Physics and Imaging in Radiation Oncology 31 (2024) 100627

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

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

Safety and efficiency of a fully automatic workflow for auto-segmentation in radiotherapy using three commercially available deep learning-based applications

Hasan Cavus [a,b,c,*], Philippe Bulens [a,b], Koen Tournel [a,b], Marc Orlandini [a,b], Alexandra Jankelevitch [a,b], Wouter Crijns [d,e], Brigitte Reniers [c]

[a] Department of Radiation Oncology, Jessa Hospital, 3500 Hasselt, Belgium
[b] Limburg Oncology Center, 3500 Hasselt, Belgium
[c] Faculty of Engineering Technology, Hasselt University, B-3590 Diepenbeek, Belgium
[d] Department Oncology, Laboratory of Experimental Radiotherapy, KU Leuven, Belgium
[e] Department of Radiation Oncology, UZ Leuven, Belgium

ARTICLE INFO

Keywords:
Automation
Deep learning
ESAPI
Segmentation

ABSTRACT

Advancements in radiotherapy auto-segmentation necessitate reliable and efficient workflows. Therefore, a standardized fully automatic workflow was developed for three commercially available deep learning-based auto-segmentation applications and compared to a manual workflow for safety and efficiency. The workflow underwent safety evaluation with failure mode and effects analysis. Notably, eight failure modes were reduced, including seven with severity factors ≥7, indicating the effect on patients, and two with Risk Priority Number value >125, which assesses relative risk level. Efficiency, measured by mouse clicks, showed zero clicks with the automatic workflow. This automation illustrated improvement in both safety and efficiency of workflow.

1. Introduction

To allow conformal and accurate treatment planning in radiotherapy, precise delineation of the target volume and organs at risk (OARs) is of utmost importance [1]. Manual contouring on the Computed Tomography (CT) images used for planning (planning-CT) is considered the gold standard [2].

In the last two decades, significant progress has been made in implementing multi-modality imaging with image registration [3,4] and dynamic image series [5,6], allowing for much higher precision in the delineation of target volumes and OARs. However, even with all tools at hand, delineation remains time-consuming. Additionally, delineation is subjective and dependent on the observer, as has been shown by numerous articles on inter-observer [7,8], intra-observer [9], as well as inter- and intra-institute [10] contouring variations. To overcome these issues, auto-contouring (or auto-segmentation) methods have been developed to complement manual delineation [11]. Furthermore, a modern radiotherapy workflow, including contouring, is complex and requires highly qualified staff. This complexity arises from the advancement of techniques aimed at precisely delivering target volumes while minimizing damage to surrounding healthy tissue. Therefore, the time needed to apply this workflow increases significantly. Automation of this workflow offers a possible solution to manage resource constraints [12].

Recently, several commercially available **d**eep **l**earning-based auto-**segmentation** (DL-Segmentation) applications have been developed. Clinical evaluation of these applications has already been performed [13–16]. However, their integration in a fully automated workflow has not yet been widely reported. The purpose of this paper is twofold. Firstly, to design a standardized, fully automatic workflow for three DL-Segmentation applications, and secondly, to evaluate the safety and efficiency of this automated workflow in comparison to the manual workflow.

2. Materials and methods

2.1. Deep-learning based auto-segmentation applications

The fully automatic workflow was evaluated using three DL-Segmentation applications. Contour ProtégéAI (Application A),

* Corresponding author.
E-mail address: hasan.cavus@jessazh.be (H. Cavus).

https://doi.org/10.1016/j.phro.2024.100627
Received 21 March 2024; Received in revised form 8 August 2024; Accepted 8 August 2024
Available online 13 August 2024

developed by MIM Software Inc. (Cleveland, OH, US), is an artificial neural network model based on the U-Net architecture composed of a convolutional encoding and decoding unit [17]. Syngo.via (Application B), from SIEMENS Healthineers (Erlangen, Germany), is a Deep Image-to-Image Network based on a convolutional encoder–decoder architecture combined with multi-level feature concatenation. Additional convolutional layers are used compared to traditional U-Net [18]. MVISION (Application C) (Helsinki, Finland) is a convolutional neural network that utilizes on an encoder-decoder architecture, similar to 3D U-Net and with more recent residual block [19].

