European Radiology (2020) 30:1297–1305
https://doi.org/10.1007/s00330-019-06467-3

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer

Yu-Chun Lin[1,2] · Chia-Hung Lin[1] · Hsin-Ying Lu[1,2,3] · Hsin-Ju Chiang[1,2,3] · Ho-Kai Wang[1] · Yu-Ting Huang[1,4] · Shu-Hang Ng[1,2] · Ji-Hong Hong[2,4,5] · Tzu-Chen Yen[4,6] · Chyong-Huey Lai[2,7] · Gigin Lin[1,2,3,4]

Received: 22 May 2019 / Revised: 20 August 2019 / Accepted: 19 September 2019 / Published online: 11 November 2019

Abstract

Objective To develop and evaluate the performance of U-Net for fully automated localization and segmentation of cervical tumors in magnetic resonance (MR) images and the robustness of extracting apparent diffusion coefficient (ADC) radiomics features.

Methods This retrospective study involved analysis of MR images from 169 patients with cervical cancer stage IB–IVA captured; among them, diffusion-weighted (DW) images from 144 patients were used for training, and another 25 patients were recruited for testing. A U-Net convolutional network was developed to perform automated tumor segmentation. The manually delineated tumor region was used as the ground truth for comparison. Segmentation performance was assessed for various combinations of input sources for training. ADC radiomics were extracted and assessed using Pearson correlation. The reproducibility of the training was also assessed.

Results Combining b0, b1000, and ADC images as a triple-channel input exhibited the highest learning efficacy in the training phase and had the highest accuracy in the testing dataset, with a dice coefficient of 0.82, sensitivity 0.89, and a positive predicted value 0.92. The first-order ADC radiomics parameters were significantly correlated between the manually contoured and fully automated segmentation methods ($p < 0.05$). Reproducibility between the first and second training iterations was high for the first-order radiomics parameters (intraclass correlation coefficient = 0.70–0.99).

Conclusion U-Net-based deep learning can perform accurate localization and segmentation of cervical cancer in DW MR images. First-order radiomics features extracted from whole tumor volume demonstrate the potential robustness for longitudinal monitoring of tumor responses in broad clinical settings.

Summary U-Net-based deep learning can perform accurate localization and segmentation of cervical cancer in DW MR images.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06467-3) contains supplementary material, which is available to authorized users.

✉ Gigin Lin
giginlin@cgmh.org.tw

[1] Department of Medical Imaging and Intervention, Chang Gung Memorial Hospital at Linkou, 5 Fuhsing St., Guishan, Taoyuan, Taiwan 33382

[2] Imaging Core Laboratory, Institute for Radiological Research, Chang Gung Memorial Hospital at Linkou and Chang Gung University, 5 Fuhsing St., Guishan, Taoyuan, Taiwan 33382

[3] Clinical Metabolomics Core Laboratory, Chang Gung Memorial Hospital at Linkou, 5 Fuhsing St., Guishan, Taoyuan, Taiwan 33382

[4] Department of Obstetrics and Gynecology and Gynecologic Cancer Research Center, Chang Gung Memorial Hospital at Linkou and Chang Gung University, 5 Fuhsing St., Guishan, Taoyuan, Taiwan 33382

[5] Department of Radiation Oncology, Chang Gung Memorial Hospital at Linkou and Chang Gung University, 5 Fuhsing St., Guishan, Taoyuan, Taiwan 33382

[6] Department of Nuclear Medicine and Center for Advanced Molecular Imaging and Translation, Chang Gung Memorial Hospital and Chang Gung University, 5 Fuhsing St., Guishan, Taoyuan, Taiwan 33382

[7] Clinical Trial Center, Chang Gung Memorial Hospital at Linkou and Chang Gung University, 5 Fuhsing St., Guishan, Taoyuan, Taiwan 33382

Key Points

- *U-Net-based deep learning can perform accurate fully automated localization and segmentation of cervical cancer in diffusion-weighted MR images.*
- *Combining b0, b1000, and apparent diffusion coefficient (ADC) images exhibited the highest accuracy in fully automated localization.*
- *First-order radiomics feature extraction from whole tumor volume was robust and could thus potentially be used for longitudinal monitoring of treatment responses.*

Keywords Apparent diffusion coefficient · Diffusion-weighted imaging · Uterine cervical neoplasm · Deep learning · Radiomics

Abbreviations

ADC	Apparent diffusion coefficient
DSC	Dice similarity coefficient
DW	Diffusion-weighted
MR	Magnetic resonance
PPV	Positive predictive value
ROI	Region of interest
T2W	T2-weighted

Introduction

Magnetic resonance (MR) imaging is standard practice for care of cervical cancers of International Federation of Gynecology and Obstetrics (FIGO) stage IB1 or higher [1, 2]. MR imaging is useful for defining the extent of tumor involvement, including the depth of cervical stromal or parametrial invasion [1]. Tumor contour and volume are important for radiation therapy planning, and MR is considered the gold standard for the evaluation of local tumor extension [3]. However, discrepancies in gross tumor volume between manual measurements were observed following the recommendations from Gynaecological Groupe Européen de Curiethérapie and the European Society for Radiotherapy & Oncology working group (GYN GEC-ESTRO) [4]. In addition to tumor localization, diffusion-weighted (DW) MR imaging provides quantitative estimates, namely the apparent diffusion coefficient (ADC), which reflects the cellular microenvironment of a tumor to assist treatment decision-making [5]. Radiomics features derived from the ADC of a volumetric analysis of a whole tumor showed promise for tumor grading [6] and predicting the recurrence of cervical cancer [7–9]. However, radiomics approach requires the development of a strategy for high-throughput extraction of quantitative features to enable the conversion of images into mineable data in a precise and timely fashion [10].

