Original Article

Integrating a deep neural network and Transformer architecture for the automatic segmentation and survival prediction in cervical cancer

Shitao Zhu[1], Ling Lin[2], Qin Liu[3], Jing Liu[2], Yanwen Song[4], Qin Xu[2]

[1]College of Computer and Data Science, Fuzhou University, Fuzhou, China; [2]Department of Gynecology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China; [3]Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; [4]Department of Radiation Oncology, Xiamen Humanity Hospital, Xiamen, China

Contributions: (I) Conception and design: Q Xu, S Zhu; (II) Administrative support: Q Xu, L Lin, Q Liu, J Liu; (III) Provision of study materials or patients: Q Xu, L Lin, J Liu; (IV) Collection and assembly of data: S Zhu, L Lin, Y Song; (V) Data analysis and interpretation: Q Xu, S Zhu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Qin Xu, MD, PhD. Department of Gynecology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, No. 420 Fuma Road, Jin'an District, Fuzhou 350011, China. Email: 1379423879@qq.com.

Background: Automated tumor segmentation and survival prediction are critical to clinical diagnosis and treatment. This study aimed to develop deep-learning models for automatic tumor segmentation and survival prediction in magnetic resonance imaging (MRI) of cervical cancer (CC) by combining deep neural networks and Transformer architecture.

Methods: This study included 406 patients with CC, each with comprehensive clinical information and MRI scans. We randomly divided patients into training, validation, and independent test cohorts in a 6:2:2 ratio. During the model training, we employed two architecture types: one being a hybrid model combining convolutional neural network (CNN) and ransformer (CoTr) and one of pure CNNs. For survival prediction, the hybrid model combined tumor image features extracted by segmentation models with clinical information. The performance of the segmentation models was evaluated using the Dice similarity coefficient (DSC) and 95% Hausdorff distance (HD95). The performance of the survival models was assessed using the concordance index.

Results: The CoTr model performed well in both contrast-enhanced T1-weighted (ceT1W) and T2-weighted (T2W) imaging segmentation tasks, with average DSCs of 0.827 and 0.820, respectively, which outperformed other the CNN models such as U-Net (DSC: 0.807 and 0.808), attention U-Net (DSC: 0.814 and 0.811), and V-Net (DSC: 0.805 and 0.807). For survival prediction, the proposed deep-learning model significantly outperformed traditional methods, yielding a concordance index of 0.732. Moreover, it effectively divided patients into low-risk and high-risk groups for disease progression ($P<0.001$).

Conclusions: Combining Transformer architecture with a CNN can improve MRI tumor segmentation, and this deep-learning model excelled in the survival prediction of patients with CC as compared to traditional methods.

Keywords: Magnetic resonance imaging (MRI); Transformer; deep learning; cervical cancer (CC); survival prediction

Submitted Mar 19, 2024. Accepted for publication May 24, 2024. Published online Jul 16, 2024.
doi: 10.21037/qims-24-560
View this article at: https://dx.doi.org/10.21037/qims-24-560

Introduction

Cervical cancer (CC) is a leading cause of female cancer mortality worldwide, particularly in medically underserved countries (1). Medical imaging plays an essential role in oncology treatment, with conventional radiomics allowing for the extraction of tumor features and quantitative imaging information from magnetic resonance imaging (MRI) (2,3). Studies indicate that imaging features are valuable for prognostic analysis, clinical classification, and the diagnosis of various cancers (4-6). The imaging features of CC are widely used for grading, diagnosis (7,8), and prognostic analysis (9,10).

Accurate tumor segmentation is essential for precise radiomics analysis. Traditional methods rely on manual delineation by experts, which is both time-consuming and subjective. Over the past few years, deep-learning models have evolved considerably in their ability to segment the tumor lesions of various cancers (11-14). The most common approach involves employing the U-Net architecture for segmentation (15), which consists of a convolutional neural network (CNN) with a classical encoder-decoder structure. Although CNNs are widely adopted, their reliance on local focusing involve limitations that make it challenging to capture global contextual information (16-18).

Transformer is a sequence-to-sequence prediction architecture with a self-attention mechanism that dynamically adjusts the receptive field according to the input content, thus outperforming convolutional operations in long-range dependency (19). Transformer has demonstrated excellent performance in image-processing tasks, with significant advantages over traditional CNNs (20,21). The integration of CNN with Transformer is better adapted to the characteristics of medical images (16,17,22,23). According to certain reports (24,25), tumor segmentation tasks can capture prognostically relevant image features and combine them with clinical information to analyze survival prognosis.

Proportional hazard regression models have been used to estimate survival rates. However, these models assume linear relationships and fail to capture the nonlinear relationships that may exist in real life (26,27). Therefore, Katzman *et al.* (28) developed a deep-learning model, DeepSurv, which combines deep neural networks with the Cox proportional hazard (CPH) (29), and demonstrated that this deep-learning model is comparable to other traditional survival models.

In this study, we aimed to determine whether the Transformer structure can improve the segmentation performance of CNN models in CC and extract multiscale image features in the segmentation task to predict survival of patients with CC. We present this article in accordance with the TRIPOD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-24-560/rc).

Methods

Patients

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of Fujian Cancer Hospital (No. SQ2022-191). Individual consent for this retrospective analysis was waived.

This study included 406 patients diagnosed with CC at Fujian Cancer Hospital between 2009 and 2016 who underwent preoperative MRI treatment. Patients had complete clinical information, MRI, pathology diagnosis, and prognostic data, with the survival endpoint being overall survival (OS). The inclusion criteria for adult patients were (I) confirmation of visible tumors on pelvic MRI by experienced oncologists, (II) willingness to accept chemotherapy as standard treatment, and (III) availability of MRI and clinical information. The exclusion criteria were noncompliance with treatment and unavailability for follow-up.

