Clinical evaluation of the convolutional neural network-based automatic delineation tool in determining the clinical target volume and organs at risk in rectal cancer radiotherapy

YANGYANG HUANG, RUI SONG, TINGTING QIN, MENGLIN YANG and ZONGWEN LIU

Department of Radiation Oncology, The Second Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450014, P.R. China

Received March 4, 2024; Accepted July 16, 2024

DOI: 10.3892/ol.2024.14672

Abstract. Delineating the clinical target volume (CTV) and organs at risk (OARs) is crucial in rectal cancer radiotherapy. However, the accuracy of manual delineation (MD) is variable and the process is time consuming. Automatic delineation (AD) may be a solution to produce quicker and more accurate contours. In the present study, a convolutional neural network (CNN)-based AD tool was clinically evaluated to analyze its accuracy and efficiency in rectal cancer. CT images were collected from 148 supine patients in whom tumor stage and type of surgery were not differentiated. The rectal cancer contours consisted of CTV and OARs, where the OARs included the bladder, left and right femoral head, left and right kidney, spinal cord and bowel bag. The MD contours reviewed and modified together by a senior radiation oncologist committee were set as the reference values. The Dice similarity coefficient (DSC), Jaccard coefficient (JAC) and Hausdorff distance (HD) were used to evaluate the AD accuracy. The correlation between CT slice number and AD accuracy was analyzed, and the AD accuracy for different contour numbers was compared. The time recorded in the present study included the MD time, AD time for different CT slice and contour numbers and the editing time for AD contours. The Pearson correlation coefficient, paired-sample t-test and unpaired-sample t-test were used for statistical analyses. The results of the present study indicated that the DSC, JAC and HD for CTV using AD were 0.80±0.06, 0.67±0.08 and 6.96±2.45 mm, respectively. Among the OARs, the highest DSC and JAC using AD were found for the right and left kidney, with 0.91±0.06 and 0.93±0.04, and 0.84±0.09 and 0.88±0.07, respectively, and HD was lowest for the spinal cord with 2.26±0.82 mm. The lowest accuracy was found for the bowel bag. The more CT slice numbers, the higher the accuracy of the spinal cord analysis. However, the contour number had no effect on AD accuracy. To obtain qualified contours, the AD time plus editing time was 662.97±195.57 sec, while the MD time was 3294.29±824.70 sec. In conclusion, the results of the present study indicate that AD can significantly improve efficiency and a higher number of CT slices and contours can reduce AD efficiency. The AD tool provides acceptable CTV and OARs for rectal cancer and improves efficiency for delineation.

Correspondence to: Professor Zongwen Liu, Department of Radiation Oncology, The Second Affiliated Hospital of Zhengzhou University, 2 Jingba Road, Zhengzhou, Henan 450014, P.R. China
E-mail: liuzwhh@sina.com

Key words: automatic delineation, convolutional neural network-based, clinical target volume, organs at risk, rectal cancer, clinical evaluation

Introduction

Rectal cancer is a severe malignancy that endangers the lives of patients, which has an incidence rate ranking third in the world and fifth in China (1,2). Radiotherapy is recognized as an essential method in the treatment of rectal cancer (3). It has been demonstrated that the accuracy of the determined clinical target volume (CTV) and organs at risk (OARs) when treating patients is closely related to tumor control and radiation toxicity (4-6), even beyond the impact of inadequate planning and relative significant positioning errors (7,8). The specific process of manual delineation (MD) is composed of the contouring by a junior physician and the review and revision by a senior physician in the same group. Therefore, manual contours can vary markedly due to the differences in the knowledge level and clinical experience among physicians and the anatomical structure among patients (9-13). Meanwhile, it is time-consuming to perform MD for pelvic CTV and OARs, which fills a large proportion of the preparation process before radiotherapy (9,14).

With the development of computer technology, automatic delineation (AD) of CTV and OARs for rectal cancer has been increasingly applied in clinical practice (15-17). The AD tool was developed with the aid of specific software that was trained based on a large sample and can complete the AD of contours for patients in a very short time (18). Meanwhile, the accuracy of AD is increasing (19), and it is independent of the body size, shape and age of patients (20). Among various AD tools, convolutional neural network (CNN)-based tools have been increasingly applied in clinical practice owing to their higher accuracy, and these CNN-based tools have significantly improved working efficiency while ensuring acceptable automatic contours (21-24). Liu *et al* (25) evaluated a deep neural network-based tool for automatic prostate segmentation based on a large cohort of patient images. The authors made comparisons between AD and MD using the Dice similarity coefficient (DSC), Hausdorff distances (HD) and center of mass distances (CMD) and found that the mean DSC and HD were >0.85 and >7.0, respectively, and the mean CMD was within 5 mm. Based on these results, it was concluded that the AD tool used achieved a high level of accuracy in the contours of the prostate gland compared with the consensus contours, thus exhibiting a promising application prospect. Liu *et al* (26) introduced a CNN-based segmentation model that could provide accurate AD in much less time compared with manual contours. The authors also confirmed that the bladder, bone marrow, left femoral head, right femoral head, rectum, small intestine and spinal cord in 105 patients with cervical cancer were delineated by the model. Compared with the corresponding manual contours set as the reference values, the mean DSC was 0.924, 0.854, 0.906, 0.900, 0.791, 0.833 and 0.827 for the bladder, bone marrow, femoral head left, femoral head right, rectum, small intestine and spinal cord, respectively; the mean HD was 5.098, 1.993, 1.390, 1.435, 5.949, 5.281 and 3.269 for the aforementioned OARs, respectively. The results corroborated that the AD contours were highly acceptable in clinical practice. To *et al* (27) and Breto *et al* (28) implemented the CNN-based AD on MR images for prostate and cervical regions, respectively, which achieved desirable contours. Bi *et al* (29) revealed that the tool resulted in a 35% decrease in time spent in the comparison between AD and MD (median, 9.59 vs. 14.81 min; P<0.001). Men *et al* (20) suggested that the test time for the CNN-based AD of the CTV, bladder, left and right femoral heads, colon and intestine in rectal cancer was 45 sec per patient.

The focus of most existing studies is placed on different CNN-based tools used in prostate and cervical cancer (25-28). Furthermore, the existing studies on AD in rectal cancer either addressed one type of patient [such as Sha *et al* (30) studied preoperative radiotherapy patients and Song *et al* (15) analyzed postoperative patients] or they do not discuss the efficiency of AD (31). For patients with rectal cancer that did not require assessment for tumor staging or surgery type, the accuracy and efficiency of CNN-based AD for CTV and OARs are rarely analyzed. Therefore, the present study was conducted to identify whether the commercial CNN-based tool could automatically delineate both CTV and OARs in rectal cancer with high accuracy and efficiency.