All three DL-Segmentation applications allow the automation of tasks based on incoming data. Subsequently, the data can be automatically exported to another location. Before integration in the workflow, the structures required by the department can be selected and their properties can be adjusted for each segmentation module.

2.2. Fully automatic workflow

The planning CT-set was automatically transferred to ARIA database via DICOM following the completion of CT simulation reconstruction. Subsequently, a binary plug-in script, created using Eclipse Scripting API^{TM} (Varian Medical Systems, Palo Alto) and Structured Query Language, was used to transfer these CT images from the ARIA database to the DL-Segmentation software database. To execute this process, a database daemon was created on the server hosting the ARIA database, enabling selective retrieval of CT images for each patient and their export to a dedicated database.

Once the images reached the DL-Segmentation application database, the server was configured to automatically trigger the relevant segmentation module, based on the anatomical region specified in the series description within the CT DICOM-header. After segmentation, only the structure set containing the segmented structures was automatically transferred to the ARIA database. Finally, this structure set was automatically associated with the corresponding CT set already existing in the ARIA database through a unique identifier in the DICOM-tags.

The automated workflow was developed to be flexible to changes. For example, relocating an application to a new server requires updating DICOM server entities (IP-address, AE-title, and port) in the script and database daemon. Similarly, updating segmentation modules necessitates modifying their identifiers in the automation process. Multiple segmentation modules can also be applied to a single CT set via this automation, particularly if the region of interest covers multiple body parts.

2.3. Safety

The safety of the fully automatic workflow was assessed through a **f**ailure **m**odes and **e**ffects **a**nalysis (FMEA) [20]. This method involves