To extract the quantitative features of a whole tumor, an efficient tumor segmentation tool is critical. Manual segmentation is time-consuming and prone to interobserver variation. Torheim et al [11] employed a machine-learning method and Fisher's linear discriminant analysis to yield a Dice similarity coefficient (DSC) of 0.44 for tumor segmentation. Deep-learning algorithms based on convolutional neural networks (CNNs) have emerged to demonstrate potential for lesion detection and segmentation. Results have revealed that CNNs outperform traditional machine-learning techniques for brain [12–14], rectal [15], and prostate cancer [16] in tumor segmentation. Recently, the U-Net architecture was proposed, which requires fewer network parameters and accelerates the training process [17, 18]. Recent studies have shown that U-Net outperforms previous methods for semantic segmentation of a natural scene image in terms of performance and speed [13, 19]. To the best of our knowledge, no studies have examined the performance of deep-learning algorithms for fully automated tumor localization, volume segmentation, and the reliability of the extracted radiomics parameters of cervical cancer.

The aim of this study was therefore to develop and evaluate the performance of U-Net for fully automated localization and segmentation of cervical tumors in MR imaging, as well as the robustness of extracting ADC radiomics features.

Materials and methods

Patients

We retrospectively analyzed two consecutively collected cohorts from January 2011 to December 2016 (IRB 97-2366B and IRB 102-0620A3). Our Institutional Review Board approved this study, and informed consent was obtained from all participants. The study was conducted in a tertiary referral center with a dedicated interdisciplinary gynecological oncology team. The inclusion criteria were (a) newly diagnosed cervical cancer of FIGO stage IB–IVA and (b) willingness to undergo chemoradiation as standard treatment. The exclusion criteria were noncompliance with treatment or inaccessibility for follow-up, contraindications to MR scanning, and mental health–related inability to cooperate. One hundred thirty-four of the 169 patients have been previously reported to study the prognostic values of whole-tumor ADC values from manually delineated tumor regions [9]. In this study, we developed a deep-learning algorithm based on the reported manually delineated tumor regions and tested the algorithm on an independent dataset. From a consecutive cohort of 215 patients, we excluded 30 patients who had no visible tumors and 16 patients who were susceptible to artifacts in DW imaging.

Thus, 169 patients were eligible for the final analysis. During the model development, 144 patients were randomized into the training dataset, and another independent testing dataset comprising 25 patients, to provide an unbiased evaluation of the final model fit of the training dataset.

MR image acquisition

MR images were obtained using two 3-T MR scanners, namely a Trio TIM ($n = 96$) and Skyra ($n = 73$) (Siemens Healthineers). All patients underwent the standard MR protocol for the pelvis at our institute, including T2-weighted, T1-weighted, DW images, and contrast-enhanced T1-weighted series acquired in the sagittal and axial oblique planes perpendicular to the cervical canal. The imaging parameters of the DW images were as follows: single-shot echo-planar imaging with fat suppression; repetition time = 3600–8400 ms; echo time = 64–84 ms; section thickness = 4 mm; gap = 1 mm; matrix = 192 × 192 or 256 × 256; field of view = 20 × 20 cm^2, *b*-values = 0 and 1000 s/mm^2. ADC maps were generated automatically by an MR scanner by using a monoexponential decay model with *b*-values of 0 and 1000 s/mm^2. The sagittal DW images of b0, b1000, and the ADC maps of each slice of each patient were used as input sources for subsequent training. For each patient, 15–20 slice sections were acquired to cover the entire tumor. The contrast-enhanced series was acquired at about 120–180 s equilibrium phase after intravenous injection of 0.1 mmol/kg body weight of contrast medium but not used for analysis in this study.

Image annotation

An in-house graphical user interface was developed in MATLAB (Mathworks, Natick) to manually delineate tumor contours. Anonymous data were exported offline, and the first reader (Y. T. H., a gynecologic radiologist with 9 years' experience) drew regions of interest (ROIs) around the tumor on each section of the ADC maps with respect to the high-*b*-value DW and T2-weighted images to delineate the whole tumor volume. The second reader (G.L., a gynecologic radiologist with 11 years' experience) independently verified the ROIs. Both readers were blind to the clinical outcomes. Care was taken to avoid each ROI contaminating the adjacent normal cervical stroma or vascular structures and contamination by areas of fluid or nabothian cysts in the cervix. The labeled ROIs were regarded as the ground truth for the subsequent training process.

U-Net model development in the training dataset

Figure 1 illustrates the U-Net architecture based on a fully convolutional network [18]. The architecture combines a contracting (downsampling) path to capture context and a symmetrical expending (upsampling) path to enable precise localization [17]. The detail of U-Net architecture is described in the Supplementary Material. The matrices of all images were resized to 256 × 256. All images were standardized such that the intensity was distributed to have mean = 0 and standard deviation = 1 [15]. Data augmentation was performed on each training image set such that the training model could be robust against the degree of enlargement, rotation, and parallel shift. The central part (200 × 200 pixels) of each image was cropped; rotated 90°, 180°, and 270°; and shifted by 28 pixels in four directions (up, down, left, and right). Through those processes, 24 additional image sets were generated from one image set, yielding 61,320 images in each set.