Materials and methods

Patient cohort. The planning CT images of 148 patients who were diagnosed with rectal cancer without the distinguishment between tumor stages and surgery were collected from March, 2021 to June, 2024. The present study was approved by the Ethics Committee of The Second Affiliated Hospital of Zhengzhou University (Zhengzhou, China; approval no. 20210302). Written informed consent was obtained from the patients for the use of their anonymized data in the present study. All methods were implemented following relevant guidelines and regulations. The inclusion criteria were as follows: i) Preoperative and postoperative patients with rectal cancer; ii) patients who underwent pelvic CT for rectal cancer radiotherapy; iii) patients who received the examination in a supine position; iv) planning CT with or without intravenous contrast; and v) patients who urinated and then drank 500 ml water 30 min before the CT scan. Patients who had a radiotherapy history and received the examination in a prone position were excluded from the present study. The purpose of radiotherapy in the 105 preoperative patients was to reduce the tumor stage and thus remove the lesions more completely upon surgery; therefore, the radiotherapy administered was not aggressive. The purpose of radiotherapy in the 43 postoperative patients was to irradiate subclinical lesions to reduce the postoperative recurrence rate; therefore, the radiotherapy administered was also not aggressive in this instance. Of the 43 postoperative patients, anastomotic fistula occurred in 4 patients, bleeding occurred in 3, bowel obstruction occurred in 2, urinary retention occurred in 3 and no other postoperative complications were recorded. All patients were immobilized using a radiotherapy-specific thermoplastic mold. A summary of the patient characteristics is shown in Table I.

The planning CT process was divided into 2 days. The first day involved individualized customization of the thermoplastic mold, which was left in place for 24 h to ensure fixation and repeatability throughout the radiotherapy period. CT images were acquired on the second day. All patients were scanned using a Philips Brilliance Big Bore CT (Philips Healthcare). The scan parameters were as follows: i) Scanned slice thickness, 5 mm; ii) tube voltage, 120 kV; and iii) reconstructed with a 512x512 voxel matrix. The field of view was adapted to the size of the patient.

CNN architectures. The CNN architecture used was designed based on the modified U-Net and termed 'RT-Mind' (version 1.1; MedMind Technology Co., Ltd.) (http://www.medicalmind.cn/). The manufacturer's description of the tool is as follows: In the basic U-Net, to learn features from part to whole, the encoder aggregated semantic information by reducing spatial information. Semantic information was received by the decoder. Therefore, the feature extraction ability of the encoder was very important. However, the simple convolution layer of U-Net experienced difficulty in learning complicated features efficiently. Therefore, the entire U-net encoder was replaced by the dual path network (DPN) architecture. A large number of advanced abstract features and parameters were encoded into the input image by the DPN. The core of the DPN was the micro-block, which combined the advantages of both the Residual block and the Dense block into a dual-path architecture. Since the Residual block enabled the reuse of features, while the Dense block enabled the exploration of new features, this combination markedly improved the representation learning ability. To achieve an improved feature extraction capability, the whole DPN92 architecture (32) was used in the U-net encoder. The micro-block was embedded in the decoder to replace the standard convolution operation, thus enabling the decoder to obtain the same performance in recovering abstract features.

A CT slice can be considered a grey-scale image with only one channel. This results in a 2D model that incorporates each CT slice independently. The model used in the present study was designed as a 2.5D architecture by assigning three adjacent slices to three channels, to obtain the 3D information from

Table I. Characteristics of the included 148 patients with rectal cancer.

Characteristics	Value
Median age, years (range)	61 (32-80)
Sex, n (%)	
Male	80 (54.1)
Female	68 (45.9)
RT step, n (%)	
Preoperative RT	105 (70.9)
Postoperative adjuvant RT	43 (29.1)
Tumor stage, n (%)	
II	30 (20.3)
III	71 (48.0)
IV	47 (31.7)
Pathological type, n (%)	
Adenocarcinoma	121 (81.8)
Adenosquamous carcinoma	10 (6.8)
Undifferentiated carcinoma	17 (11.4)
Distance from the anal verge, n (%)	
<5 cm	84 (56.8)
5-10 cm	34 (23.0)
>10 cm	30 (20.2)

RT, radiotherapy.

the CT images. The output was the center slice delineation. The output block can also be modified so that a multi-class segmentation result can be output. The modified U-Net was a combination of the ResNet block and the DenseNet block, thus obtaining the ability to focus on a larger receptive field of the image and to extract more high-level semantic features for ambiguous boundary segmentation.

The learning, validation and test procedures were previously completed by the tool manufacturer at Peking Union Medical College Hospital (Beijing, China) (33,34). Therefore, only the accuracy and efficiency of its clinical use were assessed in the present study. Fig. 1 shows the main components of the CNN architecture.

Contour delineation. Planning CT images were imported into the CNN-based tool for AD. According to the delineation guidelines of Valentini *et al* (35) and the Radiation Therapy Oncology Group (36), the manual contours for all patients were re-delineated by a junior physician and then reviewed and modified together by three oncologists with >10 years collective experience in radiotherapy for pelvic tumors; these contour were set as the reference values. More specifically, in terms of the CTV: i) The high-risk areas of the primary tumor included the tumor or tumor bed, the rectal mesenteric and presacral areas and the target area of low to medium rectal cancer, which included the rectal sciatic fossa; and ii) the regional lymphatic drainage area included the lymphatic drainage area of the common iliac vessels in the true pelvis, the rectal mesenteric area, the lymphatic drainage area of the internal iliac vessels and the closed lymph node area. The MD OARs included the skin, bladder, left femoral head, right femoral head, left kidney, right kidney, spinal cord and bowel bag. The skin was considered the external contour of all CT images. The delineation range for the bladder was defined inferiorly from its base and superiorly to the dome. The delineation range for the left and right femoral heads was defined as the proximal femur inferiorly from the lowest level of the ischial tuberosities (right or left) and superiorly to the top of the ball of the femur, including the trochanters. The kidney was represented by the kidney parenchyma. The bony inner edge of the spinal canal was defined as the spinal cord, including all CT slices. A correctly delineated bowel bag encompassed all the small bowel and colon contours, and the upper bound was 4-5 CT slices further up in the uppermost layer of the CTV (36).

AD accuracy. According to the comparative method proposed by Yeghiazaryan and Voiculescu (37), DSC, Jaccard coefficient (JAC) and HD were used to compare the manual and automatic contours quantitatively. In the following formulae, A and B represent the manual and automatic contour volumes, respectively.

DSC represents the overlap degree between the automatic and manual contours in terms of volume. The range of DSC is 0-1. DSC=0 implies that the two contours do not overlap at all, and DSC=1 implies that the two contours coincide entirely. The DSC can be expressed as follows: DSC (A, B)=$\frac{2|A\cap B|}{|A|+|B|}$.