Data preprocessing

The MRI data were acquired in the initial DICOM (Digital Imaging and Communications in Medicine) format with 1.5- or 3.0-T scanners and preprocessed. Detailed imaging parameters and preprocessing procedures can be found in Appendix 1. Two experienced radiologists (J.L. and L.L., each with over 6 years of experience in pelvic MRI reading) manually delineated three-dimensional (3D) tumor contours on axial slices using ITK-SNAP (version 4.0.0 software; www.itksnap.org) (30). The segmentation results were confirmed by another gynecological oncologist (Q.X., with 20 years of clinical experience). All physicians were blinded to the patient's clinical information but could ensure the tumor location using diffusion-weighted imaging (DWI). Interobserver variability between the 2 radiologists was assessed using MRI scans from 30 patients.

Deep-learning network

We use a hybrid CNN-Transformer 3D segmentation

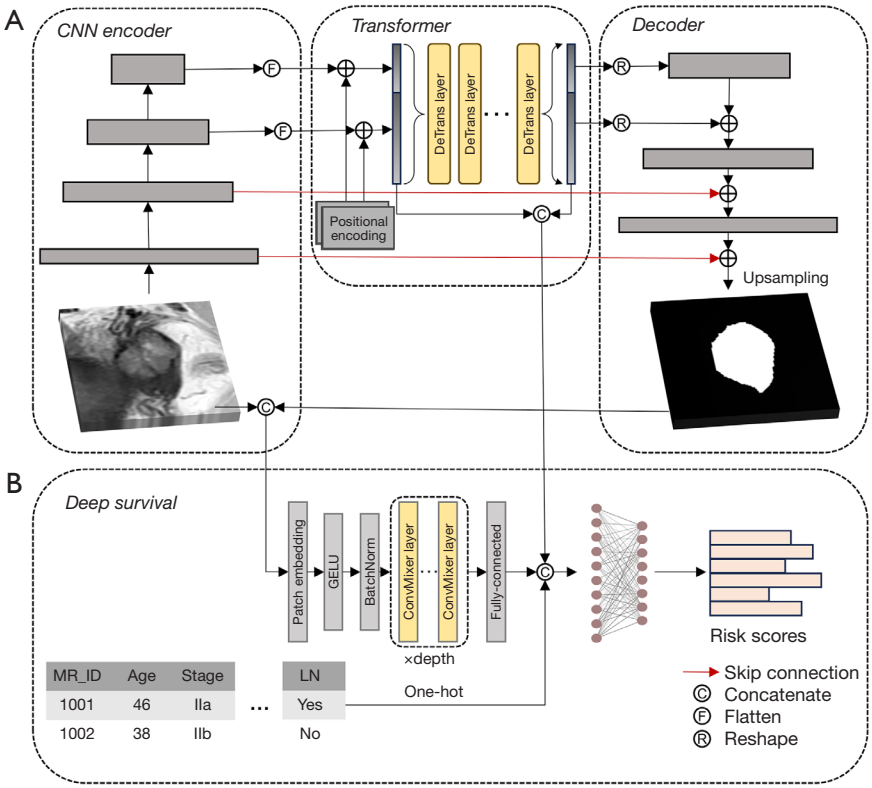

Figure 1 Deep-learning model architecture for segmentation and survival prediction. (A) The improved fundamental architecture is based on the CoTr model. Gray rectangles denote convolutional blocks, and yellow rectangles indicate 3D deformable Transformer layers. (B) Deep neural network prognostic model. CNN, convolutional neural network; GELU, Gaussian Error Linear Unit; CoTr, combined convolutional network and transformer model; 3D, three-dimensional.

model called *CoTr*, designed and proposed by Xie *et al.* (16). As illustrated in *Figure 1A*, the CoTr incorporates a hybrid structure of a CNN and Transformer architecture, consisting of three modules: CNN encoder, Transformer, and decoder. The CNN encoder includes a convolution batch normalization leaky rectified linear unit (CBL) block and three 3D residual blocks, with the CBL and the first residual block connected to the decoder through a residual connection. To accommodate the sequential input format of the Transformer, the model flattens the output features of the last two residual blocks. Positional encoding is applied to supplement 3D spatial information into the input sequence to preserve spatial information. Following Transformer processing, the output sequence is reshaped into feature maps. Segmentation outcomes are derived by upsampling the feature maps through three deconvolution layers and integrating residual inputs. The visualization of feature maps is achieved through the learning weights of the CNN encoder using an improved structure based on gradient-weighted class activation mapping (Grad-CAM) (31). Details of the segmentation model architecture and parameters can be found in Appendix 2.

As depicted in *Figure 1B*, we used the lightweight CNN architecture ConvMixer (32) for feature extraction in the prognostic task. In precisely determining the location of the tumor region more accurately, ConvMixer takes the concatenation of the image and the segmentation results as input. Subsequently, the extracted multiscale features from the segmentation model are integrated with clinical information and input into the deep-learning prognostic network, which primarily comprises four fully connected layers, ultimately producing a prediction score for each patient. The specific network structure and details are available in Appendix 3.

Model training

During the training phase of the network, we employed a

series of preprocessing steps on the data using the MONAI library (version 1.1.0). "NormalizeIntensityd" was used to normalize the image intensities, which reduced the computational overhead during training. The following enhancement operations were then performed with a probability of 0.3 for each batch of training cohort: random offset field transformation, the addition of Gibbs noise (α=0.3), random contrast adjustment ($\gamma \in [1.5, 2]$), random affine transformation, and random scaling with nearest-neighbor interpolation in the range of (0.9–1.1). For the deep-learning model, clinical information was transformed into trainable data by the one-hot coding method and combined with multiscale image features extracted by the segmentation model as inputs to the network.

Training for all segmentation models consisted of 100 epochs with a batch size of 16. The AdamW optimizer had an initial learning rate of 0.001, and the weight decay parameter was set to 0.1, decaying every 15 epochs. The total training epochs for the prognostic model were 500, and the entire cohort was processed in each batch: the AdamW was used as the optimizer at an initial learning rate of 0.001; the weight decay parameter was 0.5, decaying every 50 epochs; and the number of early-stopping rounds was 50. All experiments were conducted on a local server equipped with 8 RTX T4 16 GB GPUs (Nvidia Corp., Santa Clara, CA, USA). The code is available online (https://github.com/khuanging/seg-surv-CC).

