose of this work, a Failure Mode and Effect Analysis (FMEA) was conducted to identify and evaluate potential risks (i.e. the failure modes) in the deployed CNN workflow [11–14]. This is the standard risk analysis in our department, aligning with TG-100 guidelines [14].

The risk analysis team comprised multidisciplinary experts, including three physicians specialised in breast radiotherapy, two physicists (one specialised in breast radiotherapy), a dosimetrist, a radiotherapy technologist, two quality managers, and two research and development team members.

The CNN workflow, new to the clinic, was unfamiliar to many team members. A movie/screen recording manual was created by the first author and shared to enhance understanding, leveraging the workflow's integration with the routinely used TPS.

The risk analysis process involved four parts. The first consisted of **defining the workflow's steps and substeps**. This was initially outlined by the first author, but could be further adapted by the other team members during the team meetings. Our team decided that the workflow involved four main steps, as detailed in Table 1, each comprising several substeps, performed once per patient. The second part in the FMEA process involved **identification of failure modes**. A preliminary list was provided by the first author, enhanced through team feedback during the team meetings facilitated by the movie/screen recording explaining the workflow. The team concluded that there were no differences between the failure modes in the clinical and CNN workflow in the 'Creation initial plan' step. During the third part, **assessment of risks**, each failure mode's occurrence, severity, and lack of detectability was assessed by all team members, see Table 2, in relation to the final output when the failure would not have been detected in time. The final output was the treatment of the patient and effects were defined in relation to this outcome, based on TG-100 [14]. The team focused on failures specific to the CNN workflow, excluding common issues with the clinical workflow. Scores were defined by taking into account that the workflow could be used for approximately 30

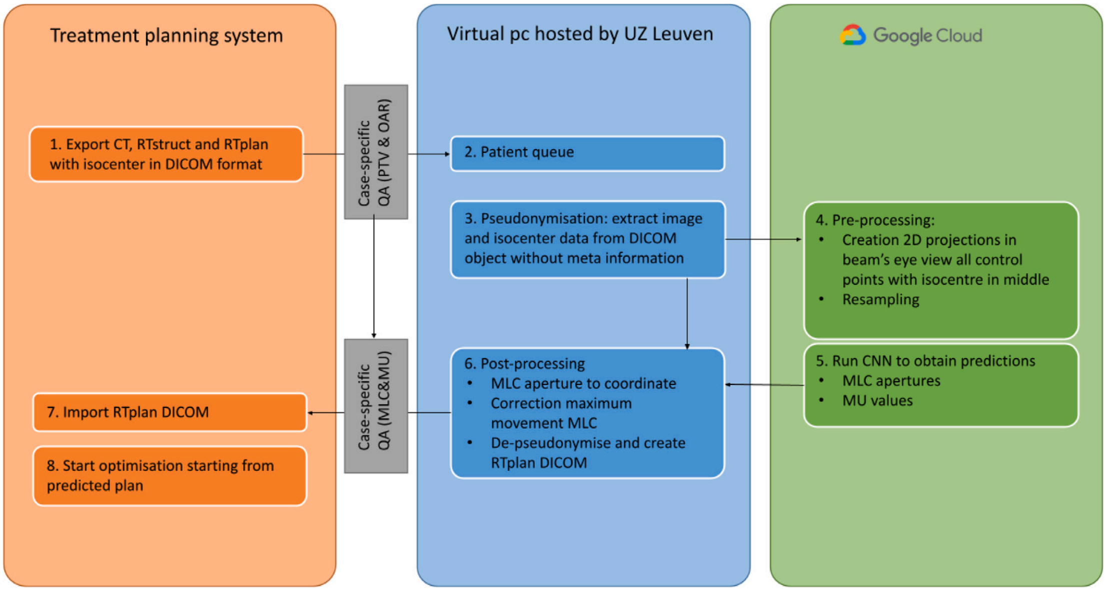

Treatment planning system
1. Export CT, RTstruct and RTplan with isocenter in DICOM format
7. Import RTplan DICOM
8. Start optimisation starting from predicted plan
Case-specific QA (PTV & OAR)
Case-specific QA (MLC&MU)
Virtual pc hosted by UZ Leuven
2. Patient queue
3. Pseudonymisation: extract image and isocenter data from DICOM object without meta information
6. Post-processing
• MLC aperture to coordinate
• Correction maximum movement MLC
• De-pseudonymise and create RTplan DICOM
Google Cloud
4. Pre-processing:
• Creation 2D projections in beam's eye view all control points with isocentre in middle
• Resampling
5. Run CNN to obtain predictions
• MLC apertures
• MU values

Fig. 2. Workflow to embed the CNN based automatic planning tool in clinical practice. The workflow can be started from our clinically used TPS. The model itself is executed on the cloud.