JAC represents the ratio of the intersection of the automatic and manual contours to their combination, in which the similarity and differences between the two sets of contours are compared. The range of JAC is 0-1. JAC=0 indicates that the automatic contours are entirely inconsistent with the manual contours, and JAC=1 indicates that the automatic contours are completely consistent with the manual contours. The JAC can be expressed as follows: JAC $=\frac{A\cap B}{A\cup B}$.

HD can be used to quantify the maximum distance between the surface of the automatic and manual contours. In addition, this metric can be employed to quantify the maximum distance between two contours by calculating the distance between the closest points in both directions, from contour A to B and vice versa. The HD can be expressed as follows: HD (A, B)=max [h (A, B), h (B, A)], where h (A, B) represents the Euclidean distance between voxels a and b belonging to contours A and B, which can be expressed as follows, h (A, B)=$\max_{b\epsilon B}(\min_{a\epsilon A}\|a-b\|)$.

It is necessary to analyze the variations in AD accuracy in case of differences in the contour and CT slice numbers of these patients. The effect of the CT slice number on the AD accuracy was mainly analyzed based on correlation analyses, while the effect of the contour number on the AD accuracy was mainly achieved by comparing DSC, JAC and HD for two groups with different numbers of contours. To analyze the effect of the contour number on AD accuracy, the AD contours were divided into two groups. Contours such as the CTV, spinal cord, left kidney, right kidney, bowel bag, bladder, left femoral head, right femoral head, left femoral head-neck and right femoral head-neck were assigned to Group C. Contours such as the skin, CTV, bladder, left femoral head, right femoral head, left femoral head-neck, and right femoral head-neck were assigned to Group D. The accuracy of the identical contours between the two groups was compared.

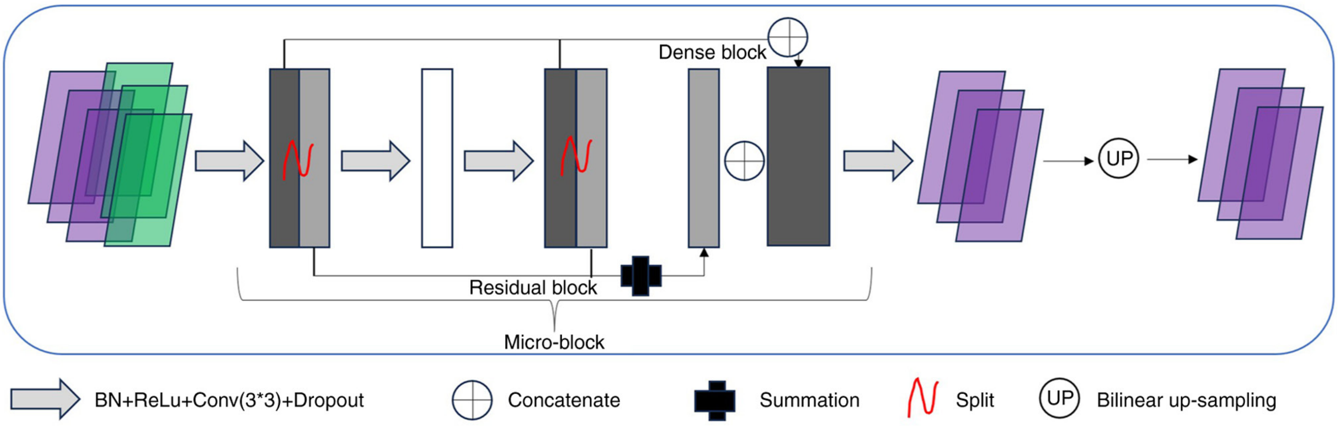

Figure 1. Main components of the convolutional neural network-based tool.

AD efficiency. To analyze the AD efficiency, the duration of the AD and MD were recorded. The MD time included the delineation time of the junior physicians and the review time of the senior physicians. The contours included the skin, CTV, bladder, left femoral head, right femoral head, left kidney, right kidney, spinal cord and bowel bag. The editing time for the AD contours by the senior physicians was also recorded.

To evaluate the influence of the CT slice number on delineation efficiency, the AD time of all 148 patients was counted, and the correlation between CT slice number and delineation time was analyzed. Meanwhile, the influence of the contour number on the AD time was also analyzed. The contour number was divided based on Group C vs. Group D, as described in the *AD accuracy* subsection.

Statistical analysis. Statistical analysis was performed using SPSS 26 (IBM Corp.). The Pearson correlation test was used to analyze the effects of the CT slice number and contour number on the AD time, as well as the effects of the CT slice number on the AD accuracy. The paired t-test was used to compare between Group C to Group D. The unpaired t-test was used to compare the MD time and the AD + editing time. $P<0.05$ was considered to indicate a statistically significant difference.

Results

Accuracy evaluation of the AD contours. The DSC, JAC and HD between the manual (reference values) and the automatic contours are shown in Table II. According to the results, for CTV the DSC was 0.80±0.06, the JAC was 0.67±0.08 and the HD was 6.96±2.45 mm. For the OARs, the left kidney had the highest DSC (0.93±0.04) and JAC (0.88±0.07), and the spinal cord had the lowest HD (2.26±0.82 mm). By contract, the bowel bag had the worst performance in terms of DSC (0.64±0.12), JAC (0.50±0.14) and HD (12.84±4.70 mm). The other OARs had a DSC of >0.83±0.07 (left and right femur), JAC of >0.71±0.10 (left femur) and HD <3.86±1.66 mm (bladder). These results were also pictorially shown in Fig. 2. Therefore, minor modifications (only when needed) of the automatic contours were required before clinical application, except for the bowel bag.

Table II also shows the correlation between the CT slice number and the AD accuracy indices. Most AD accuracy indices were not statistically related to the CT slice number, except for the DSC ($r=0.662$; $P=0.027$) and HD ($r=-0.638$; $P=0.019$) of the spinal cord. The effect of the contour number on the AD accuracy was also collected. The DSC and JAC of all contours were 1.00 and 1.00, respectively, and the HD of all contours ranged from 0.000 to 0.002 mm. Therefore, the contour number had no impact on AD accuracy.

To present the accuracy of AD when determining the CTV, an example was shown in Fig. 3. The automatic contours (cyan) of the CTV indicated a smaller volume than the manual contours (pink) and insufficient layers in the head and foot direction, expect for this, the AD of the CTV was accurate. Fig. 4A, D and E illustrate typical examples of the consistency between the automatic and manual contours of the bladder, left and right kidneys and spinal cord. A match for the femoral head and the femoral head-neck is shown in Fig. 4B and C. In the present study, both the femoral head and the femoral head-neck were delineated by this AD tool. Hence, the results showed a good match, regardless of the contour style selected by physicians. Fig. 4F shows an example of the bowel bag.