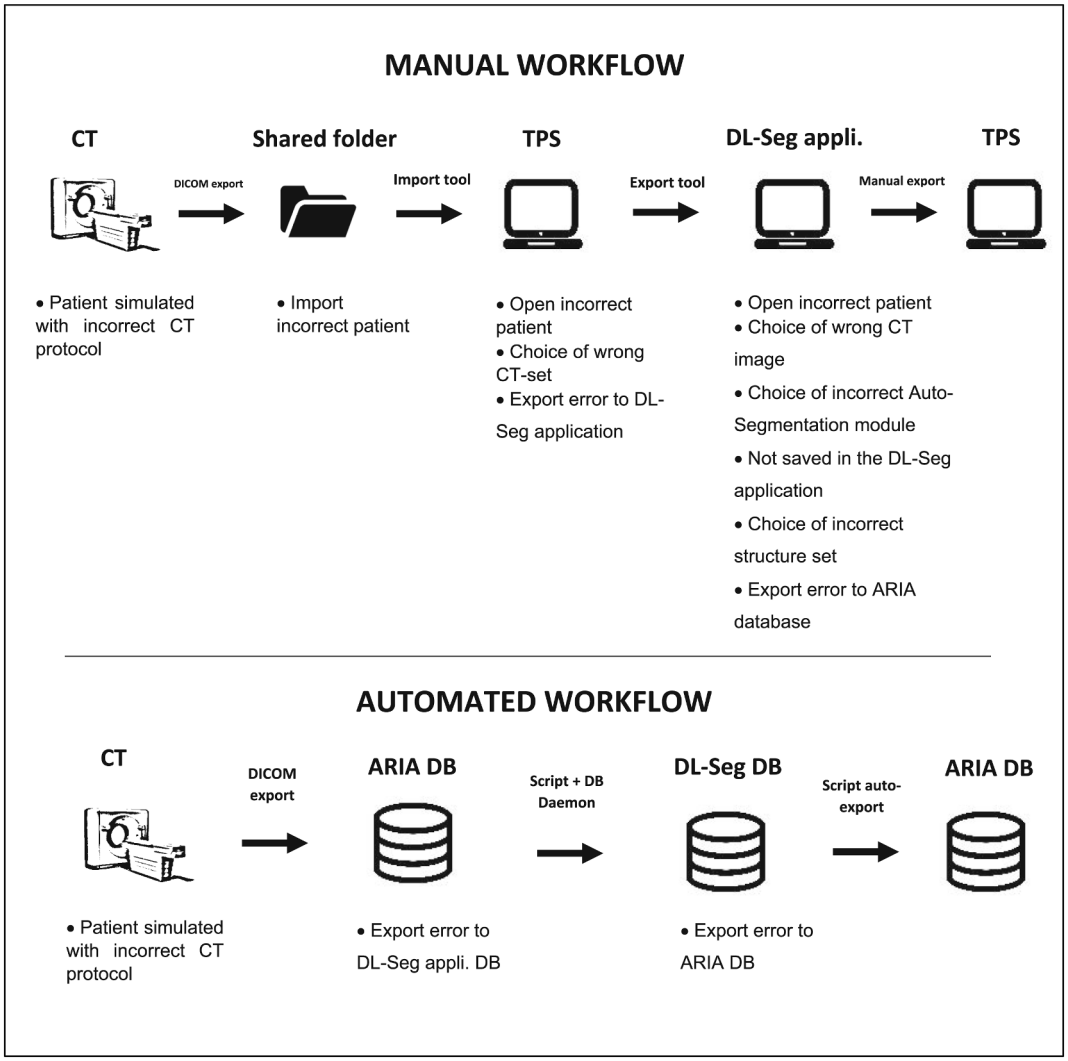

Fig. 1. Manual and automated workflow with their potential identified failure modes listed under each steps. CT: Computed Tomography; DL-Seg appli: DL-Segmentation application; DB: database.

Table 1
Potential failure modes identified and their occurrence (O), severity (S), detectability (D) and Risk Priority Number (RPN) values for manual and automatic workflow. DL-Seg: DL-Segmentation.

Location	Failure Modes	Manual Workflow				Automatic Workflow			
		O	S	D	RPN	O	S	D	RPN
CT	Patient simulated with incorrect CT protocol	1	4	2	8	1	7	2	14
Shared Folder	Import incorrect patient	2	10	1	20	/	/	/	/
ARIA	Open incorrect patient	1	10	1	10	/	/	/	/
	Choice of wrong CT-set	5	7	9	315	/	/	/	/
	Export error to DL-Seg application	8	2	1	16	8	2	1	16
DL-Seg	Open incorrect patient	4	10	1	40	/	/	/	/
	Choice of wrong CT images	2	7	2	28	/	/	/	/
	Choice of incorrect Auto-Segmentation module	7	7	2	98	/	/	/	/
	Not saved in the DL-Seg application	6	2	1	12	/	/	/	/
	Choice of incorrect structure set	5	7	8	280	/	/	/	/
	Export error to ARIA database	1	2	1	2	1	2	1	2

Table 2
Number of required mouse clicks for each subprocess within the manual workflow of the three DL-Segmentation (DL-Seg) applications and the automatic workflow.

Sub-processes	Manual Workflow			Automatic Workflow (clicks)
	Application A (clicks)	Application B (clicks)	Application C (clicks)	
Import CT-set from shared folder to ARIA database	11	11	11	/
Export to DL-Segmentation database	8	8	8	/
Open patient and CT	3	11	4	/
Auto-Segmentation	5	6	5	/
Save and export to ARIA database	10	9	2	/
TOTAL	37	45	30	0

evaluating the potential failure modes and their subsequent impact on a process component if preventive actions are not taken into account. The analysis was carried out by a multidisciplinary team consisting of a quality coordinator, three medical physicists, one radiation oncologist, and four radiation therapists.