Table 1
Steps, substeps and failure modes defined in the CNN workflow before applying the corrective actions.

Step name	Failure modes			
	Total	Acceptable	Tolerable	Not acceptable
1. Creation inputs model	9	9	0	0
1.1 CT	4	4	0	0
1.2 Contours physicians	3	3	0	0
1.3 Auxiliary contours planning	2	2	0	0
2. Creation initial plan	0	0	0	0
2.1 Position isocentre	0	0	0	0
2.2 Connect structure set	0	0	0	0
3. Get prediction of baseline plan	21	19	2	0
3.1 Export CT, structure set and isocentre	8	8	0	0
3.2 Autoplanning model	9	7	2	0
3.3 Import RTplan	4	4	0	0
4. Further optimisation	9	5	3	1
4.1 Adapt calculation models and options	3	1	2	0
4.2 Insert correct dose prescription	1	0	1	0
4.3 Verify MLC leaf positions	1	1	0	0
4.4 Use RapidPlan	0	0	0	0
4.5 Optimisation	4	3	0	1

patients per year. This was relevant since failures such as 'protocol not clear', 'inattention of the staff' or 'inadequate training of the staff' will occur more often when the workflow is not often used. Furthermore, the size of the training set (about 100) was taken into account to choose an occurrence score for outlier patients in both anatomy and positioning on CT, estimating the probability to 1%.

The final part was defining **corrective actions**. For failure modes deemed 'Tolerable' or 'Not acceptable', according to the risk matrix in Fig. 3, preventive actions and barriers were proposed by the team. Preventions prevent the failure mode from occurring, while barriers prevent the failure mode effect to occur when a failure mode did already happen. When a corrective action was identified, its effectiveness was re-evaluated, and the risk scores were updated.

2.3. Corrective actions

The first corrective action consisted of **case-specific Quality Assurance (QA)**, which was defined as an evaluation of patient-specific inputs/outputs from the model as in [9,15], aimed at identifying those patients who (1) fell outside the model's training scope or (2) whose predictions deviated from expectations. A report in PDF format was automatically generated, as shown in the supplementary materials, detailing the quality of the inputs/outputs of the model relative to the training, validation, and test data and sending this to the dosimetrist to inform them.

The inputs to the CNN model included of projections of electron density within different OARs/PTVs in the BEV of the control points, centring around the isocentre. Therefore, three different metrics were calculated to detect deviations, including (1) the volume of OARs/PTVs, (2) mean Hounsfield Units (HU) within these areas, and (3) the distance between the plan's isocentre and the centres of mass of OARs/PTVs in the transversal, longitudinal, and vertical directions. For each metric, the range of the training data (101 patients) was shown together with the position of the metric calculated on the new patient data.

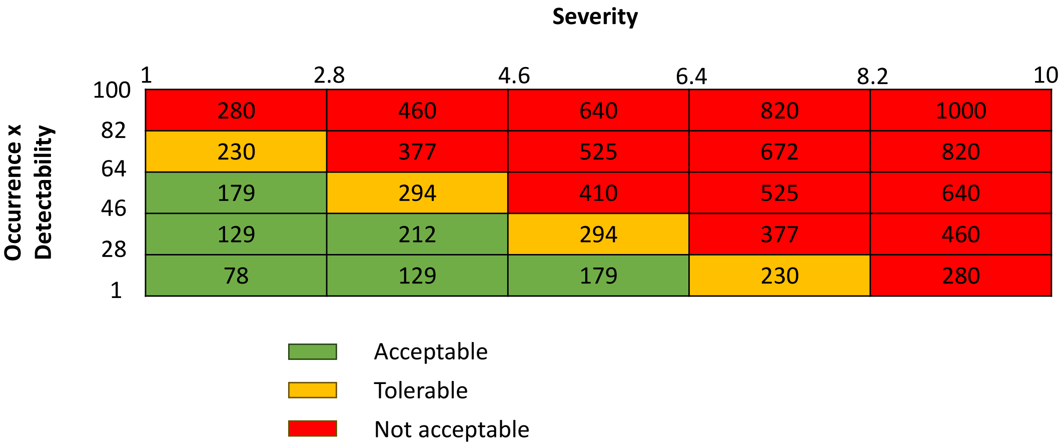
The outputs from the CNN based planning tool, namely the MU (Monitor Unit) values and MLC (Multi-Leaf Collimator) apertures per control point, were analysed using summarised metrics due to the large number of variables to check. Metrics such as Aperture Area Variability

Table 2
Description of Severity, Occurrence and (lack of) Detectability scores used in TG-100 [14].