Evaluation of AD Efficiency. As shown in Table III, there was a significant difference between the AD + editing time (sum of delineation and review time) and the MD time on the same contours in the same group of patients ($P<0.001$). To obtain accurate contours, the AD + editing time was 662.97±195.57 sec, while the MD time was 3294.29±824.70 sec. In the AD + editing process, the editing time had a predominant proportion with 584.57±193.79 sec. The AD efficiency increased by 5 times. The contours with the most editing time were the bowel bag and CTV.

Additionally, there was a significant positive correlation between the CT slice number and the AD time ($r=0.912$; $P<0.001$), as shown in Table IV. The lower the CT slice number, the shorter the AD time. When these contours were applied in Group C, the AD time was 80.83±9.88 sec in an average CT slice number of 91. When these contours were applied in Group D, the AD time was 57.96±7.90 sec in an average CT slice number of 91. Given that Group D had a lower contour number than Group C, it can therefore be considered that the lower the contour number, the shorter the AD time. This difference was statistically significant ($P<0.001$; Table V).

Table II. DSC, JAC and HD between the automatic and manual contours and their correlation with the CT slice number.

	CT slice number, 90.68±14.72					
Contour	Mean DSC ± SD	r, P-value	Mean JAC ± SD	r, P-value	Mean HD ± SD, mm	r, P-value
CTV	0.80±0.06	0.081, 0.614	0.67±0.08	0.074, 0.604	6.96±2.45	-0.081, 0.688
Bladder	0.90±0.06	-0.143, 0.435	0.83±0.10	-0.147, 0.482	3.86±1.66	0.184, 0.253
L femur	0.83±0.07	-0.265, 0.110	0.71±0.10	-0.299, 0.077	3.32±1.32	0.174, 0.284
R femur	0.83±0.07	-0.327, 0.056	0.71±0.11	-0.284, 0.086	2.88±0.98	0.175, 0.356
Left kidney	0.93±0.04	-0.053, 0.886	0.88±0.07	-0.079, 0.882	3.29±1.65	0.031, 0.924
Right kidney	0.91±0.06	0.029, 0.952	0.84±0.09	0.014, 0.980	3.52±1.60	-0.374, 0.255
Spinal cord	0.84±0.05	0.662, 0.027[a]	0.73±0.08	0.514, 0.091	2.26±0.82	-0.638, 0.019[a]
Bowel bag	0.64±0.12	0.533, 0.087	0.50±0.14	0.486, 0.111	12.84±4.70	-0.557, 0.185

[a]$P<0.05$. DSC, Dice similarity coefficient; JAC, Jaccard coefficient; HD, Hausdorff distance; SD, standard deviation; CTV, clinical target volume; femur L/R, left/right femoral head.

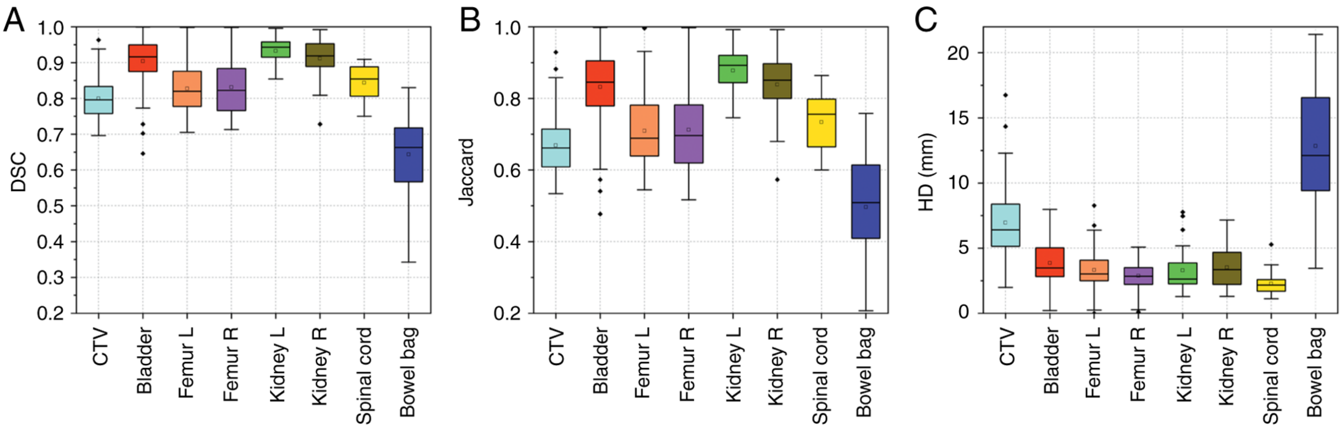

Figure 2. Box plots of the (A) DSC, (B) JAC and (C) HD for the CTV, bladder, femur L, femur R, kidney L, kidney R, spinal cord and bowel bag. CTV, clinical target volume; Dice similarity coefficient; JAC, Jaccard coefficient; HD, Hausdorff distance; femur L/R, left/right femoral head.

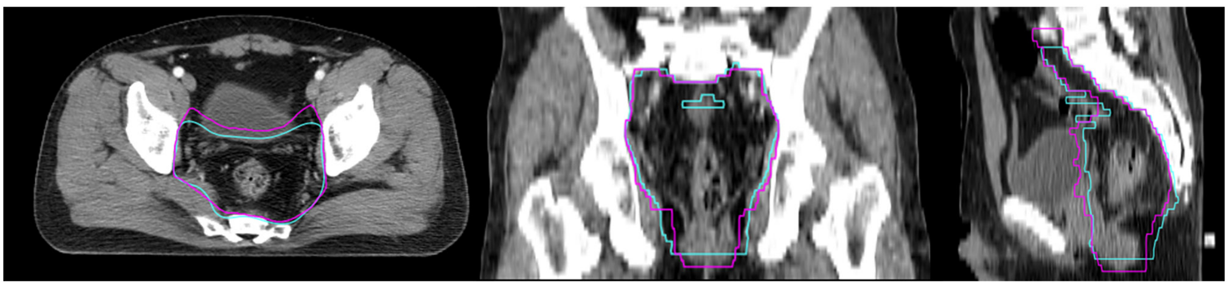

Figure 3. Typical example of the clinical target volume automatic contours (Dice similarity coefficient=0.721, Jaccard coefficient=0.564 and Hausdorff distance=6.596 mm), where the automatic contours are represented in cyan and manual contours are represented in pink.

Discussion

The delineation of the CTV and OARs is essential in the radiotherapy of rectal cancer. The inaccurate delineation of the CTV and OARs is one of the primary factors limiting the feasibility and effectiveness of radiotherapy. At present, contour delineation is still achieved manually, which may infer variable accuracy between observers. Most notably, contour delineation may be the most time-consuming step in radiotherapy (38,39).