The application of the FMEA method comprised multiple steps. Initially, a process tree was created, identifying all subprocesses starting from the completion of the CT simulation until the creation of a structure set ready for review. In the following step, potential failure modes, along with their causes and effects, were identified within each subprocess.

In accordance with the FMEA methodology, three parameters were quantitatively evaluated for each potential failure mode. Occurrence (O) is the likelihood of a failure mode occurring; Severity (S) is the degree of severity of an undetected failure mode; Detectability (D) is the likelihood of a failure mode not being detected.

Each parameter was rated on a scale from 1 to 10. The product of these three parameters yielded a metric known as the Risk Priority Number (RPN): $RPN = O \times S \times D$. The highest RPN value assigned to a failure mode signified it as the most frequent, the most severe, and the least detectable.

2.4. Efficiency

The metric used to assess the efficiency of the fully automatic workflow was the number of mouse clicks required from the conclusion of the CT simulation to the point where the structures were ready for review.

3. Results

In the process tree of the manual workflow, eleven potential failure modes were identified from the completion of the CT simulation until the structures were ready for review (supplementary material). The automated workflow reduced certain subprocesses from the manual workflow, leading to a decrease of eight identified failure modes (Fig. 1).

Furthermore, the manual workflow exhibited seven failure modes with a severity factor equal to or greater than 7. All of them were eliminated using the automatic workflow. Additionally, two failure modes had an RPN exceeding 125 with the manual workflow. After implementation of the automatic workflow, both were eliminated. Nonetheless, the severity of one individual failure mode, namely the selection of an incorrect CT protocol during the simulation, increased within the automated workflow, resulting in the value rising from 4 to 7 (Table 1).

The required number of mouse clicks, from the completion of the CT simulation until the structures were ready for review, decreased from 37, 45, and 30 in the manual workflow for Application A, B, and C, respectively, to zero clicks in the fully automatic workflow (Table 2).

4. Discussion

In our department, a standardized fully automatic workflow was implemented utilizing three DL-Segmentation applications for all tumor localizations. This workflow allowed to use a recent auto-segmentation technique for all patients without increasing the initial radiotherapy workflow. The automated process ran completely in the background, allowing concurrent TPS use for other tasks. Compared to the manual workflow, this automated approach proved safer and more efficient, as evaluated by the FMEA method and the number of mouse clicks.

The manual workflow used in clinical practice for modern radiotherapy is complex and time-consuming [21,22]. Users need to understand each step thoroughly and apply it cautiously to prevent failure modes. Despite appropriate training and awareness, avoiding failure modes entirely remains challenging. Furthermore, efficiency gains are

desired as a manual workflow is time- and resource-consuming. To overcome these issues, an in-house automated workflow was developed and implemented.

Improving safety is important when applying new workflows in clinical practice [23,24]. Using the FMEA method, we showed an enhancement in safety when automating a previous manual workflow. In this automated process, importantly, two failure modes with the highest RPN values (>125) were eliminated. A failure mode with an RPN value higher than 125 is identified as high risk and it is recommended to take preventive action [20]. Furthermore, it is reported that the potential for serious toxicity for organs at risk begins when a failure mode has a severity factor of 7 [20]. An automated workflow could successfully address and avoid such failure modes in seven cases. A possible explanation of these result is that automation reduces user intervention and therefore increase safety. Veronese et al. [25] similarly concluded that minimizing the amount of tools and software used per operator in the radiotherapy process contributes to reducing the frequency of failure modes and, consequently, decreasing the overall risk of accidents.

Although the automated workflow effectively resolved many failure modes, an unexpected observation was that the severity of one particular failure mode increased. This particular issue originated from the selection of an incorrect CT protocol during the simulation, leading to the incorrect application of the auto-segmentation module associated to the anatomical region. In the automated process, the corresponding auto-segmentation module was assigned using the series description in the DICOM- header, which was directly linked to the identification of the CT protocol. Considering that the occurrence factor and the detectability factor value were at their minimum, resulting in a small RPN, the likelihood of this failure mode occurring on a patient is low. In the future clearer description of the scan protocols in combination with specific training for the operators may reduce this risk factor.