Performance evaluation

We employed the Dice similarity coefficient (DSC) and 95% Hausdorff distance (HD95) as quantitative evaluation metrics for the independent test cohort (33,34). DSC measures accuracy by evaluating the overlap between predicted contours and manual segmentations. In contrast, HD95 measures the maximum distance between two segmentation contours, representing the largest surface-to-surface distance among the closest 95% surface voxels. The specific formula for the metrics are provided in Appendix 4. For the deep-learning survival models, performance was assessed using the concordance index, which ranges from 0 to 1, with 0 indicating random prediction and 1 indicating perfect prediction. The "survivalROC" package in R software (The Foundation for Statistical Computing) was used to determine the optimal cutoff value for the prediction value of the training cohort. Patients were then divided into groups of low and high risk for disease progression. Kaplan-Meier curves were used to assess the reliability of this risk stratification system. Inter-reader agreement was also evaluated for 30 lesions using the DSC.

Statistical analysis

To compare patient characteristics between the training and independent test cohort, the *t*-test was applied for continuous variables, while the chi-square test (n>5 in either cohort) or the Yates correction test (n<5 in either cohort) were used for categorical variables, with P>0.05 indicating no significant difference between groups (10,35). The Wilcoxon rank test was employed to assess differences in medians across the model evaluation results. The differences between DSCs and tumor histology were evaluated using the Mann-Whitney test. All results with a P value <0.05 were considered to be statistically significant.

Results

Patient characteristics

Table 1 presents the clinical and demographic characteristics of the training cohort (n=243) and the independent test cohort (n=82). The mean patient age was 47±8 years, and the median follow-up time of surviving patients was 103 months and 110 months for the training and independent test cohort, respectively. The results indicate no significant statistical differences in the clinical characteristics of between the two groups of patients. The inter-observer agreement for the 30-lesion volumetric segmentation analysis showed a mean DSC of 0.896 for contrast-enhanced T1-weighted (ceT1W) images and 0.867 for T2-weighted (T2W) images.

Segmentation performance in the independent test cohort

Figure 2 displays the partial results of each model's automatic segmentation of tumors in the ceT1W and T2W images. *Table 2* provides the DSC and HD95 for each model on the independent test cohort for both sequences. In the ceT1W sequence, the CoTr model exhibited superior performance in overall tumor segmentation and tumor contour segmentation, with a DSC of 0.827 and an HD95 of 0.854 cm. This performance was significantly better than that of the other traditional CNN models (all P values <0.01; Figure S3). In the T2W sequence, the CoTr model achieved the best DSC (0.820). However, its performance

Table 1 Patient characteristics of the training cohort and the test cohort

Characteristic	Training (n=243)	Test (n=82)	P value
Age (years)[a]	47.1±8.0	47.0±8.4	0.886[d]
Stage (FIGO 2009), n (%)			0.116[e]
IIa	110 (45.3)	41 (50.0)	
Ib2	23 (9.5)	2 (2.4)	
IIb	110 (45.3)	39 (47.6)	
Macroscopic type, n (%)			0.506[e]
Nodular type	129 (53.1)	47 (57.3)	
Cauliflower-like type	114 (46.9)	35 (42.7)	
Tumor size (cm), n (%)[b]			0.868[e]
≤4	113 (46.5)	39 (47.6)	
>4	130 (53.5)	43 (52.4)	
Postoperative pathological, n (%)			0.792[e]
Squamous cell carcinoma	223 (91.8)	76 (92.7)	
Other[c]	20 (8.2)	6 (7.3)	
Differentiation degree, n (%)			0.263[f]
Medium	186 (76.5)	57 (69.5)	
Low	51 (21.0)	24 (29.3)	
High	6 (2.5)	1 (1.2)	
Depth of tumor invasion, n (%)			0.708[e]
Deep	145 (59.7)	47 (57.3)	
Shallow	98 (40.3)	35 (42.7)	
Corpus invasion, n (%)			0.628[e]
No	221 (90.9)	76 (92.7)	
Yes	22 (9.1)	6 (7.3)	
Parametrial invasion, n (%)			0.368[f]
No	233 (95.9)	81 (98.8)	
Yes	10 (4.1)	1 (1.2)	
Lymph node metastasis, n (%)			0.757[e]
No	61 (25.1)	22 (26.8)	
Yes	182 (74.9)	60 (73.2)	

[a], mean ± standard deviation; [b], clinical diagnostic results; [c], the training cohort comprising 6 cases of adenosquamous carcinoma and the remaining cases of adenocarcinoma and the test set exclusively containing adenocarcinoma cases; [d], *t*-test; [e], Pearson Chi-square test; [f], Yates correction for continuity. FIGO, International Federation of Gynecology and Obstetrics.

in tumor contour segmentation was unsatisfactory, with an HD95 value of 1.519 cm, which was only slightly higher than the value of 1.821 cm yielded by U-Net. The V-Net model performed relatively well in the contour boundary segmentation of both sequences, with HD95 values of 0.915 and 1.009 cm, respectively. In the comparison of model performance metrics for different sequences, CoTr outperformed the other CNN models, as shown in *Figure 2*.