To evaluate the effect of input image type on the training efficacy, we evaluated seven combinations from three image

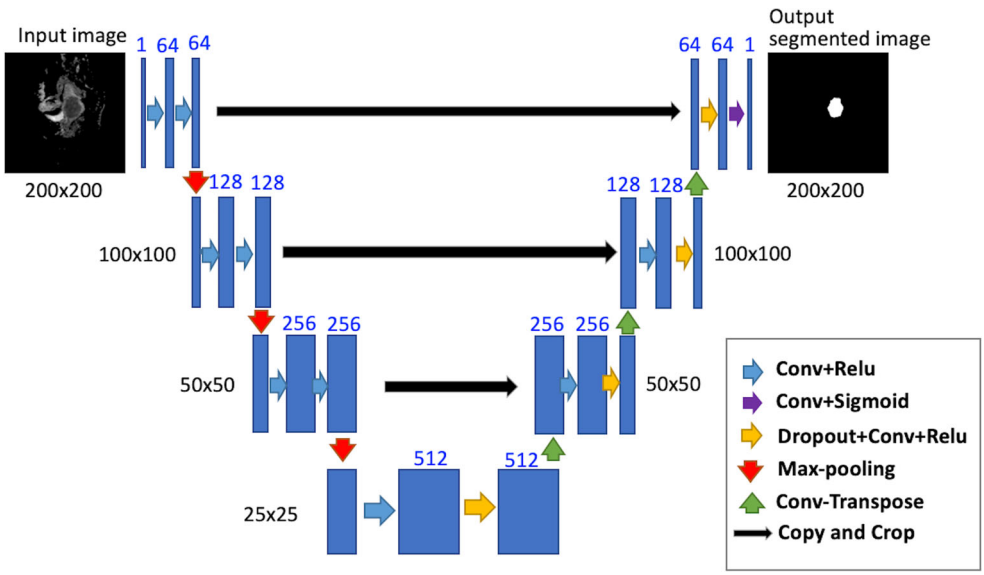

Fig. 1 U-Net architecture with input image matrix size of 200 × 200. Each blue box corresponds to a multichannel feature map. The black text at the left of the box denotes the matrix size. The blue text on top of the box denotes the number of filters. The arrows of different colors indicate different operations

sets (b0, b1000, and ADC) as input sources for training, described as follows: (a) single channel—b0, b1000, and ADC are independent; (b) dual channel—b0 + b1000, b0 + ADC, and b1000 + ADC; (c) triple channel—b0 + b1000 + ADC. The training was performed on a personal computer equipped with GeForce GTX1080Ti (NVIDIA) graphics processing unit. Python programming language 3.5.4 with Keras 2.1.4 and TensorFlow 1.5.0 for neural networks were used for the training. The networks were trained using the stochastic gradient descent Adam Optimizer method [20], with the DSC serving as an accuracy measure of the segmentation procedure, and minus DSC serving as a loss function that was backpropagated through the CNN. The batch size was set to 35. The learning rate was set to 10^{-5}. The model was trained for 100 epochs for each dataset combination. Segmentation was generated by setting the threshold of the probability map to $p = 0.5$. To test the reproducibility of the predicted ROI using the training model, we repeated the entire training procedure with the identical parameters.

ADC radiomics extraction

To assess the robustness of the predicted ROIs, we examined the radiomics parameters of the ADC values of the tumors. The calculation of MR radiomics was performed using the MR Radiomics Platform (www.ym.edu.tw/~cflu/MRP_MLinglioma.html), which was constructed in the MATLAB programming environment [21]. In the extracted 1073 radiomics features, we analyzed the 51 texture features in our data because the higher order features (local binary pattern, and scale invariant feature transform) did not show significant result in our preliminary test. The final ADC texture features employed in this study included intensity-based features (first-order statistics), shape- and size-based features, and textural features (gray-level co-occurrence matrix-based features, and gray-level run-length matrix-based features).

Performance evaluation in the testing dataset

After building the models, we examined their performance in predicting the ROIs of the tumors in the independent testing dataset. The segmentation accuracy was evaluated using (a) the DSC, a statistical measure of spatial overlap defined as $DSC = \frac{2TP}{FP+2TP+FN}$, where TP, FP, and FN are the numbers of true positive, false positive, and false negative detections, respectively; (b) the positive predictive value (PPV), a measure of the numbers of FP and TP detections defined as $PPV = \frac{TP}{TP+FP}$; and (c) sensitivity, which evaluates the numbers of TP and FN detections and is defined as $Sensitivity = \frac{TP}{TP+FN}$. We provided additional 12 metrics to evaluate concordance between labeled and predicted ROIs in our testing dataset [22]. The radiomics parameters of tumor ADCs extracted by the proposed algorithm were correlated with those that had been manually labeled using Spearman's correlation. The Bland-Altman analysis was performed for the first-order features extracted by labeled and predicted ROIs. The reproducibility of the training model was assessed using the intraclass correlation coefficient (ICC) between the radiomics parameters of the ADCs of ROIs extracted from the first and second training algorithms. The stability of the model was assess by a fivefold cross-validation using analysis of variance (ANOVA) on DSCs between labeled and predicted ROI by each trained model. The differences in DSCs with regard to tumor size and tumor histology were assessed using Mann-Whitney *U* test. Univariate and multivariate logistic regression with stepwise procedure was used to identify the independent factor accounting for the DSC below 0.9. The data were analyzed using the Statistical Package for Social Sciences, version 11 (SPSS, IBM), MedCalc for Windows, Version 9.2.0.0 (MedCalc Software), or R Package for Statistical Computing (www.r-project.org).