Additionally, the manual workflow required a different number of mouse clicks for each application. This variation was principally due to the procedures of opening, closing, and exporting CT-set within each application. With the automatic workflow, the number of mouse clicks decreased for each application, which demonstrates an enhanced efficiency [26,27]. Nowadays, radiotherapy departments are increasing their focus on efficiency as modern radiotherapy is time- and resource-consuming, as was mentioned before. Note that there is a relationship between the number of mouse clicks and the potential failure mode considering that the reduction of these clicks not only saved time but could also lead to a reduction in potential failure modes, as less manual interference is needed.

The major strength of this work lies in the general and rapid applicability in daily clinic of the automated workflow we developed, and this for all users of the TPS used in this study. Small modifications in the script need to be made in the DICOM server entities, such as the IP-address, AE-title, and port for different DL-Segmentation applications. Consequently, the proposed workflow can be also easily adapted for another application or an in-house developed auto-segmentation software at the condition that the software has a database and a database daemon. Additionally, the enhanced safety is important, as more and more daily interest is directed towards quality assurance. We acknowledge however that the FMEA method used is a subjective method of analysis, as this is expert-opinion driven. Nevertheless, this analysis was the most suited for assessment of patient safety of workflows composed of subprocesses [28]. Other more quantitative assessments of risk, like the analysis of Ford et al. [29] are mainly useful for the prevention of pretreatment error, as they demonstrated how employing a clinical error database can identify common failure modes and assess the effectiveness of quality control strategies for standard radiotherapy. To further enhance this automated workflow, in the future more work is needed to establish periodic quality assurance by applying well-known test CT-set to the automatic workflow. Additionally, it would be of interest to assess each DL-Segmentation application using an FMEA method because deep learning models are not perfect and can by itself generate failure modes [30].

In conclusion, an in-house fully automated workflow for three DL-Segmentation applications was developed. This automation, reducing manual subprocesses, improved both safety and efficiency compared to the manual workflow. This was shown through enhanced risk and severity of failure modes using the FMEA analysis, and the reduction of mouse clicks respectively. However, selecting the wrong CT protocol slightly increased the risk of one failure mode. This emphasizes the importance of starting processes correctly and recognizing the factors and triggers that affect their effectiveness.

CRediT authorship contribution statement

Hasan Cavus: Conceptualization, Methodology, Software, Writing – original draft, Investigation, Data curation. **Philippe Bulens:** Project administration, Writing – review & editing, Methodology. **Koen Tournel:** Writing – review & editing, Methodology. **Marc Orlandini:** Writing – review & editing, Methodology, Resources. **Alexandra Jankelevitch:** Writing – review & editing, Methodology. **Wouter Crijns:** Supervision, Writing – review & editing, Methodology. **Brigitte Reniers:** Supervision, Writing – review & editing.

Declaration of Competing Interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2024.100627.