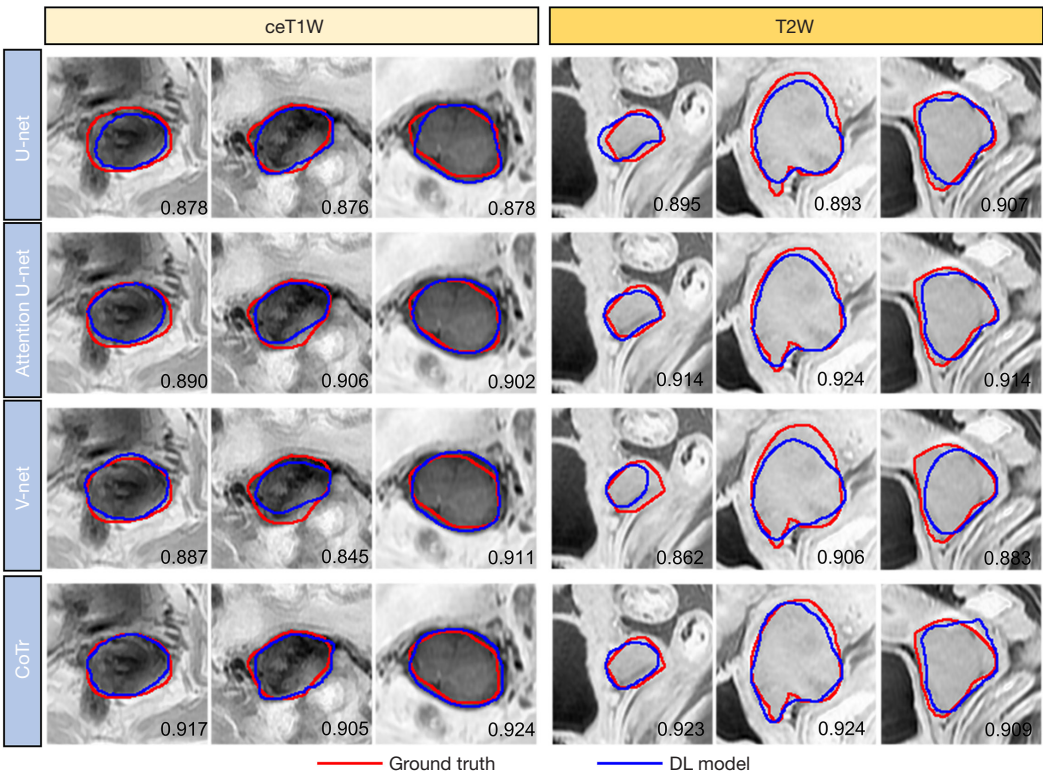

Figure 2 Example of well-performed tumor segmentation on axial slices at the middle of ceT1W and T2W images for three patients with cervical cancer. The manually segmented tumors are highlighted in red, while the model predictions are indicated in dark blue. Numerical values represent the scores of the DSC. ceT1W, contrast-enhanced T1-weighted; T2W, T2-weighted; CoTr, combined convolutional network and transformer model; DL, deep learning; DSC, Dice similarity coefficient.

Table 2 Results of quantitative DSC and HD95 metrics for automated tumor segmentation of the ceT1W and T2W images in the test cohort

Model	ceT1W		T2W	
	DSC	HD95 (cm)	DSC	HD95 (cm)
U-Net	0.807	1.678	0.808	1.821
Attention U-Net	0.814	1.377	0.811	1.122
V-Net	0.805	0.915	0.807	1.009
CoTr	0.827	0.854	0.820	1.519

DSC, Dice similarity coefficient; HD95, 95% Hausdorff distance; ceT1W, contrast-enhanced T1-weighted; T2W, T2-weighted; CoTr, combined convolutional network and transformer model.

Error analysis of the independent test cohort

As seen in *Figure 3*, the deep-learning model erroneously identified the cervix in the ceT1W images as a tumor region and had difficulty identifying larger or smaller tumor regions in the T2W images. An in-depth analysis of each 3D model's performance revealed that around 12% to 23% (n=10 to 19) of the independent test cohort had results showing a DSC below 0.750. *Figure 4* shows the relationship between the DSC and manually segmented tumor volume in the independent test cohort. There was a strong correlation between the DSC and tumor volume in all models (all R values>0.35; P<0.05). In the

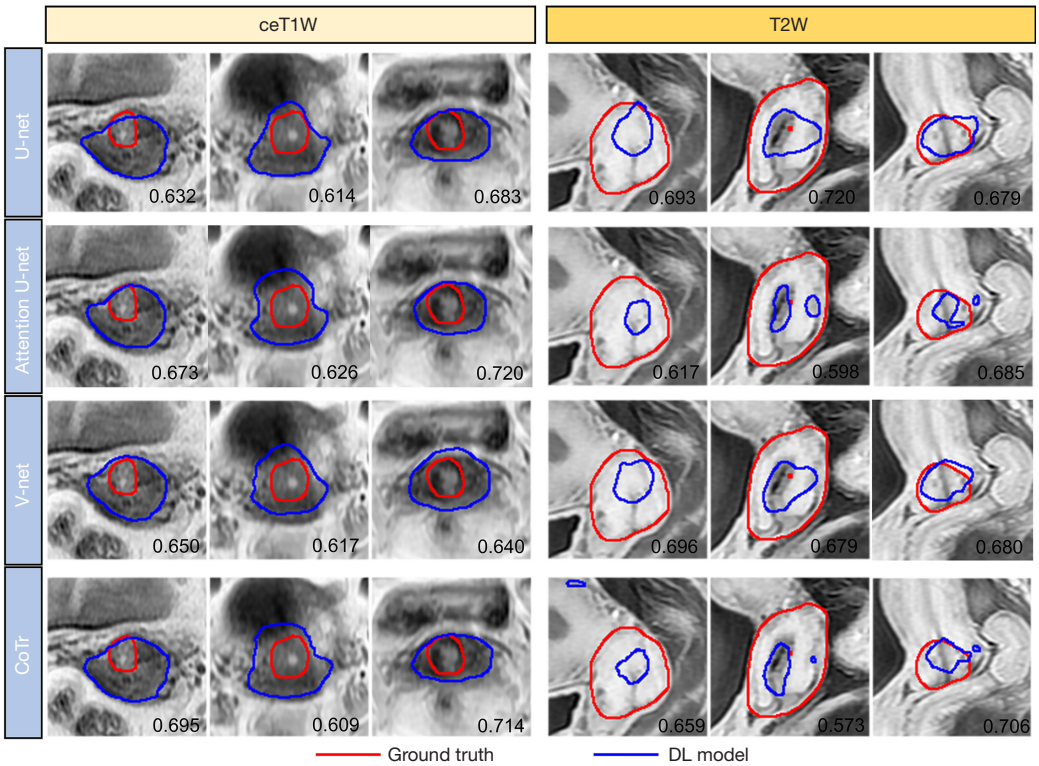

Figure 3 Example of poorly performed tumor segmentation on axial slices at the middle of ceT1W and T2W images for three patients with cervical cancer. The manually segmented tumors are highlighted in red, while the model predictions are indicated in dark blue. Numerical values represent the scores of the DSC. ceT1W, contrast-enhanced T1-weighted; T2W, T2-weighted; CoTr, combined convolutional network and transformer model; DL, deep learning; DSC, Dice similarity coefficient.

ceT1W images, all models performed poorly in small-volume tumors, but CoTr had a better DSC compare to the other models. The DSC of lymph node metastasis was significantly lower than the nonmetastatic score in ceT1W images (0.816 *vs.* 0.857; P=0.023), and there was a nonsignificant difference in the DSC for the T2W images (0.810 *vs.* 0.848; P=0.085). No significant correlation was observed between the DSC and other clinical characteristics (P>0.05).