As one of the technical solutions of deep learning, CNN-based tools have been applied increasingly in medical image analyses. Hence, it is essential to explore the accuracy of the clinical implementation of these tools. In a previous study, AD and MD were compared in cervical cancer (40). The results revealed that CNN-based AD exhibited improved evaluation results. Moreover, the authors argued that the higher accuracy of their model was attributed to the large number of training cases, which may improve the accuracy

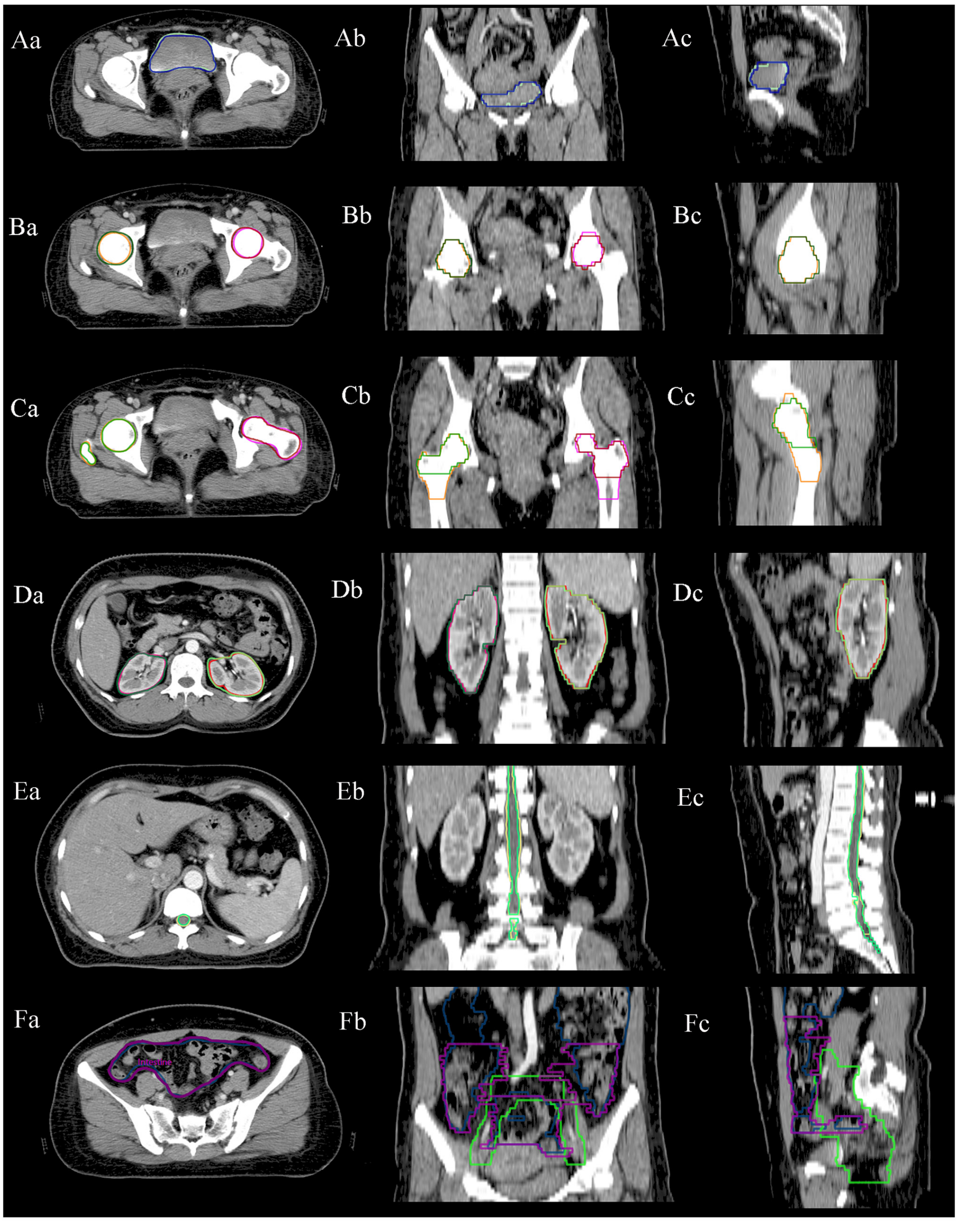

Figure 4. Examples of AD contours vs. MD contours for the (A) bladder (AD in green, MD in blue), (B) left and right femoral heads (AD in yellow and pink, MD in green and red), (C) femoral head-neck (AD in yellow and pink, MD in green and red), (D) left and right kidneys (AD in magenta and red, MD in cyan and green), (E) spinal cord (AD in green, MD in yellow) and (F) bowel bag (AD in blue, MD in purple, clinical target volume in green). MD, manual delineation; AD, automatic delineation.

and robustness of AD in medical centers in developing countries. Zabihollahy *et al* (41) also evaluated the AD of the CTV in cervical cancer. It was found that this tool achieved a DSC of 0.85±0.03 and a 95th percentile HD of 3.70±0.35 mm in the testing cases, which significantly outperformed other novel tools (P<0.05). Meanwhile, the tool generated an uncertainty map using Monte Carlo techniques to draw the attention of physicians to highly uncertain regions where careful review and manual editing may be required. Mohammadi *et al* (42) trained the ResU-Net model based on 72 patients with cervical cancer, which was further verified using 10 patients and tested using 30 patients. The DSC of the testing data set was 95.7±3.7, 96.6±1.5 and 92.2±3.3% for the bladder, rectum and sigmoid, respectively, and the HD was 4.05±5.17, 1.96±2.19 and 3.15±2.03 mm, respectively. The average symmetric surface distance was 1.04±0.97, 0.45±0.09 and 0.79±0.25 mm, respectively, which achieved a good agreement between the automatic and manual contours and improved the robustness of AD. The aforementioned results supported the conclusion that the CNN-based AD tool can provide accurate contours.

Among the contours in the present study, except for the bowel bag, the most significant difference was observed in the CTV. The DSC (0.80±0.06), JAC (0.67±0.08) and HD (6.96±2.45 mm) were somewhat inferior to those reported in other previous studies. Ju *et al* (43) found that the DSC, JAC and HD of automatic contours of the CTV in cervical cancer were 0.82, 0.30 and 1.86 mm, respectively. Men *et al* (20) obtained an average DSC of 0.877 in the automatic contours of the CTV in rectal cancer. Song *et al* (15) concluded that the DSC of the CTV was 0.88 in rectal cancer. As reported

Table III. Duration of AD and MD on the same contours in the same group of patients.