Results

Patient characteristics

A total of 169 patients were eligible for the final analysis. The patients were aged from 23 to 91 years (median, 52 years). Table 1 lists the clinical and demographic features for the

Table 1 Participant demographics

Variable	Training	Testing	*p* value
n	144	25	
Age (year)*	53.27 (14.50)	53.68 (10.72)	0.893
Histopathology			1.000
Squamous cell	119 (82.6)	21 (84.0)	
Non-squamous cell	25 (17.4)	4 (16.0)	
Grade of differentiation			0.935
Well/moderate	68 (47.2)	11 (44.0)	
Poorly	76 (52.8)	14 (56.0)	
T stage			0.821
1	47 (32.6)	7 (28.0)	
2, 3, 4	97 (67.4)	18 (72.0)	
N stage			0.569
0	87 (60.4)	13 (52.0)	
1	57 (39.6)	12 (48.0)	
M stage			0.708
0	119 (83.0)	22 (88.0)	
1	25 (17.0)	3 (12.0)	

Data in parentheses are percentages

* represents mean (standard deviation)

training and testing datasets. The training and testing datasets demonstrated no statistically significant differences in clinical or demographic characteristics. The histopathology types comprised squamous cell carcinoma ($n = 160$, 95%), adenocarcinoma ($n = 2$, 1%), adenosquamous carcinoma ($n = 4$, 2%), and small-cell carcinoma ($n = 3$, 2%). Most of the tumors were poorly differentiated ($n = 90$, 53%) or moderately differentiated ($n = 69$, 41%). Tumor size ranged from 0.21 to 83.78 cm^3 (median, 9.51 cm^3). The time interval and any clinical interventions between MR study and histopathology were 0–30 days (median, 13 days). No adverse events due to performance of the MR study were observed.

U-Net for fully automated tumor localization and segmentation in the training phase

The segmentation performance varied when different combinations of source images were used for training (Table 2). All data from the b0, b1000, and ADC maps were highly correlated. During the training phase, the use of the triple-channel input (ADC + b0 + b1000) exhibited the highest learning efficacy, reaching a plateau of 0.95 (95% confidence interval [CI], 0.95–0.96). The single channel of b0 had the lowest learning efficacy, reaching a plateau for accuracy of 0.89 (95% CI, 0.88–0.89) (Fig. 2).

Evaluation of segmentation performance in the independent testing dataset

Figure 3 shows an example of fully automated tumor segmentation in the testing dataset. The segmentation accuracy is summarized in Table 2. Among the combinations of source images that served as input sources, the triple-channel model (ADC + b0 + b1000) exhibited the highest DSC of 0.82 (95% CI, 0.79–0.85), the highest sensitivity of 0.89 (95% CI, 0.86–0.91), and a PPV of 0.92 (95% CI, 0.90–0.93). The triple-channel model also outperformed the other models regarding similarity and distance-based metrics (Supplementary Table 1). The difference

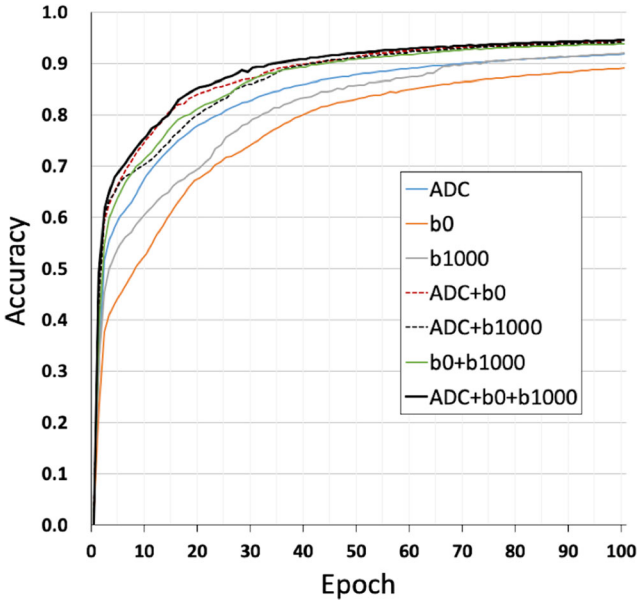

Fig. 2 Plot of training accuracy against epoch produced using various combinations of image sources

of DSCs among the fivefold partition groups was not statistically significant ($p = 0.34$) (Supplementary Table 2).