Interpretability of the deep-learning model

Figure 5 intuitively shows the feature activation maps of the CNN encoder in the CoTr model for the feature extraction phase in the ceT1W and T2W images. Initially, the model focuses uniformly on features at all locations. As the model is updated, it progressively suppresses the unrelated regions and intensifies its attention on local tumor regions, gradually approximating the tumor contour.

Prognostic model performance

Table 3 displays the outcomes of the traditional survival analysis methods with the deep-learning survival prediction model. Due to model structure limitations, the validation cohort was merged into the training cohort while the independent test cohort was left unchanged, and the CPH used only clinical data. With only clinical information, the concordance index of the CPH and random survival forest (RSF) were 0.692 and 0.707, respectively. Subsequently, we fused the multiscale features extracted from the ceT1W and T2W images and validated them in the RSF model, resulting in an improved concordance index of 0.713. Meanwhile, our deep-learning model integrated image features and clinical information, yielding an average concordance index of 0.733, which was significantly better than that of the other models. The cutoff value (0.171) from the training cohort was applied to both the validation cohort and the independent test cohort, yielding

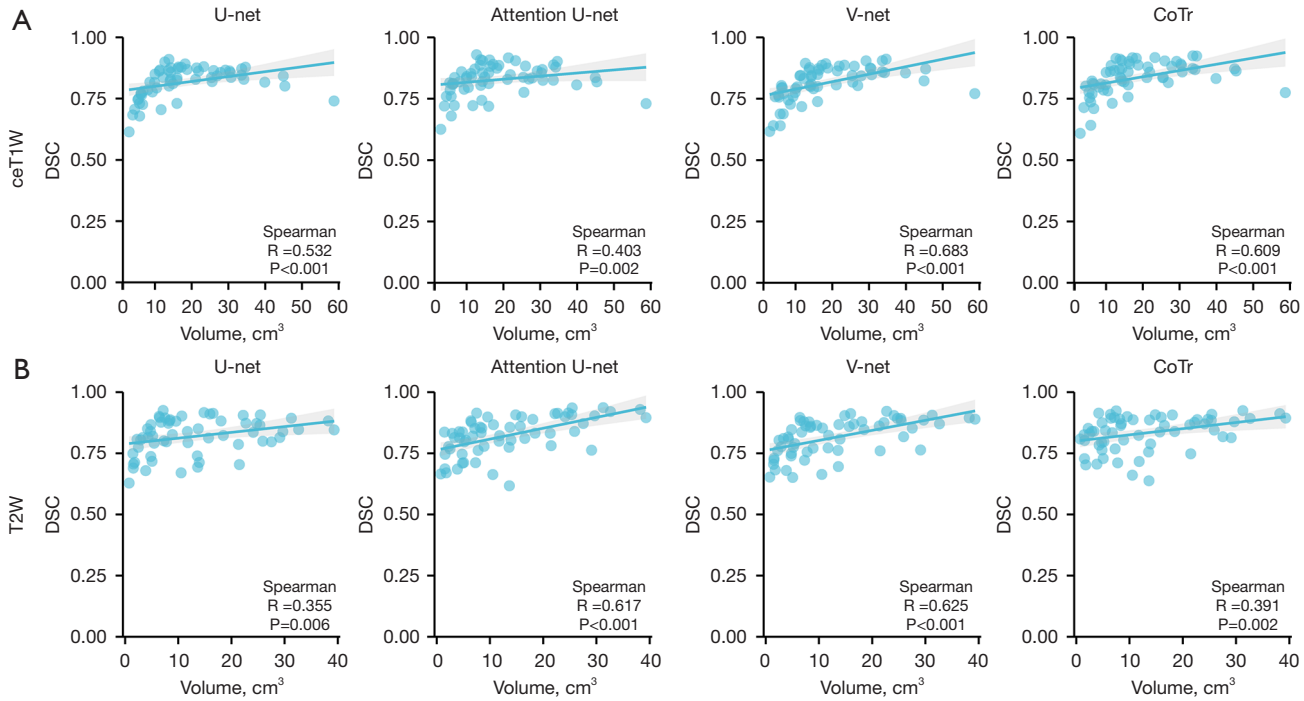

Figure 4 Scatter plot of the DSC and manually segmented tumor volumes of patients in the independent test cohort for different sequences. (A) ceT1W imaging. (B) T2W imaging. R: Spearman rank correlation coefficient, the absolute value represents the degree of correlation (weak correlation: 0.3–0.5; moderate correlation: 0.5–0.8). P<0.05 indicates statistical significance. ceT1W, contrast-enhanced T1-weighted; T2W, T2-weighted; CoTr, combined convolutional network and transformer model; DSC, Dice similarity coefficient.