Item	Mean ± SD (range)	t-test	P-value
MD time, sec	3294.29±824.70 (2100-4920)	19.832[a]	<0.001[a]
AD + editing time, sec	662.97±195.57 (246-1172)		
AD time, sec	78.40±10.32 (65-101)		
Editing time for AD contours, sec	584.57±193.79 (180-1080)		

[a]MD time vs. AD + editing time. SD, standard deviation; AD, automatic delineation; MD, manual delineation.

Table IV. Correlation analysis of the CT slice number and AD time.

Item	Mean ± SD (range)	r	P-value
CT slice number, n	90.68±14.72 (70-135)	0.912	<0.001
AD time, sec	80.83±9.88 (67-102)		

SD, standard deviation; AD, automatic delineation.

Table V. Comparison of the AD time for both groups of contour numbers.

Item	Mean ± SD (range)	t	P-value
Group C	80.83±9.88 (67-102)	30.221	<0.001
Group D	57.96±7.90 (48-89)		

SD, standard deviation; AD, automatic delineation.

in certain previous studies (15,20,43), this evidence may be attributed to the perception and bias among observers and the difficulty in differentiating the soft tissues of CTV structures on CT images. It may also be related to the fact that the AD tool used in the present study did not differentiate between preoperative and postoperative rectal cancer options, and the same delineation option was used for all patients. To achieve a larger sample size and to assess the delineation efficiency of the tool, both preoperative and postoperative patients were included in the present study. Although the accuracy of the CTV was somewhat compromised, it remained reasonably accurate. The tool also provided criteria that were consistent with the CTV contours of rectal cancer from leading experts in China (33,34), which made the editing contours and the subsequent clinical work more consistent. To visualize the accuracy of automatic CTV determination, the automatic and manual contours of the CTV of a single patient were also compared in the present study. The AD volume was smaller than the MD volume, but the difference was not notable. In general, necessary modifications are needed after the completion of AD.

In the present study, except for the bowel bag, the DSC of all OARs was >0.83±0.07 (left and right femur), with the bladder reaching 0.90±0.06 and the left and right kidneys reaching 0.93±0.04 and 0.91±0.06, respectively. The JAC was >0.71±0.10 (left femur), particularly for the left kidney, which was 0.88±0.07. The maximum HD was 3.86±1.66 mm for the bladder. The bladder, left and right kidneys and spinal cord had a higher accuracy, exhibiting a larger DSC and JAC and a lower HD, which may be related to there being little inconsistency in the boundary of these OARs and these OARs exhibiting a notably different density from the surrounding tissues. The contouring accuracy of physicians may have also varied, which could be improved by AD tools (29,44,45). For instance, whether the physicians chose to delineate the femoral head or the femoral head-neck, the AD tool had a corresponding contour to match it. The results of the OARs in the present study were consistent with those reported in previous studies (26,40-42), which can be regarded as notable support for the high accuracy of this AD tool in rectal cancer.

In the present study, the effect of the CT slice number on AD accuracy was also analyzed. It was found that only the DSC and HD of the spinal cord were correlated with the CT slice number, and the absolute values of the correlation coefficients were both >0.6 with statistically significant correlations. This indicated that the accuracy of the AD of the spinal cord was strongly correlated with the CT slice number. Based on the fact that the spinal cord is generally only evaluated for the maximum dose, physicians delineate all the CT slices with the spinal cord. This may be related to AD of the spinal cord only having a larger error at the cauda equina division and a smaller error at other parts. Thus, the larger the CT slice number, the smaller the proportion of errors. This suggested that physicians should pay attention to the spinal cord cauda equina junction in the modification of the AD of the spinal cord. In the present study, the AD accuracy was also compared with different contour numbers. The results demonstrated that the contour number had no impact on the AD accuracy, proving the robustness of the AD tool.

The most marked difference between the manual and automatic contours was observed in the bowel bag. The AD of the bowel bag had the smallest DSC and JAC, the largest HD and the greatest dispersion of data. However, there were no marked differences in the number of CT slices in the manual and automatic contours of the bowel bag. In one case, when there were identical CT slices between the AD and MD of the bowel bag, the DSC, JAC and HD were 0.876, 0.780 and 6.764 mm, respectively, which was not notably different from the accuracy of other AD contours. In our department, due to the delineation efficiency and dose volume limits of the bowel bag, physicians only delineated 4-5 CT slices further up in the uppermost layer of the CTV, while the AD of the bowel bag appeared in all CT slices where the bowel bag was present. Hence, the difference in the contour volume naturally resulted in poor accuracy of the bowel bag. When the dose-volume parameters for the bowel bag, such as $V_{50\ Gy}$ <10%, were evaluated, the over-delineation of the bowel bag can lower the evaluation accuracy and may ultimately lead to the emergence of bowel bag toxicity (46). At present, there is no option in the AD tool to decide the range of the bowel bag, and it is suggested that the tool manufacturer should arrange relevant options regarding the delineation range in the future.

In the present study, the editing time for AD contours including CTV and OARs was 584.57±193.79 sec. Although relatively poor delineation accuracy may be obtained, physicians can still modify existing contours to increase the delineation efficiency (15,47). This conclusion was also validated in the present study. To obtain accurate contours, the MD time of contours including CTV and OARs was 3294.29±824.70 sec, while the AD + editing time was 662.97±195.57 sec, and the efficiency increased 5-fold. Owing to the time-saving advantage, AD may increase clinical efficiency and reduce the waiting times for initial treatment. In terms of the advantages of AD, Lustberg *et al* (48) found a median time reduction of 10 min for deep-learning AD. Ginn *et al* (45) analyzed an AD tool for head, neck and pelvic OAR delineation. The time of AD plus the modification process was less than that of MD alone, with an average time reduction of 43.4% or 11.8 min per patient. Hu *et al* (49) analyzed the efficiency of a cloud-based solution for AD. Based on the difference between the average time for AD and MD, the average time reduction was 291 sec for the male pelvic cavity and 210 sec for the female pelvic cavity. In the present study, the mean AD time was 78.40±10.32 sec when CTV, OARs (including skin, spinal cord, left and right kidneys, duodenum, bowel bag, intestinal tube, bladder, left and right femoral heads and left and right femoral head-neck), and an average CT slice number of 91 were selected, which was much shorter than the time (1,560-2,880 sec) for the MD of the pelvic CTV (9) and the MD time (3,294.29±824.70 sec) for delineating the CTV, bladder, left and right femoral heads, left and right kidneys, spinal cord and bowel bag in the present study. To explore the factors affecting the MD efficiency, the CT slice number and the contour number were further analyzed. The CT slice number displayed a significant positive correlation with AD time (r=0.912; P<0.001), indicating that as the CT slice number decreased, the AD time also decreased. Different contour numbers also significantly affected the AD time (P<0.001), indicating that as the contour number decreased, the AD time also decreased.