Insights into error analysis from the testing dataset

Based on stepwise selection procedure, it showed that tumor size ≥4 cm was an independent predictor for the DSC < 0.9 (odds ratio, 4.30; 95% confidence interval 1.74–11.39; $p = 0.002$). We found that the DSC was statistically significant at tumor diameters of < 4 cm compared with those of ≥ 4 cm (DSC 0.58 vs. 0.87, $p = 0.039$). Figure 4 plots the relationship between the DSC and patient-based tumor volume in the testing dataset. The DSC was significantly correlated with the logarithm of tumor volume ($r = 0.82$, $p < 0.001$). No significant correlation was observed between the DSC and patient age ($p = 0.337$). Notably, the five patients with very low DSCs of < 0.30 demonstrated low tumor volumes of < 3.42 cm^3. We observed that the DSC was

Table 2 Accuracy of predicted ROIs generated using various combinations of source images from the training and testing datasets

Input image	Training		Testing					
	DSC		DSC		Sensitivity		PPV	
ADC	0.90	(0.89, 0.91)	0.77	(0.74, 0.81)	0.62	(0.58, 0.66)	0.91	(0.89, 0.93)
b0	0.89	(0.88, 0.89)	0.72	(0.69, 0.76)	0.44	(0.40, 0.48)	0.88	(0.85, 0.91)
b1000	0.92	(0.91, 0.93)	0.78	(0.75, 0.82)	0.71	(0.66, 0.75)	0.88	(0.85, 0.91)
ADC + b0	0.95	(0.94, 0.95)	0.80	(0.77, 0.83)	0.71	(0.68, 0.75)	0.89	(0.87, 0.92)
ADC + b1000	0.94	(0.94, 0.94)	0.79	(0.76, 0.83)	0.88	(0.85, 0.90)	0.89	(0.86, 0.91)
b0 + b1000	0.94	(0.94, 0.95)	0.79	(0.76, 0.82)	0.74	(0.70, 0.77)	0.89	(0.86, 0.91)
ADC + b0 + b1000	0.95	(0.95, 0.96)	0.82	(0.79, 0.85)	0.89	(0.86, 0.91)	0.92	(0.90, 0.93)

ADC = apparent diffusion coefficient, *b* = *b*-value, *DSC* = dice similarity coefficient, *PPV* = positive predictive value

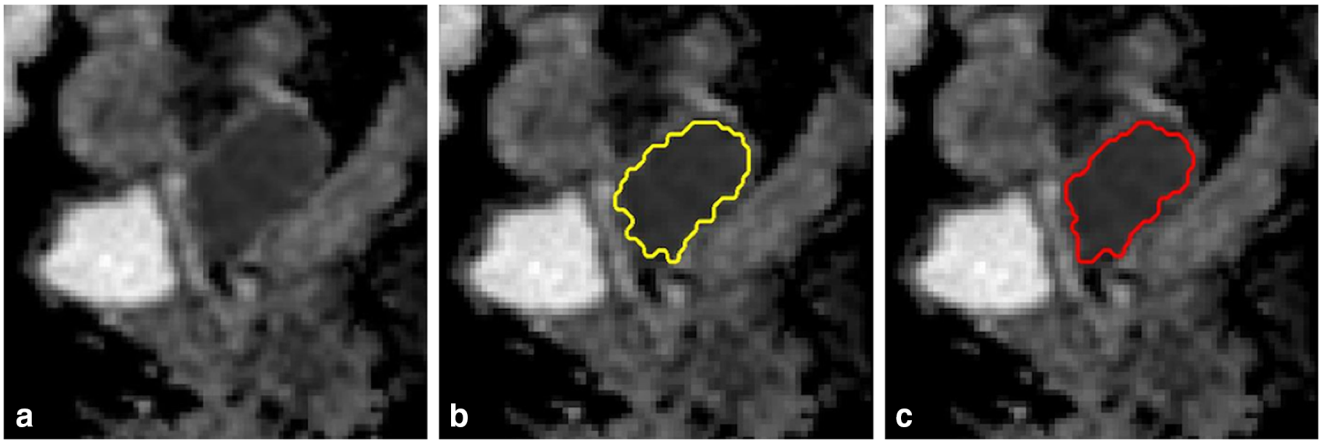

Fig. 3 A 50-year-old woman with cervical cancer. The tumor exhibited hypointensity on the ADC map (**a**). The ROI of the tumor was labeled manually (yellow contour) and was overlaid with the predicted ROI (red contour) generated by the trained triple-channel model (**b**, **c**). The predicted ROI was segmented at the corresponding tumor location with DSC = 0.928

statistically significantly lower in non-squamous cell types compared with squamous cell carcinomas (DSC 0.44 vs. 0.80, $p = 0.025$), whereas tumor grade was not an influential factor ($p = 0.618$). The DSC was statistically significantly lower in tumor stage T1 compared with stages T2 and above (DSC 0.41 vs. 0.87, $p = 0.018$) and in stage N0 compared with stage N1 (DSC 0.62 vs. 0.87, $p = 0.031$) but not in stage M0 compared with stage M1 ($p = 0.337$).

Robustness of extracting ADC radiomics features

The effects of the predicted ROIs on the ADC radiomics parameters are illustrated in Fig. 5, Table 3, and Supplementary Table 3. Tumor volumes were significantly correlated between the manually contoured and fully automated segmentation methods ($r = 0.94$, $p < 0.001$), as were the first-order ADC radiomics parameters ($r = 0.432–0.986$, $p < 0.05$). The reproducibility scores of the radiomics parameters between the first and second training runs were high for the first-order statistics (ICC = 0.70–0.99, $p < 0.05$). However, the correlation and reproducibility were less significant for the high-order features, namely the textural features (gray-level co-occurrence matrix-based features and gray-level run-length matrix-based features; ICC = − 0.12 to 0.38). The Bland-Altman plots demonstrated that the first-order features mostly distributed within the mean ± 1.96 standard deviation (Supplementary Fig. 1), implying that the labeled and predicted ROIs were in agreement without systematic errors.