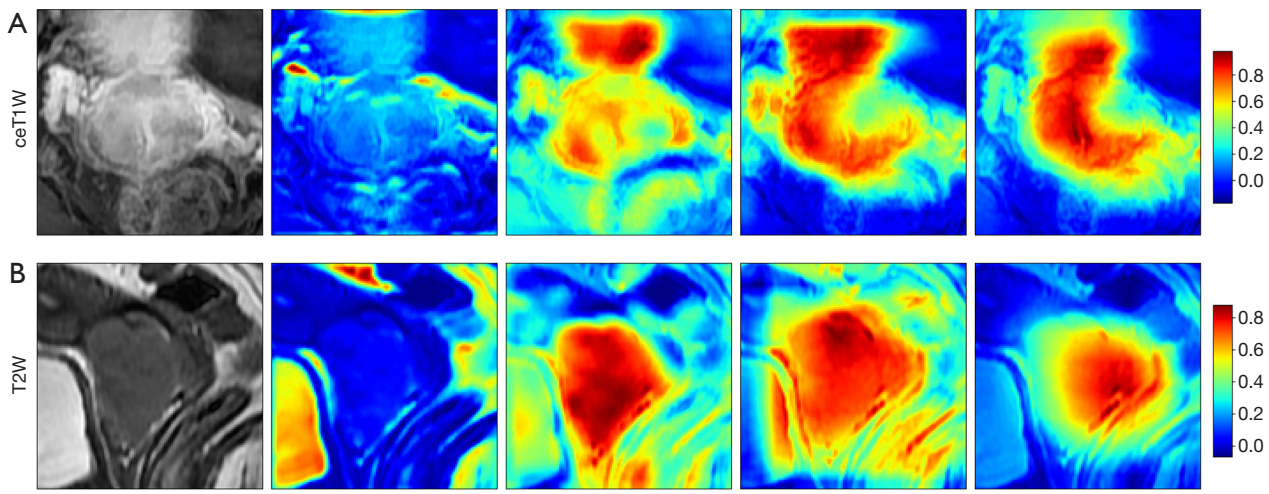

Figure 5 Class activation maps for the CoTr model's encoder section. (A) Axial plane of ceT1W images. (B) Axial plane of T2W images. The 3D convolutional layer, layer 1, layer 2, and layer 3 of the CNN encoder are shown from left to right, where red indicates the regions the model focused on. ceT1W, contrast-enhanced T1-weighted; T2W, T2-weighted; CoTr, combined convolutional network and transformer model; 3D, three-dimensional; CNN, convolutional neural network.

Table 3 Prognostic performance of different models in the test cohort.

Method	CPH	RSF	RSF*	CoTr*
Concordance index	0.692	0.707	0.713	0.733

*, the multiscale image features and clinical features were used for training and testing. CPH, Cox proportional hazard; RSF, random survival forest; CoTr, combined convolutional network and transformer model.

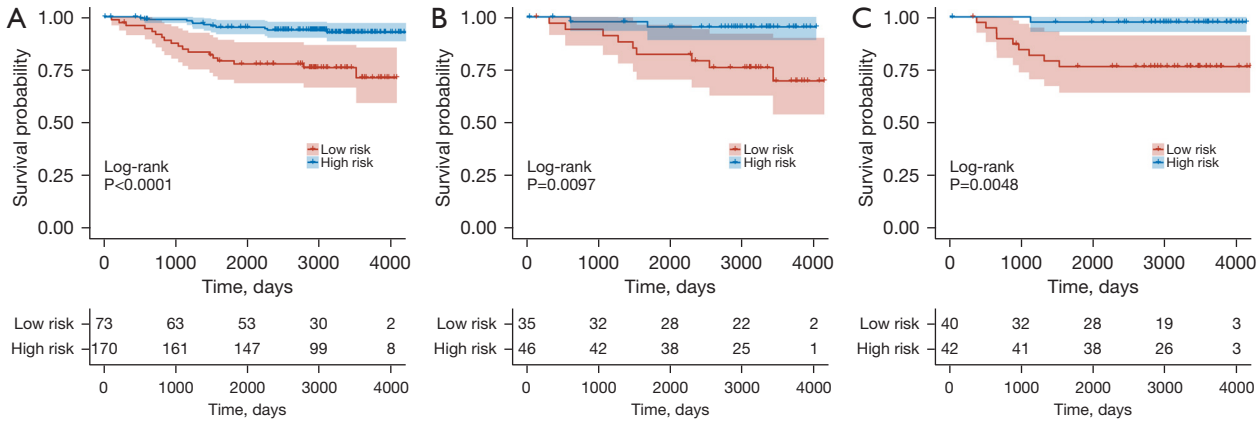

Figure 6 Kaplan-Meier analysis of OS prediction for the training cohort (A), validation cohort (B), and independent test cohort (C) based on risk scores predicted by the deep-learning model. OS, overall survival.

a statistically significant patient risk stratification (validation: log-rank P=0.0097; test: log-rank P=0.0048; *Figure 6*).

Discussion

MRI is essential for diagnosing and treating CC, aiding physicians in assessing the extent of the primary tumor, local staging, and lymph node invasion (7,10). However, accurate primary lesion delineation is often time-consuming and demands specialized expertise (36). Although deep-learning models have found widespread application in automating tumor segmentation, the prevalent architectures have predominantly relied on traditional CNNs such as U-Net (11,37,38). However, recent research suggests that networks combining CNNs and Transformer architecture exhibit superior performance in medical image segmentation (16,17,23). In this study, we collected MRI scans from 406 patients with confirmed CC to evaluate the segmentation results of different models on ceT1W and T2W imaging. Meanwhile, we used the trained segmentation model to extract multiscale features and combined them with clinical information to conduct survival prediction with the deep-learning model. Our results show that the CNN-Transformer model outperformed the CNN models in segmentation, and the deep-learning survival prediction network was superior to the traditional methods.

CNN faces challenges in capturing the long-range dependency in images due to the limitations of the convolution sensory field. Integrating the self-attention mechanism in Transformer with CNNs has demonstrated efficacy in the extraction of 3D contextual information from images (23). In our study, the hybrid network (CoTr) achieved the best DSC on the independent test cohort for both ceT1W and T2W images (0.827 and 0.820, respectively), significantly outperforming the other CNN models (all P values <0.05).

Several studies have examined the automatic segmentation of MRI scans for CC. Lin *et al.* (39) reported a maximum DSC of 0.820 (on multiparametric MRI) (n=25) using the U-Net model. Similarly, Hodneland *et al.* (10) used the U-Net model to segment T2W images, yielding an adjusted DSC of only 0.780 (n=26). In contrast, other studies (40,41) have reported higher DSCs (all >0.80); however, these studies employed k-fold cross-validation without the inclusion of an independent test cohort and thus are not directly comparable to our study.