Severity of effect (S) Qualitative description	Occurrence (O) Pocc [%]	(Lack of) Detectability (D) Pmiss [%]	Related S/O/D score
No effect	0.01	0.01	1
Inconvenience	0.02	0.2	2
Inconvenience	0.05	0.5	3
Minor dosimetric error	0.1	1	4
Limited toxicity or tumour underdose	0.2	2	5
Limited toxicity or tumour underdose	0.5	5	6
Serious toxicity or tumour underdose	1	10	7
Serious toxicity or tumour underdose	2	20	8
Very serious toxicity or tumour underdose	5	50	9
Catastrophic	100	100	10

Table 3
Failure modes that fall in the ‘Tolerable’ and ‘Not acceptable’ region of the used risk matrix before the corrective action and their new scores after the corrective action (O’, D’ and Risk matrix’).

Step	Failure mode	Cause	Effect	S	O	D	Risk matrix	Corrective action	O’	D’	Risk matrix’
3.2	Prediction is bad	Patient falls outside the range of the training data of the deep learning model	Sub-optimal plan	4	7	8.17	Tolerable	Barrier: Case-specific QA	7	3.51	Acceptable
3.2	De-pseudonymisation with wrong patient data	Software failure (bug)	Wrong volume	7	1	8.17	Tolerable	Barrier: Time stamps	1	4.51	Tolerable
4.1	Multiresolution level not changed to MR3 before optimisation	Protocol not clear/inattention staff/inadequate training staff	Sub-optimal plan	4	8.33	7	Tolerable	Prevention: Pop-up window	3.79	7	Acceptable
4.1	Aperture shape controller not changed before optimisation	Protocol not clear/inattention staff/inadequate training staff	Sub-optimal plan	4	8.33	7	Tolerable	Prevention: Pop-up window	3.79	7	Acceptable
4.2	Wrong dose prescription inserted	Protocol not clear/miscommunication/ inattention staff	Wrong dose distribution	8	2	2	Tolerable	Prevention: Scripted checks and properties	1	2	Tolerable
4.5	No automatic optimisation mode and intermediate dose calculation checked	Protocol not clear/inattention staff/inadequate training staff	Sub-optimal plan	4	8.33	9	Not acceptable	Prevention: Pop-up window	3.79	9	Acceptable



🟥 280	🟥 460	🟥 640	🟥 820	🟥 1000
🟧 230	🟥 377	🟥 525	🟥 672	🟥 820
🟩 179	🟧 294	🟥 410	🟥 525	🟥 640
🟩 129	🟩 212	🟧 294	🟥 377	🟥 460
🟩 78	🟩 129	🟩 179	🟧 230	🟥 280

🟩 Acceptable
🟧 Tolerable
🟥 Not acceptable

Fig. 3. Risk matrix used in our department to assess the failure modes that need corrective actions. The values inside the matrix are the multiplication of the occurrence, lack of detectability and severity scores.

(AAV) and Leaf Sequence Variability (LSV) [16,17] per arc could assess the complexity of MLC movements, while total MU and its standard deviation per arc could gauge MU consistency. These output parameters were visualised across the validation and test datasets (47 patients).

The second corrective action consisted of **time stamps**. In Section 2.1, a patient queue on the virtual pc in UZ Leuven ensured individual patient processing. It was decided to use time stamps to verify the absence of data switching between patients by comparing the interval between pseudonymisation and de-pseudonymisation steps. Anomalies, where time exceeds typical durations observed in validation/test datasets, could be flagged in red in the case-specific QA report.

Another corrective action, was to use ESAPI to **script some checks and properties**. It was decided to use ESAPI to automatically adapt the calculation models and dose prescription when the predicted plan was imported in the TPS.

For the properties that could not be scripted using ESAPI, It was decided to implement a **pop-up reminder** that appeared when the plan was imported, prompting the user to adjust the aforementioned properties manually.

The proposed methodology in the current section follows from the results of the risk analysis. The full implementation of the corrective actions is outside the scope of this work.