References

[1] Landberg T, Chavaudra J, Dobbs J, Gerard J-P, Hanks G, Horiot J-C, Johansson K-A, et al. ICRU Report 62: Prescribing, recording and reporting photon beam therapy (Supplement to ICRU Report 50). ICRU 1999.
[2] Mercieca S, Belderbos JSA, van Herk M. Challenges in the target volume definition of lung cancer radiotherapy. Transl Lung Cancer Res 2021;10:1983–98. https://doi.org/10.21037/tlcr-20-627.
[3] Combs SE, Baumert BG, Bendszus M, Bozzao A, Brada M, Fariselli L, et al. ESTRO ACROP guideline for target volume delineation of skull base tumors. Radiother Oncol 2021;156:80–94. https://doi.org/10.1016/j.radonc.2020.11.014.
[4] Grosu AL, Piert M, Weber WA, Jeremic B, Picchio M, Schratzenstaller U, et al. Positron emission tomography for radiation treatment planning. Strahlenther Onkol 2005;181:483–99. https://doi.org/10.1007/s00066-005-1422-7.
[5] Dumas M, Laugeman E, Sevak P, Snyder KC, Mao W, Chetty IJ, et al. Comparison of the internal target volume (ITV) contours and dose calculations on 4DCT, average CBCT, and 4DCBCT imaging for lung stereotactic body radiation therapy (SBRT). J Appl Clin Med Phys 2020;21:288–94. https://doi.org/10.1002/acm2.13041.
[6] Wong Yuzhen N, Barrett S. A review of automatic lung tumour segmentation in the era of 4DCT. Rep Pract Oncol Radiother 2019;24:208–20. https://doi.org/10.1016/j.rpor.2019.01.003.
[7] Damato AL, Townamchai K, Albert M, Bair RJ, Cormack RA, Jang J, et al. Dosimetric consequences of interobserver variability in delineating the organs at risk in gynecologic interstitial brachytherapy. Int J Radiat Oncol Biol Phys 2014;89:674–81. https://doi.org/10.1016/j.ijrobp.2014.03.005.
[8] Falcinelli L, Palumbo I, Radicchia V, Arcidiacono F, Lancellotta V, Montesi G, et al. Prostate cancer: contouring target and organs at risk by kilovoltage and megavoltage CT and MRI in patients with and without hip prostheses. Br J Radiol 2015;88:20150509. https://doi.org/10.1259/bjr.20150509.
[9] Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. Radiother Oncol 1998;47:285–92. https://doi.org/10.1016/s0167-8140(98)00021-8.
[10] Fiorino C, Vavassori V, Sanguineti G, Bianchi C, Cattaneo GM, Piazzolla A, et al. Rectum contouring variability in patients treated for prostate cancer: impact on rectum dose-volume histograms and normal tissue complication probability. Radiother Oncol 2002;63:249–55. https://doi.org/10.1016/s0167-8140(01)00469-8.
[11] Gibbons E, Hoffmann M, Westhuyzen J, Hodgson A, Chick B, Last A. Clinical evaluation of deep learning and atlas-based auto-segmentation for critical organs at