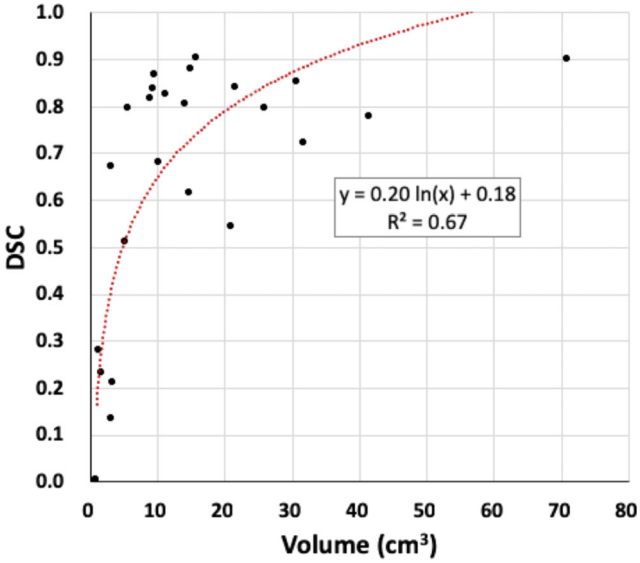

Fig. 4 Scatter plot of the DSC and tumor volumes of patients in the testing dataset. A logarithmic regression was plot showing that the DSC was correlated with the logarithm of tumor volume

Discussion

We exploited the potential of deep neural networks for fully automated segmentation of DW MR imaging for cervical cancer and examined the robustness of extracting ADC radiomics features. By using the fully convolutional network of U-Net architecture with DW MR images, cervical tumors could undergo fully automated segmentation in standard DICOM images without manual interference. All metrics for model evaluation, namely the DSC, sensitivity, and the PPV, demonstrated more than twofold improvement compared with the previously reported automated segmentation methods of machine learning and Fisher's linear discriminant analysis [11]. Related studies on pelvic MR imaging have used a patch-based approach, but operating on image patches is computationally time-consuming, and performance is influenced manually [12, 15]. Trebeschi et al [15] utilized patch-based deep learning to segment rectal cancer on T2-weighted and DW images and reported a DSC of 0.70. However, the performance of the patch-based approach was influenced by the size and location of the manually selected patches. The U-Net architecture is

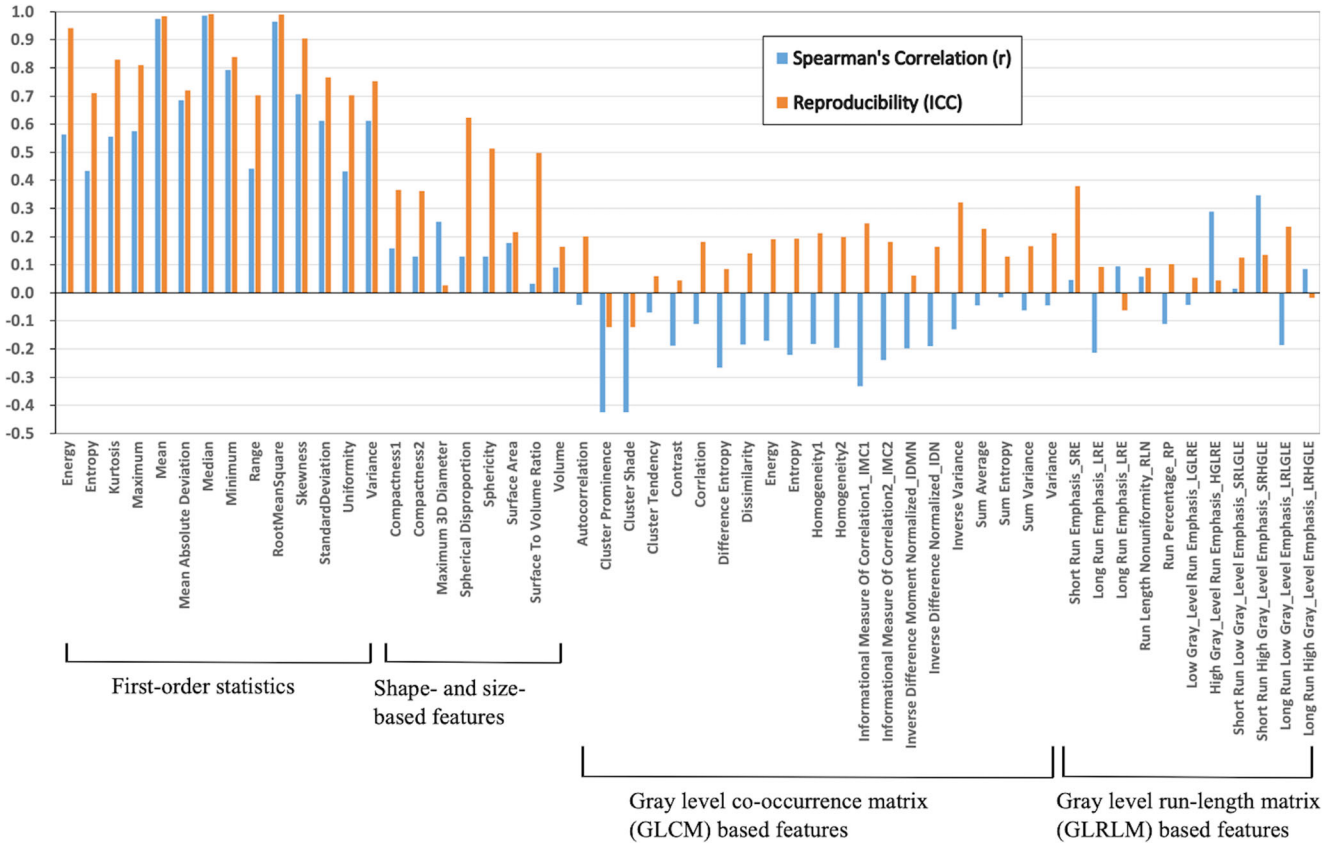

Fig. 5 Correlations and reproducibilities of ADC radiomics parameters of predicted ROIs using the triple-channel training model