3. Results

3.1. Failure modes

Across the workflow's substeps, 39 failure modes were identified (see Table 1). Of these, 33 were 'Acceptable', five 'Tolerable', and one 'Not acceptable' according to the risk matrix (see Fig. 3 and Table 3).

Nine different failure modes were defined in the 'Creation inputs model' step, which were all acceptable. Failures modes consisted of incorrect posture during the CT scan, inappropriate scan range (e.g., missing liver), or suboptimal positioning of the radio-opaque wire. Failures like missing contours or incorrect contour names were countered by using contouring templates and the automated tool for auxiliary structure creation. Additional potential failures involved the need to create an extra bolus structure in the CNN workflow. However, this had a low lack of detectability score since the CNN workflow would throw an error and would not run further, serving as a barrier.

The 'get prediction of baseline plan' step identified 21 failure modes, two of which were classified as 'Tolerable' on the risk matrix. Common issues included wrong data export (e.g., wrong structure set exported), incomplete data export, software bugs, or errors during the import of the predicted plan. These generally caused disruptions by halting the CNN workflow or triggering errors, with scripts crashing acting as a barrier to prevent more severe outcomes.

Detailed in Table 3, one 'Tolerable' failure mode was bad predictions by the DL model for atypical cases, such as patients with unusually large breast volumes or changes in CT scanner specifications. The estimated occurrence was not that high (1%) leading to an occurrence score of seven. However, the lack of detectability was high, and while the resulting plan might have failed to meet clinical constraints – noticeable to dosimetrists or physicians – the specific cause might have been hard to detect.

Another failure involved the de-pseudonymisation of data due to an error, swapping meta information with another patient's. This was rare due to rigorous coding, yet it was severe and hard to detect, placing it in the 'Tolerable' category of the risk matrix.

During further optimisation of the predicted plan, certain properties like multiresolution level, aperture shape controller, and dose prescription had to be manually specified, unlike in the clinical workflow. The high likelihood of incorrectly setting these properties, coupled with the difficulty in detecting such errors based on the plan's final outcome, resulted in three properties being categorised as 'Tolerable' and one as 'Not acceptable' on the risk matrix. Barriers to mitigate these risks included institutional clinical goals and supervision by physicians and medical physicists.

3.2. Corrective actions

The **case-specific QA report** was chosen as a barrier against the 'prediction is bad' failure mode (see Table 3), which decreased the lack of detectability score. Nonetheless, there remained a small risk (estimated at 3%) that the dosimetrist overlooked the report.

The risk score of 'De-pseudonymisation with wrong patient data' was decreased using **time stamps**. Although this barrier reduced the lack of detectability score, errors could still occur if the dosimetrist overlooked or misinterpreted the report. Given its potential impact, this failure mode remained classified as 'Tolerable' in the risk matrix.

The occurrence score of the failure mode of inserting a wrong dose prescription was reduced to a minimum by **scripting some checks and properties** using ESAPI. However, the failure mode remained in the 'Tolerable' zone of the risk matrix due to its high severity score.

A **pop-up** to remind the dosimetrist to adapt the multiresolution level, the aperture shape controller, the optimisation mode and intermediate dose calculation before optimisation, decreased the occurrence score of these failure modes, moving them into the 'Acceptable' zone of the risk matrix.

4. Discussion

An illustration was given of a pro-active risk analysis of a clinical workflow using an in-house DL model for breast VMAT planning prior to clinical use. A multidisciplinary team conducted a FMEA risk analysis, identifying several failure modes requiring intervention. Ideas for corrective actions included generating a case-specific QA report, collecting time stamps, scripting optimisation properties and using pop-up reminders. As a result of these corrective actions, only two failure modes remained in the 'Tolerable' zone of the risk matrix.

The case-specific QA report consisted of an input and output analysis, which has been successfully applied before for QA of automatic contours [18]. The input analysis was based on the knowledge that CNNs are most reliable when applied to data similar to their training data [15]. The output analysis was based on summarised metrics calculated on the output of the CNN, compared to the validation and test set. Including the validation set was not ideal, but increased the amount of patients obtaining a more reliable population.

ESAPI's capabilities had limitations; it could not automatically define all properties, such as multiresolution level and aperture shape controller. The automatic optimisation mode, intermediate dose calculation and continue previous optimisation could only be controlled automatically if the whole residual optimisation was performed using scripting, which was not the goal in the CNN workflow. To counteract this, the idea was to create a pop-up reminder. Although less good than scripting the properties itself, it reduced the occurrence score.