risk in radiation therapy. J Med Radiat Sci 2023;70:15–25. https://doi.org/10.1002/jmrs.618.
[12] Krishnamurthy R, Mummudi N, Goda JS, Chopra S, Heijmen B, Swamidas J. Using Artificial Intelligence for Optimization of the Processes and Resource Utilization in Radiotherapy. JCO Glob Oncol 2022;8:e2100393. https://doi.org/10.1200/GO.21.00393.
[13] Doolan PJ, Charalambous S, Roussakis Y, Leczynski A, Peratikou M, Benjamin M, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. Front Oncol 2023;13:1213068. https://doi.org/10.3389/fonc.2023.1213068.
[14] Pera Ó, Martínez Á, Möhler C, Hamans B, Vega F, Barral F et al. Clinical validation of Siemens' Syngo.via automatic contouring system. Adv Radiat Oncol 202316;8:101177. http://doi.org/10.1016/j.adro.2023.101177.
[15] Heilemann G, Buschmann M, Lechner W, Dick V, Eckert F, Heilmann M, et al. Clinical implementation and evaluation of auto-segmentation tools for multi-site contouring in radiotherapy. Phys Imaging Radiat Oncol 2023;28:100515. https://doi.org/10.1016/j.phro.2023.100515.
[16] Palazzo G, Mangili P, Deantoni C, Fodor A, Broggi S, Castriconi R, et al. Real-world validation of Artificial Intelligence-based Computed Tomography auto-contouring for prostate cancer radiotherapy planning. Phys Imaging Radiat Oncol 2023;28:100501. https://doi.org/10.1016/j.phro.2023.100501.
[17] Wan H. Automated Contouring Using Neural Networks. MIM Software Inc., Cleveland, OH, United States. 2020 (White Paper).
[18] Marschner S, Datar M, Gaasch A, Xu Z, Grbic S, Chabin G, et al. A deep image-to-image network organ segmentation algorithm for radiation treatment planning: principles and evaluation. Radiat Oncol 2022;17:129. https://doi.org/10.1186/s13014-022-02102-6.
[19] Kiljunen T, Akram S, Niemelä J, Löyttyniemi E, Seppälä J, Heikkilä J, et al. A deep learning-based automated CT segmentation of prostate cancer anatomy for radiation therapy planning-A retrospective multicenter study. Diagnostics (Basel) 2020;10:959. https://doi.org/10.3390/diagnostics10110959.
[20] 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. https://doi.org/10.1118/1.4947547.
[21] Turcas A, Leucuta D, Balan C, Clementel E, Gheara C, Kacso A, et al. Deep-learning magnetic resonance imaging-based automatic segmentation for organs-at-risk in the brain: Accuracy and impact on dose distribution. Phys Imaging Radiat Oncol 2023;27:100454. https://doi.org/10.1016/j.phro.2023.100454.
[22] Saha M, Jung JW, Lee SW, Lee C, Lee C, Mille MM. A deep learning segmentation method to assess dose to organs at risk during breast radiotherapy. Phys Imaging Radiat Oncol 2023;28:100520. https://doi.org/10.1016/j.phro.2023.100520.
[23] Nishioka S, Okamoto H, Chiba T, Sakasai T, Okuma K, Kuwahara J, et al. Identifying risk characteristics using failure mode and effect analysis for risk management in online magnetic resonance-guided adaptive radiation therapy. Phys Imaging Radiat Oncol 2022;23:1–7. https://doi.org/10.1016/j.phro.2022.06.002.
[24] Klüter S, Schrenk O, Renkamp CK, Gliessmann S, Kress M, Debus J, et al. A practical implementation of risk management for the clinical introduction of online adaptive Magnetic Resonance-guided radiotherapy. Phys Imaging Radiat Oncol 2021;17:53–7. https://doi.org/10.1016/j.phro.2020.12.005.
[25] Veronese I, De Martin E, Martinotti AS, Fumagalli ML, Vite C, Redaelli I, et al. Multi-institutional application of Failure Mode and Effects Analysis (FMEA) to CyberKnife Stereotactic Body Radiation Therapy (SBRT). Radiat Oncol 2015;10:132. https://doi.org/10.1186/s13014-015-0438-0.
[26] Zhou W, Xie Y. Interactive contour delineation and refinement in treatment planning of image-guided radiation therapy. J Appl Clin Med Phys 2014;15:4499. https://doi.org/10.1120/jacmp.v15i1.4499.
[27] Miao J, Niu C, Liu Z, Tian Y, Dai J. A practical method for predicting patient-specific collision in radiotherapy. J Appl Clin Med Phys 2020;21:65–72. https://doi.org/10.1002/acm2.12915.
[28] Liu HC, Zhang LJ, Ping YJ, Wang L. Failure mode and effects analysis for proactive healthcare risk evaluation: A systematic literature review. J Eval Clin Pract 2020;26:1320–37. https://doi.org/10.1111/jep.13317.
[29] Ford EC, Terezakis S, Souranis A, Harris K, Gay H, Mutic S. Quality control quantification (QCQ): a tool to measure the value of quality control checks in radiation oncology. Int J Radiat Oncol Biol Phys 2012;84:e263–9. https://doi.org/10.1016/j.ijrobp.2012.04.036.
[30] Kalantar R, Lin G, Winfield JM, Messiou C, Lalondrelle S, Blackledge MD, et al. Automatic segmentation of pelvic cancers using deep learning: state-of-the-art approaches and challenges. Diagnostics (Basel) 2021;11:1964. https://doi.org/10.3390/diagnostics11111964.