based on a fully convolutional network, which extracts significant features of images to generate a predicted map [17]. With the input of single-channel T2-weighted images, Wang et al [19] obtained a DSC of 0.74 for segmenting rectal cancers using the U-Net architecture. Furthermore, Dolz et al [23] obtained a DSC of 0.69 for segmenting bladder cancers using a dilated convolutional network. Our study demonstrated the triple-channel model further improving the segmentation of

Table 3 Robustness of ADC radiomics parameters (first-order statistics) of predicted ROIs obtained using the triple-channel training model

	Correlation between manual and predicted values		Reproducibility between 1st and 2nd predicted values		
Variable	*r*	*p* value	ICC*		*p* value
Energy	0.564	0.011	0.94	(0.86, 0.98)	< 0.001
Entropy	0.445	0.048	0.71	(0.40, 0.89)	0.022
Kurtosis	0.555	0.012	0.83	(0.62, 0.93)	< 0.001
Maximum	0.574	0.009	0.81	(0.59, 0.92)	< 0.001
Mean	0.974	< 0.001	0.98	(0.96, 0.99)	< 0.001
Mean absolute deviation	0.686	< 0.001	0.71	(0.41, 0.88)	< 0.001
Median	0.986	< 0.001	0.99	(0.98, 1.00)	< 0.001
Minimum	0.792	< 0.001	0.84	(0.64, 0.93)	< 0.001
Range	0.441	0.048	0.70	(0.40, 0.87)	< 0.001
Root mean square	0.964	< 0.001	0.99	(0.98, 1.00)	< 0.001
Skewness	0.707	< 0.001	0.90	(0.78, 0.96)	< 0.001
Standard Deviation	0.612	< 0.001	0.77	(0.51, 0.90)	< 0.001
Uniformity	0.432	0.049	0.70	(0.39, 0.88)	0.042
Variance	0.612	0.005	0.75	(0.48, 0.89)	< 0.001

*Data are means with 95% confidence intervals in parentheses

cervical cancers regarding the regional overlap and distance along the shapes, with a final DSC of 0.82.

DW MR imaging provides the widest available tissue contrast range between a tumor and noncancerous tissue compared with T1-weighted, T2-weighted, and contrast-enhanced MR images [24], and thus it was employed in the current study. DW MR imaging acquired b0 and b1000 images simultaneously during scanning and generated ADC maps, thereby preventing bias due to motion-related misregistration across MR pulse sequences. Our approach mimicked the clinical scenario of a radiologist delineating tumor ROIs by referencing multiparametric MR images simultaneously. Therefore, we propose that including all available DW and ADC images for training is beneficial for deep learning used for tumor segmentation.

Fully automated localization and segmentation of cervical cancer based on the present study could assist the radiation therapy planning and minimize discrepancies in gross tumor volume. The fully automated strategy could also accelerate the high-throughput extraction of quantitative features to enable the conversion of images into mineable data in a precise and timely fashion. We observed that the ADC radiomics parameters were significantly correlated between the manually contoured and fully automated segmentation methods for first-order radiomics features. This finding implies that the algorithm based on the U-Net architecture may instantaneously accelerate the extraction of radiomics features through fully automated contouring of cervical cancers, although care should be taken if using this algorithm for tumor detection because our results revealed suboptimal DSCs clustered in patients with low tumor volumes or non-squamous cell types. Nonetheless, the fully automated localization and segmentation approach may be the key to integrating imaging biomarkers into daily clinical routines and could be of particular benefit to oncologists for longitudinal monitoring of tumor responses to treatment [25].

Limitations, future directions, and recommendations

Some limitations of this study require further research. First, we focused on only DW MR imaging; multiparametric images may have demonstrated superior performance. Because organ motion leads to temporal variation during the scanning of various MR pulse sequences, we were unable to confidently include the labeled ROIs from T1-weighted, T2-weighted, and contrast-enhanced images for rigid registration. We believe that the segmentation accuracy would increase further if more images were used following a robust nonrigid registration among various parametric MR images. Second, the U-Net architecture used in this study is a basic form resulting from a recent modification for higher accuracy in segmenting prostate tumors [26]. However, the performance of the architecture warrants further verification and optimization for cervical cancer.

Conclusion

In conclusion, our results demonstrate that U-Net-based deep learning can perform accurate localization and segmentation of cervical cancer in DW MR images. First-order radiomics features extracted from the model-based tumor segmentation were robust and reproducible, whereas the reproducibility of higher order features was limited.