However, there are several limitations to the present study. First, the present study did not assess the relationship between accurate delineation and treatment effect, and this will be analyzed in our future work. Second, the present study lacks data on recurrence and surgical cure rates in postoperative patients. Third, there is a lack of localization training for the AD tool, increasing the inaccuracy of the automatic contours compared with manual contours.

In summary, the present study demonstrated through quantitative analyses that the CNN-based AD tool could provide a certain degree of clinically acceptable CTV and OARs for patients diagnosed with rectal cancer in whom tumor stage and type of surgery were not differentiated. Additionally, the AD tool may provide a valuable starting point for manual editing, which will significantly accelerate the contour delineation process. In the present study, it was also found that only the AD accuracy of the spinal cord had a positive correlation with the CT slice number, and the AD accuracy of the bowel bag was not high due to the absence of limits to the delineation range. Moreover, reducing the CT slice number and contour number may improve AD efficiency. In conclusion, it is suggested that the CNN-based AD tool should be used in radiotherapy centers, particularly in those centers with good economic conditions, to improve the accuracy and efficiency of contouring.

Acknowledgements

Not applicable.

Funding

This work was supported by Henan Province Science and Technology Development Project (grant no. 232102310091) and Henan Province Medical Science and Technology Project (grant no. LHGJ20210407).

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

ZL, YH and RS contributed to the conception and design of the study. YH performed the majority of the experiments, as well as the statistical analysis, and drafted the initial manuscript. TQ and MY were responsible for data analysis. ZL revised the final version of the manuscript. YH, RS, TQ, MY and ZL confirm the authenticity of all the raw data. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

The present study was approved by the Ethics Committee of The Second Affiliated Hospital of Zhengzhou University (Zhengzhou, China; approval no. 20210302). Written informed consent was obtained from the patients for the use of their anonymized data in the present study.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD, Fuchs HE and Jemal A: Cancer statistics, 2022. CA Cancer J Clin 72: 7-33, 2022.

2. Cao W, Chen HD, Yu YW, Li N and Chen WQ: Changing profiles of cancer burden worldwide and in China: A secondary analysis of the global cancer statistics 2020. Chin Med J (Engl) 134: 783-791, 2021.
3. Hanna CR, Slevin F, Appelt A, Beavon M, Adams R, Arthur C, Beasley M, Duffton A, Gilbert A, Gollins S, *et al*: Intensity-modulated radiotherapy for rectal cancer in the UK in 2020. Clin Oncol (R Coll Radiol) 33: 214-223, 2021.
4. Chen AM, Chin R, Beron P, Yoshizaki T, Mikaeilian AG and Cao M: Inadequate target volume delineation and local-regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer. Radiother Oncol 123: 412-418, 2017.
5. Walker GV, Awan M, Tao R, Koay EJ, Boehling NS, Grant JD, Sittig DF, Gunn GB, Garden AS, Phan J, *et al*: Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. Radiother Oncol 112: 321-325, 2014.
6. Rezaeijo SM, Jafarpoor Nesheli S, Fatan Serj M and Tahmasebi Birgani MJ: Segmentation of the prostate, its zones, anterior fibromuscular stroma, and urethra on the MRIs and multimodality image fusion using U-Net model. Quant Imaging Med Surg 12: 4786-4804, 2022.
7. Voet PWJ, Dirkx MLP, Teguh DN, Hoogeman MS, Levendag PC and Heijmen BJM: Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. Radiother Oncol 98: 373-377, 2011.
8. Daisne JF and Blumhofer A: Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: A clinical validation. Radiat Oncol 8: 154, 2013.
9. Ma CY, Zhou JY, Xu XT, Guo J, Han MF, Gao YZ, Du H, Stahl JN and Maltz JS: Deep learning-based auto-segmentation of clinical target volumes for radiotherapy treatment of cervical cancer. J Appl Clin Med Phys 23: e13470, 2022.
10. Marin T, Zhuo Y, Lahoud RM, Tian F, Ma X, Xing F, Moteabbed M, Liu X, Grogg K, Shusharina N, *et al*: Deep learning-based GTV contouring modeling inter- and intra-observer variability in sarcomas. Radiother Oncol 167: 269-276, 2022.
11. Casati M, Piffer S, Calusi S, Marrazzo L, Simontacchi G, Di Cataldo V, Greto D, Desideri I, Vernaleone M, Francolini G, *et al*: Clinical validation of an automatic atlas-based segmentation tool for male pelvis CT images. J Appl Clin Med Phys 23: e13507, 2022.
12. Vinod SK, Min M, Jameson MG and Holloway LC: A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol 60: 393-406, 2016.
13. Wang Z, Chang Y, Peng Z, Lv Y, Shi W, Wang F, Pei X and Xu XG: Evaluation of deep learning-based auto-segmentation algorithms for delineating clinical target volume and organs at risk involving data for 125 cervical cancer patients. J Appl Clin Med Phys 21: 272-279, 2020.
14. Young AV, Wortham A, Wernick I, Evans A and Ennis RD: Atlas-based segmentation improves consistency and decreases time required for contouring postoperative endometrial cancer nodal volumes. Int J Radiat Oncol Biol Phys 79: 943-947, 2011.
15. Song Y, Hu J, Wu Q, Xu F, Nie S, Zhao Y, Bai S and Yi Z: Automatic delineation of the clinical target volume and organs at risk by deep learning for rectal cancer postoperative radiotherapy. Radiother Oncol 145: 186-192, 2020.
16. Chen W, Wang C, Zhan W, Jia Y, Ruan F, Qiu L, Yang S and Li Y: A comparative study of auto-contouring softwares in delineation of organs at risk in lung cancer and rectal cancer. Sci Rep 11: 23002, 2021.
17. Piqeur F, Hupkens BJP, Nordkamp S, Witte MG, Meijnen P, Ceha HM, Berbee M, Dieters M, Heyman S, Valdman A, *et al*: Development of a consensus-based delineation guideline for locally recurrent rectal cancer. Radiother Oncol 177: 214-221, 2022.
18. Mackay K, Bernstein D, Glocker B, Kamnitsas K and Taylor A: A review of the metrics used to assess auto-contouring systems in radiotherapy. Clin Oncol (R Coll Radiol) 35: 354-369, 2023.
19. Li Y, Wu W, Sun Y, Yu D, Zhang Y, Wang L, Wang Y, Zhang X and Lu Y: The clinical evaluation of atlas-based auto-segmentation for automatic contouring during cervical cancer radiotherapy. Front Oncol 12: 945053, 2022.
20. Men K, Dai J and 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 44: 6377-6389, 2017.
21. Ronneberger O, Fischer P and Brox T (eds): U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III. Springer International Publishing, Cham, pp234-241, 2015.
22. Shen G, Jin X, Sun C and Li Q: Artificial intelligence radiotherapy planning: Automatic segmentation of human organs in CT images based on a modified convolutional neural network. Front Public Health 10: 813135, 2022.
23. Luan S, Xue X, Ding Y, Wei W and Zhu B: Adaptive attention convolutional neural network for liver tumor segmentation. Front Oncol 11: 680807, 2021.
24. Wu Y, Kang K, Han C, Wang S, Chen Q, Chen Y, Zhang F and Liu Z: A blind randomized validated convolutional neural network for auto-segmentation of clinical target volume in rectal cancer patients receiving neoadjuvant radiotherapy. Cancer Med 11: 166-175, 2022.
25. Liu C, Gardner SJ, Wen N, Elshaikh MA, Siddiqui F, Movsas B and Chetty IJ: Automatic segmentation of the prostate on CT images using deep neural networks (DNN). Int J Radiat Oncol Biol Phys 104: 924-932, 2019.
26. Liu Z, Liu X, Xiao B, Wang S, Miao Z, Sun Y and Zhang F: Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network. Phys Med 69: 184-191, 2020.
27. To MNN, Vu DQ, Turkbey B, Choyke PL and Kwak JT: Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging. Int J Comput Assist Radiol Surg 13: 1687-1696, 2018.
28. Breto A, Zavala-Romero O, Asher D, Baikovitz J, Ford J, Stoyanova R and Portelance L: A deep learning pipeline for per-fraction automatic segmentation of GTV and OAR in cervical cancer. Int J Radiat Oncol Biol Phys 105 (Suppl 1): S202, 2019.
29. Bi N, Wang J, Zhang T, Chen X, Xia W, Miao J, Xu K, Wu L, Fan Q, Wang L, *et al*: Deep learning improved clinical target volume contouring quality and efficiency for postoperative radiation therapy in non-small cell lung cancer. Front Oncol 9: 1192, 2019.
30. Sha X, Wang H, Sha H, Xie L, Zhou Q, Zhang W and Yin Y: Clinical target volume and organs at risk segmentation for rectal cancer radiotherapy using the Flex U-Net network. Front Oncol 13: 1172424, 2023.
31. Li J, Song Y, Wu Y, Liang L, Li G and Bai S: Clinical evaluation on automatic segmentation results of convolutional neural networks in rectal cancer radiotherapy. Front Oncol 13: 1158315, 2023.
32. Chen Y, Li J, Xiao H, Jin X, Yan S and Feng J: Dual path networks. Adv Neural Inf Process Syst 30, 2017.
33. Liu Z, Liu X, Guan H, Zhen H, Sun Y, Chen Q, Chen Y, Wang S and Qiu J: Development and validation of a deep learning algorithm for auto-delineation of clinical target volume and organs at risk in cervical cancer radiotherapy. Radiother Oncol 153: 172-179, 2020.
34. Liu Z, Liu F, Chen W, Liu X, Hou X, Shen J, Guan H, Zhen H, Wang S, Chen Q, *et al*: Automatic segmentation of clinical target volumes for post-modified radical mastectomy radiotherapy using convolutional neural networks. Front Oncol 10: 581347, 2021.
35. Valentini V, Gambacorta MA, Barbaro B, Chiloiro G, Coco C, Das P, Fanfani F, Joye I, Kachnic L, Maingon P, *et al*: International consensus guidelines on clinical target volume delineation in rectal cancer. Radiother Oncol 120: 195-201, 2016.
36. Gay HA, Barthold HJ, O'Meara E, Bosch WR, El Naqa I, Al-Lozi R, Rosenthal SA, Lawton C, Lee WR, Sandler H, *et al*: Pelvic normal tissue contouring guidelines for radiation therapy: A radiation therapy oncology group consensus panel atlas. Int J Radiat Oncol Biol Phys 83: e353-e362, 2012.
37. Yeghiazaryan V and Voiculescu I: Family of boundary overlap metrics for the evaluation of medical image segmentation. J Med Imaging (Bellingham) 5: 015006, 2018.
38. Geets X, Daisne JF, Arcangeli S, Coche E, De Poel M, Duprez T, Nardella G and Grégoire V: Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI. Radiother Oncol 77: 25-31, 2005.
39. Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, Van Herk M, Lee A, Maingon P, Nutting C, *et al*: CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG oncology and TROG consensus guidelines. Radiother Oncol 117: 83-90, 2015.
40. Rhee DJ, Jhingran A, Rigaud B, Netherton T, Cardenas CE, Zhang L, Vedam S, Kry S, Brock KK, Shaw W, *et al*: Automatic contouring system for cervical cancer using convolutional neural networks. Med Phys 47: 5648-5658, 2020.