Previous reports have indicated there to be a correlation between the segmentation accuracy of deep-learning models

and tumor size, with relatively larger tumor volumes being more easily segmented (10,35). In this study, all models encountered challenges in segmenting tumors of different sizes, including oversized and small tumors with unclear boundaries. Meanwhile, as can be seen from the scatter plot in *Figure 4*, the tumor segmentation performance was positively correlated with tumor volume.

Deep-learning models are being increasingly applied to cancer survival prediction and have shown superiority over traditional methods (24,26,28). Matsuo *et al.* (26) achieved a concordance index of 0.616 (n=768) in predicting the OS of those with CC using a deep-learning model, which represents a slight improvement over the CPH model (concordance index =0.607). In our study, the concordance index of the deep-learning model combining multiscale image features with clinical characteristics was 0.733, while the concordance index values of the traditional CPH and RSF methods were 0.692 and 0.707, respectively. This indicates that deep learning is a feasible method for CC survival prediction and outperforms conventional methods.

There were several limitations in our study that should be acknowledged. First, despite examining a relatively large cohort, we employed a single-center, retrospective design, which involves inherent biases and lacks generalizability. Second, we did not consider the fusion training of multiparametric MRI, and further exploration is needed to reduce outliers in the results through multisequence fusion and multimodel voting. Third, the CoTr model remains limited in some regards, and there is a need for further improvements in the Transformer structure to balance model performance and computational overhead.

Conclusions

Our study indicates that the CoTr model outperforms the CNN models in the segmentation of tumors in CC. In addition, the deep-learning model provides superior performance in the survival prediction of patients with CC compared to traditional methods.

Acknowledgments

Funding: This work was supported by Joint Funds for the Innovation of Science and Technology, Fujian Province (No. 2023Y9449), Joint Funds for the National Clinical Key Specialty Construction Program (2021), the Natural Science Foundation of Fujian Province (No. 2023J011273), and Fujian Cancer Hospital High-Level Talent Training Program (No. 2024YNG09).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-24-560/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-24-560/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of Fujian Cancer Hospital (No. SQ2022-191). Individual consent for this retrospective analysis was waived. This diagnostic study was not registered on any clinical trial platform.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Varghese BA, Cen SY, Hwang DH, Duddalwar VA. Texture Analysis of Imaging: What Radiologists Need to Know. AJR Am J Roentgenol 2019;212:520-8.
3. Wang Y, Wang M, Cao P, Wong EMF, Ho G, Lam TPW, Han L, Lee EYP. CT-based deep learning segmentation of ovarian cancer and the stability of the extracted radiomics features. Quant Imaging Med Surg 2023;13:5218-29.

4. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, Mao R, Li F, Xiao Y, Wang Y, Hu Y, Yu J, Zhou J. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 2020;11:1236.
5. Cui Y, Zhang J, Li Z, Wei K, Lei Y, Ren J, Wu L, Shi Z, Meng X, Yang X, Gao X. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: A multicenter cohort study. EClinicalMedicine 2022;46:101348.
6. Lefebvre TL, Ueno Y, Dohan A, Chatterjee A, Vallières M, Winter-Reinhold E, Saif S, Levesque IR, Zeng XZ, Forghani R, Seuntjens J, Soyer P, Savadjiev P, Reinhold C. Development and Validation of Multiparametric MRI-based Radiomics Models for Preoperative Risk Stratification of Endometrial Cancer. Radiology 2022;305:375-86.
7. Xiao M, Ma F, Li Y, Li Y, Li M, Zhang G, Qiang J. Multiparametric MRI-Based Radiomics Nomogram for Predicting Lymph Node Metastasis in Early-Stage Cervical Cancer. J Magn Reson Imaging 2020;52:885-96.
8. Zhu Q, Che P, Li M, Guo W, Ye K, Yin W, Chu D, Wang X, Li S. Artificial intelligence for segmentation and classification of lobar, lobular, and interstitial pneumonia using case-specific CT information. Quant Imaging Med Surg 2024;14:579-91.
9. Zhou Y, Gu HL, Zhang XL, Tian ZF, Xu XQ, Tang WW. Multiparametric magnetic resonance imaging-derived radiomics for the prediction of disease-free survival in early-stage squamous cervical cancer. Eur Radiol 2022;32:2540-51.
10. Hodneland E, Kaliyugarasan S, Wagner-Larsen KS, Lura N, Andersen E, Bartsch H, Smit N, Halle MK, Krakstad C, Lundervold AS, Haldorsen IS. Fully Automatic Whole-Volume Tumor Segmentation in Cervical Cancer. Cancers (Basel) 2022;14:2372.
11. Rodrigues NM, Silva S, Vanneschi L, Papanikolaou N. A Comparative Study of Automated Deep Learning Segmentation Models for Prostate MRI. Cancers (Basel) 2023;15:1467.
12. Ramesh KK, Xu KM, Trivedi AG, Huang V, Sharghi VK, Kleinberg LR, Mellon EA, Shu HG, Shim H, Weinberg BD. A Fully Automated Post-Surgical Brain Tumor Segmentation Model for Radiation Treatment Planning and Longitudinal Tracking. Cancers (Basel) 2023;15:3956.
13. Isaksson LJ, Summers P, Mastroleo F, Marvaso G, Corrao G, Vincini MG, Zaffaroni M, Ceci F, Petralia G, Orecchia R, Jereczek-Fossa BA. Automatic Segmentation with Deep Learning in Radiotherapy. Cancers (Basel) 2023;15:4389.
14. Priya S, Dhruba DD, Perry SS, Aher PY, Gupta A, Nagpal P, Jacob M. Optimizing Deep Learning for Cardiac MRI Segmentation: The Impact of Automated Slice Range Classification. Acad Radiol 2024;31:503-13.
15. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Springer, 2015;9351:234-41.
16. Xie Y, Zhang J, Shen C, Xia Y. CoTr: Efficiently Bridging CNN and Transformer for 3D Medical Image Segmentation. In: de Bruijne M, Cattin PC, Cotin S, Padoy N, Speidel S, Zheng Y, Essert C. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Lecture Notes in Computer Science, Springer 2021;12903;171-80.
17. Wang W, Chen C, Ding M, Yu H, Zha S, Li J. TransBTS: Multimodal Brain Tumor Segmentation Using Transformer. In: de Bruijne M, Cattin PC, Cotin S, Padoy N, Speidel S, Zheng Y, Essert C. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Lecture Notes in Computer Science, Springer, 2021;12901:109-19.
18. Khan RF, Lee BD, Lee MS. Transformers in medical image segmentation: a narrative review. Quant Imaging Med Surg 2023;13:8747-67.
19. Ashish A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is All you Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.
20. Yao T, Li Y, Pan Y, Wang Y, Zhang XP, Mei T. Dual Vision Transformer. IEEE Trans Pattern Anal Mach Intell 2023;45:10870-82.
21. Azad R, Kazerouni A, Heidari M, Aghdam EK, Molaei A, Jia Y, Jose A, Roy R, Merhof D. Advances in medical image analysis with vision Transformers: A comprehensive review. Med Image Anal 2024;91:103000.
22. Dai Y, Gao Y, Liu F. TransMed: Transformers Advance Multi-Modal Medical Image Classification. Diagnostics (Basel) 2021;11:1384.
23. Shaker AM, Maaz M, Rasheed H, Khan S, Yang MH, Khan FS. UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation. IEEE Trans Med Imaging 2024. [Epub ahead of print]. doi: 10.1109/TMI.2024.3398728.
24. Starke S, Zwanenburg A, Leger K, Lohaus F, Linge