Acknowledgments The authors acknowledge the assistance provided by the Cancer Center and Clinical Trial Center (Statistician Dr. Lan-Yan Yang), Chang Gung Memorial Hospital, Linkou, Taiwan, which was founded by the Ministry of Health and Welfare of Taiwan MOHW106-TDU-B-212-113005.

Funding information This study was supported by the Chang Gung Medical Foundation grant nos. CPRPG3G0021-3, CIRPG3H0011, CIRPG3D0163, and CMRPG3I0141; Ministry of Science and Technology (Taiwan) grant no. MOST 106-2314-B-182A-016-MY2; and Chang Gung IRB97-2366B, IRB102-0620A3, IRB201800412B0, and IRB201702204B0.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Gigin Lin, MD, PhD.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry Dr. Lan-Yan Yang kindly provided statistical advice for this manuscript.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- performed at one institution

References

1. Sala E, Rockall AG, Freeman SJ, Mitchell DG, Reinhold C (2013) The added role of MR imaging in treatment stratification of patients with gynecologic malignancies: what the radiologist needs to know. Radiology 266:717–740
2. Bhatla N, Berek JS, Cuello Fredes M et al (2019) Revised FIGO staging for carcinoma of the cervix uteri. Int J Gynaecol Obstet 145: 129–135
3. Ma DJ, Zhu JM, Grigsby PW (2011) Tumor volume discrepancies between FDG-PET and MRI for cervical cancer. Radiother Oncol 98:139–142
4. Dimopoulos JC, De Vos V, Berger D et al (2009) Inter-observer comparison of target delineation for MRI-assisted cervical cancer brachytherapy: application of the GYN GEC-ESTRO recommendations. Radiother Oncol 91:166–172

5. Lin YC, Lin G, Hong JH et al (2017) Diffusion radiomics analysis of intratumoral heterogeneity in a murine prostate cancer model following radiotherapy: Pixelwise correlation with histology. J Magn Reson Imaging 46:483–489
6. Schob S, Meyer HJ, Pazaitis N et al (2017) ADC histogram analysis of cervical cancer aids detecting lymphatic metastases-a preliminary study. Mol Imaging Biol 19:953–962
7. Liu Y, Zhang Y, Cheng R et al (2019) Radiomics analysis of apparent diffusion coefficient in cervical cancer: a preliminary study on histological grade evaluation. J Magn Reson Imaging 49:280–290
8. Meng J, Zhu L, Zhu L et al (2017) Whole-lesion ADC histogram and texture analysis in predicting recurrence of cervical cancer treated with CCRT. Oncotarget 8:92442–92453
9. Lin G, Yang LY, Lin YC et al (2019) Prognostic model based on magnetic resonance imaging, whole-tumour apparent diffusion coefficient values and HPV genotyping for stage IB-IV cervical cancer patients following chemoradiotherapy. Eur Radiol 29:556–565
10. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
11. Torheim T, Malinen E, Hole KH et al (2017) Autodelineation of cervical cancers using multiparametric magnetic resonance imaging and machine learning. Acta Oncol 56:806–812
12. Cui S, Mao L, Jiang J, Liu C, Xiong S (2018) Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. J Healthc Eng 2018:4940593
13. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO (2018) Automatic lesion detection and segmentation of 18F-FET PET in gliomas: a full 3D U-Net convolutional neural network study. PLoS One 13:e0195798
14. Perkuhn M, Stavrinou P, Thiele F et al (2018) Clinical evaluation of a multiparametric deep learning model for glioblastoma segmentation using heterogeneous magnetic resonance imaging data from clinical routine. Invest Radiol 53:647–654
15. Trebeschi S, van Griethuysen JJM, Lambregts DMJ et al (2017) Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. Sci Rep 7:5301
16. Alkadi R, Taher F, El-Baz A, Werghi N (2018) A deep learning-based approach for the detection and localization of prostate cancer in T2 magnetic resonance images. J Digit Imaging. https://doi.org/10.1007/s10278-018-0160-1
17. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer International Publishing, Cham, pp 234–241
18. Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39:640–651
19. Wang J, Lu J, Qin G et al (2018) Technical note: a deep learning-based autosegmentation of rectal tumors in MR images. Med Phys 45:2560–2564
20. Arcos-Garcia A, Alvarez-Garcia JA, Soria-Morillo LM (2018) Deep neural network for traffic sign recognition systems: an analysis of spatial transformers and stochastic optimisation methods. Neural Netw 99:158–165
21. Lu CF, Hsu FT, Hsieh KL et al (2018) Machine learning-based radiomics for molecular subtyping of gliomas. Clin Cancer Res 24:4429–4436
22. Taha AA, Hanbury A (2015) Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 15:29
23. Dolz J, Xu X, Rony J et al (2018) Multiregion segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks. Med Phys 45:5482–5493
24. Huang YT, Chang CB, Yeh CJ et al (2018) Diagnostic accuracy of 3.0T diffusion-weighted MRI for patients with uterine carcinosarcoma: assessment of tumor extent and lymphatic metastasis. J Magn Reson Imaging. https://doi.org/10.1002/jmri.25981
25. Jalaguier-Coudray A, Villard-Mahjoub R, Delouche A et al (2017) Value of dynamic contrast-enhanced and diffusion-weighted MR imaging in the detection of pathologic complete response in cervical cancer after neoadjuvant therapy: a retrospective observational study. Radiology 284:432–442
26. Zhu Y, Wei R, Gao G et al (2019) Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. J Magn Reson Imaging 49:1149–1156