It is important to notice that the failure modes and scores are highly dependent on the local deployment [13]. Different deployments might lead to different failure modes and scores. The performed risk analysis is therefore only valuable for our own clinical implementation, but can be used as an illustration for other clinical implementations. Furthermore, the scores are dependent on the persons that perform the risk analysis. Other teams might reveal other failure modes requiring intervention. However, to our opinion, the most important failure modes would remain the same.

The risk analysis for the CNN tool was conducted before user implementation, potentially overlooking certain failure modes that might emerge with actual use. Periodic retrospective analysis and updates to the FMEA are crucial parts of our in-house risk management strategy to address any new failure modes [9].

Clinical implementation of automated planning for breast radiotherapy using ML/DL has gained a lot of interest in literature recent years. Several methods exist to perform this task. Models predict either DVH [19] or 3D dose distributions [20,21] after which the plan parameters as the MLC coordinates and MU values are obtained during an optimisation process. It has been shown that both methods lead to similar dosimetric plans [22]. Our method works differently. The DL model directly predicts the MLC aperture and MU value per control point. The parameters can then be given as initialisation to the optimisation, reducing optimisation time in this way. During the retrospective analysis, it was found that planning with the created DL model takes about half of the time of a workflow based on a DVH prediction model with similar plan quality [2]. However, a prospective clinical evaluation is important to capture the real-world clinical decisions [23]. With the clinical implementation now available, we have the option to explore the clinical applicability of the model.

A risk analysis is not the only challenge of using in-house created models in clinical routine. Following the Medical Device Regulation (MDR) [24], every in-house created medical device, including our tool, must meet extensive requirements before clinical use. These include proving the tool's working principle (already performed in [2]), full documentation of the tool, providing a user manual detailing intended use, and establishing post-market surveillance mechanisms [25]. Additionally, pro-active risk analysis is required. While these steps enhance patient safety, they significantly increase the paperwork and complexity involved in validating new image processing algorithms.

To conclude, the pro-active risk analysis revealed possible risks in the implemented workflow for automated planning and led to the implementation of mitigation strategies that decreased the risk scores for safer clinical use.

CRediT authorship contribution statement

Liesbeth Vandewinckele: Designed the study, Implemented the workflow to the cloud, Set up, guided and analysed the risk analysis, Part of the multidisciplinary team that performed the risk analysis, Writing – original draft. **Chahrazad Benazzouz:** Set up, guided and analysed the risk analysis, Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Laurence Delombaerde:** Created the GUI, Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Laure Pape:** Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Truus Reynders:** Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Aline Van der Vorst:** Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Dylan Callens:** Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Jan Verstraete:** Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Adinda Baeten:** Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Caroline Weltens:** Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing. **Wouter Crijns:** Designed the study, Part of the multidisciplinary team that performed the risk analysis, Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to help shorten the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

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

Acknowledgements

The authors would like to thank Cédric Devroye and Jo Vanvoorden of the IT team of UZ Leuven for their help in the implemented workflow. Liesbeth Vandewinckele is supported by a Ph.D. fellowship of the research foundation – Flanders (FWO), under grant number 1SA6123N.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.phro.2024.100677.

References

[1] 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. http://dx.doi.org/10.1016/j.ijrobp.2010.12.007.

[2] Vandewinckele L, Reynders T, Weltens C, Maes F, Crijns W. Deep learning based MLC aperture and monitor unit prediction as a warm start for breast VMAT optimisation. Phys Med Biol 2023;68:225013. http://dx.doi.org/10.1088/1361-6560/ad07f6.

[3] Cai W, Ding S, Li H, Zhou X, Dou W, Zhou L, et al. Automatic IMRT treatment planning through fluence prediction and plan fine-tuning for nasopharyngeal carcinoma. Radiat Oncol 2024;19:39. http://dx.doi.org/10.1186/s13014-024-02401-0.

[4] 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. http://dx.doi.org/10.1002/mp.13271.

[5] Nguyen D, Jia X, Sher D, Lin M, 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;64:065020. http://dx.doi.org/10.1088/1361-6560/ab039b.

[6] Li X, Zhang J, Sheng Y, Chang Y, Yin F-F, Ge Y, et al. Automatic IMRT planning via static field fluence prediction (AIP-SFFP): a deep learning algorithm for real-time prostate treatment planning. Phys Med Biol 2020;65:175014. http://dx.doi.org/10.1088/1361-6560/aba5eb.