A, Kalinauskaite G, et al. Multitask Learning with Convolutional Neural Networks and Vision Transformers Can Improve Outcome Prediction for Head and Neck Cancer Patients. Cancers (Basel) 2023;15:4897.
25. Saeed N, Sobirov I, Al Majzoub R, Yaqub M. TMSS: An End-to-End Transformer-Based Multimodal Network for Segmentation and Survival Prediction. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Lecture Notes in Computer Science, Springer, 2022;13437:319-29.
26. Matsuo K, Purushotham S, Jiang B, Mandelbaum RS, Takiuchi T, Liu Y, Roman LD. Survival outcome prediction in cervical cancer: Cox models vs deep-learning model. Am J Obstet Gynecol 2019;220:381.e1-381.e14.
27. Matsuo K, Purushotham S, Moeini A, Li G, Machida H, Liu Y, Roman LD. A pilot study in using deep learning to predict limited life expectancy in women with recurrent cervical cancer. Am J Obstet Gynecol 2017;217:703-5.
28. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18:24.
29. Cox DR. Regression models and life-tables. J Royal Stat Soc B (Methodological) 1972;34:187-202.
30. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116-28.
31. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017:618-26.
32. Trockman A, Kolter JZ. Patches Are All You Need? Available online: https://arxiv.org/abs/2201.09792v1
33. Choi MS, Choi BS, Chung SY, Kim N, Chun J, Kim YB, Chang JS, Kim JS. Clinical evaluation of atlas- and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer. Radiother Oncol 2020;153:139-45.
34. Rouvière O, Moldovan PC, Vlachomitrou A, Gouttard S, Riche B, Groth A, Rabotnikov M, Ruffion A, Colombel M, Crouzet S, Weese J, Rabilloud M. Combined model-based and deep learning-based automated 3D zonal segmentation of the prostate on T2-weighted MR images: clinical evaluation. Eur Radiol 2022;32:3248-59.
35. Lin YC, Lin G, Pandey S, Yeh CH, Wang JJ, Lin CY, Ho TY, Ko SF, Ng SH. Fully automated segmentation and radiomics feature extraction of hypopharyngeal cancer on MRI using deep learning. Eur Radiol 2023;33:6548-56.
36. Buelens P, Willems S, Vandewinckele L, Crijns W, Maes F, Weltens CG. Clinical evaluation of a deep learning model for segmentation of target volumes in breast cancer radiotherapy. Radiother Oncol 2022;171:84-90.
37. Chen X, Mumme RP, Corrigan KL, Mukai-Sasaki Y, Koutroumpakis E, Palaskas NL, Nguyen CM, Zhao Y, Huang K, Yu C, Xu T, Daniel A, Balter PA, Zhang X, Niedzielski JS, Shete SS, Deswal A, Court LE, Liao Z, Yang J. Deep learning-based automatic segmentation of cardiac substructures for lung cancers. Radiother Oncol 2024;191:110061.
38. Wu B, Zhang F, Xu L, Shen S, Shao P, Sun M, Liu P, Yao P, Xu RX. Modality preserving U-Net for segmentation of multimodal medical images. Quant Imaging Med Surg 2023;13:5242-57.
39. Lin YC, Lin CH, Lu HY, Chiang HJ, Wang HK, Huang YT, Ng SH, Hong JH, Yen TC, Lai CH, Lin G. Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer. Eur Radiol 2020;30:1297-305.
40. Torheim T, Malinen E, Hole KH, Lund KV, Indahl UG, Lyng H, Kvaal K, Futsaether CM. Autodelineation of cervical cancers using multiparametric magnetic resonance imaging and machine learning. Acta Oncol 2017;56:806-12.
41. Bnouni N, Rekik I, Rhim MS, Ben Amara NE. Context-Aware Synergetic Multiplex Network for Multi-organ Segmentation of Cervical Cancer MRI. In: Rekik I, Adeli E, Park SH, Valdés Hernández MdC. editors. Predictive Intelligence in Medicine. PRIME 2020. Lecture Notes in Computer Science, Springer, 2020;12329:1-11.

Cite this article as: Zhu S, Lin L, Liu Q, Liu J, Song Y, Xu Q. Integrating a deep neural network and Transformer architecture for the automatic segmentation and survival prediction in cervical cancer. Quant Imaging Med Surg 2024;14(8):5408-5419. doi: 10.21037/qims-24-560