41. Zabihollahy F, Viswanathan AN, Schmidt EJ and Lee J: Fully automated segmentation of clinical target volume in cervical cancer from magnetic resonance imaging with convolutional neural network. J Appl Clin Med Phys 23: e13725, 2022.
42. Mohammadi R, Shokatian I, Salehi M, Arabi H, Shiri I and Zaidi H: Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer. Radiother Oncol 159: 231-240, 2021.
43. Ju Z, Guo W, Gu S, Zhou J, Yang W, Cong X, Dai X, Quan H, Liu J, Qu B and Liu G: CT based automatic clinical target volume delineation using a dense-fully connected convolution network for cervical Cancer radiation therapy. BMC Cancer 21: 243, 2021.
44. Chan JW, Kearney V, Haaf S, Wu S, Bogdanov M, Reddick M, Dixit N, Sudhyadhom A, Chen J, Yom SS and Solberg TD: A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. Med Phys 46: 2204-2213, 2019.
45. Ginn JS, Gay HA, Hilliard J, Shah J, Mistry N, Möhler C, Hugo GD and Hao Y: A clinical and time savings evaluation of a deep learning automatic contouring algorithm. Med Dosim 48: 55-60, 2023.
46. Åström LM, Behrens CP, Calmels L, Sjöström D, Geertsen P, Mouritsen LS, Serup-Hansen E, Lindberg H and Sibolt P: Online adaptive radiotherapy of urinary bladder cancer with full re-optimization to the anatomy of the day: Initial experience and dosimetric benefits. Radiother Oncol 171: 37-42, 2022.
47. D'Aviero A, Re A, Catucci F, Piccari D, Votta C, Piro D, Piras A, Di Dio C, Iezzi M, Preziosi F, *et al*: Clinical validation of a deep-learning segmentation software in head and neck: An early analysis in a developing radiation oncology center. Int J Environ Res Public Health 19: 9057, 2022.
48. Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, van Elmpt W and Dekker A: Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol 126: 312-317, 2018.
49. Hu Y, Nguyen H, Smith C, Chen T, Byrne M, Archibald-Heeren B, Rijken J and Aland T: Clinical assessment of a novel machine-learning automated contouring tool for radiotherapy planning. J Appl Clin Med Phys 24: e13949, 2023.