[7] Vandewinckele L, Willems S, Lambrecht M, Berkovic P, Maes F, Crijns W. Treatment plan prediction for lung IMRT using deep learning based fluence map generation. Phys Med 2022;99:44–54. http://dx.doi.org/10.1016/j.ejmp.2022.05.008.

[8] Ni Y, Chen S, Hibbard L, Voet P. Fast VMAT planning for prostate radiotherapy: dosimetric validation of a deep learning-based initial segment generation method. Phys Med Biol 2022;67:155016. http://dx.doi.org/10.1088/1361-6560/ac80e5.

[9] Vandewinckele L, Claessens M, Dinkla AM, Brouwer CL, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. Radioth Oncol 2020;153:55–66. http://dx.doi.org/10.1016/j.radonc.2020.09.008.

[10] Lizondo M, Latorre-Musoll A, Ribas M, Carrasco P, Espinosa N, Coral A, et al. Pseudo skin flash on VMAT in breast radiotherapy: optimization of virtual bolus thickness and HU values. Phys Med 2019;63:56–62. http://dx.doi.org/10.1016/j.ejmp.2019.05.010.

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

[12] Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr 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:S170–3. http://dx.doi.org/10.1016/j.ijrobp.2007.06.081.

[13] 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. Radioth Oncol 2010;94:367–74. http://dx.doi.org/10.1016/j.radonc.2009.12.040.

[14] Huq MS, Fraass BA, Dunscombe PB, Gibbons Jr 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 2016;43:4209–62. http://dx.doi.org/10.1118/1.4947547.

[15] Claessens M, Oria CS, Brouwer CL, Ziemer BP, Scholey JE, Lin H, et al. Quality assurance for AI-based applications in radiation therapy. Semin Radiat Oncol 2022;32:421–31. http://dx.doi.org/10.1016/j.semradonc.2022.06.011.

[16] McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. Med Phys 2010;37:505–15. http://dx.doi.org/10.1118/1.3276775.

[17] Crijns W, Defraene G, Van Herck H, Depuydt T, Haustermans K, Maes F, et al. Online adaptation and verification of VMAT. Med Phys 2015;42:3877–91. http://dx.doi.org/10.1118/1.4921615.

[18] Hui CB, Nourzadeh H, Watkins WT, Trifiletti DM, Alonso CE, Dutta SW, et al. Quality assurance tool for organ at risk delineation in radiation therapy using a parametric statistical approach. Med Phys 2018;45:2089–96. http://dx.doi.org/10.1002/mp.12835.

[19] Fogliata A, Nicolini G, Bourgier C, Clivio A, De Rose F, Fenoglietto P, et al. Performance of a knowledge-based model for optimization of volumetric modulated arc therapy plans for single and bilateral breast irradiation. PLoS One 2015;10:e0145137. http://dx.doi.org/10.1371/journal.pone.0145137.

[20] Hedden N, Xu H. Radiation therapy dose prediction for left-sided breast cancers using two-dimensional and three-dimensional deep learning models. Phys Med 2021;83:101–7. http://dx.doi.org/10.1016/j.ejmp.2021.02.021.

[21] Bakx N, Bluemink H, Hagelaar E, van der Sangen M, Theuws J, Hurkmans C. Development and evaluation of radiotherapy deep learning dose prediction models for breast cancer. Phys Imaging Radiat Oncol 2021;17:65–70. http://dx.doi.org/10.1016/j.phro.2021.01.006.

[22] Portik D, Clementel E, Krayenbühl J, Bakx N, Andratschke N, Hurkmans C. Knowledge-based versus deep learning based treatment planning for breast radiotherapy. Phys Imaging Radiat Oncol 2024;29:100539. http://dx.doi.org/10.1016/j.phro.2024.100539.

[23] Bakx N, van der Sangen M, Theuws J, Bluemink J, Hurkmans C. Evaluation of a clinically introduced deep learning model for radiotherapy treatment planning of breast cancer. Phys Imaging Radiat Oncol 2023;28:100496. http://dx.doi.org/10.1016/j.tipsro.2023.100211.

[24] Regulation (EU) 2017/745 of the European Parliament and of the Council - of 5 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.

[25] Beckers R, Kwade Z, Zanca F. The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics. Phys Med 2021;83:1–8. http://dx.doi.org/10.1016/j.ejmp.2021.02.011.