



# The Effects of Automatic Segmentations on Preoperative Lymph Node Status Prediction Models With Ultrasound Radiomics for Patients With Early Stage Cervical Cancer

Technology in Cancer Research & Treatment  
Volume 21: 1-7  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/15330338221099396  
journals.sagepub.com/home/tct  


Yinyan Teng, MS<sup>1,\*</sup>, Yao Ai, MS<sup>2,\*</sup>, Tao Liang, BS<sup>2</sup>, Bing Yu, MS<sup>2</sup>,  
Juebin Jin, MS<sup>3</sup>, Congying Xie, PhD<sup>2,4</sup>, and Xiance Jin, PhD<sup>2,5</sup> 

## Abstract

**Introduction:** The purpose of this study is to investigate the effects of automatic segmentation algorithms on the performance of ultrasound (US) radiomics models in predicting the status of lymph node metastasis (LNM) for patients with early stage cervical cancer preoperatively. **Methods:** US images of 148 cervical cancer patients were collected and manually contoured by two senior radiologists. The four deep learning-based automatic segmentation models, namely U-net, context encoder network (CE-net), Resnet, and attention U-net were constructed to segment the tumor volumes automatically. Radiomics features were extracted and selected from manual and automatically segmented regions of interest (ROIs) to predict the LNM of these cervical cancer patients preoperatively. The reliability and reproducibility of radiomics features and the performances of prediction models were evaluated. **Results:** A total of 449 radiomics features were extracted from manual and automatic segmented ROIs with Pyradiomics. Features with an intraclass coefficient (ICC) > 0.9 were all 257 (57.2%) from manual and automatic segmented contours. The area under the curve (AUCs) of validation models with radiomics features extracted from manual, attention U-net, CE-net, Resnet, and U-net were 0.692, 0.755, 0.696, 0.689, and 0.710, respectively. Attention U-net showed best performance in the LNM prediction model with a lowest discrepancy between training and validation. The AUCs of models with automatic segmentation features from attention U-net, CE-net, Resnet, and U-net were 9.11%, 0.58%, -0.44%, and 2.61% higher than AUC of model with manual contoured features, respectively. **Conclusion:** The reliability and reproducibility of radiomics features, as well as the performance of radiomics models, were affected by manual segmentation and automatic segmentations.

<sup>1</sup> Department of Ultrasound imaging, Wenzhou Medical University First Affiliated Hospital, Wenzhou, People's Republic of China

<sup>2</sup> Department of Radiotherapy Center, Wenzhou Medical University First Affiliated Hospital, Wenzhou, People's Republic of China

<sup>3</sup> Department of Medical Engineering, Wenzhou Medical University First Affiliated Hospital, Wenzhou, People's Republic of China

<sup>4</sup> Department of Radiation and Medical Oncology, Wenzhou Medical University Second Affiliated Hospital, Wenzhou, People's Republic of China

<sup>5</sup> School of Basic Medical Science, Wenzhou Medical University, Wenzhou, People's Republic of China

\*The authors contributed equally.

## Corresponding Authors:

Chongying Xie, Department of Radiation and Medical Oncology, Wenzhou Medical University Second Affiliated Hospital, Wenzhou 325000, People's Republic of China.

Email: wzxiecongying@163.com

Xiance Jin, Department of Radiotherapy Center, Wenzhou Medical University First Affiliated Hospital, Shangcai Village, Wenzhou 325000, People's Republic of China.

Email: jinxc1979@hotmail.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Keywords

ultrasound, cervical cancer, lymph node metastasis, segmentation, radiomics

## Abbreviations

AG, attention gate; AUC, area under the curve; CE-net, context encoder network; DAC, dense atrous convolution; DSC, dice similarity coefficient; ECCR, Ethics Committee in Clinical Research; FIGO, International Federation of Gynecology and Obstetrics; ICC, intraclass correlation coefficients; LNM, lymph node metastasis; LASSO, least absolute shrinkage and selection operator; MR, magnetic resonance; NSCLC, non-small cell lung cancer; RMP, residual multi-kernel pooling; ROC, receiver operating characteristic; ROIs, regions of interest; US, ultrasound

Received: October 11, 2021; Revised: April 2, 2022; Accepted: April 19, 2022.

## Introduction

Ultrasound (US) is one of the most used imaging modalities in specialties of echocardiography, breast US, abdominal US, transrectal US, intravascular US, prenatal diagnosis US, etc., because of its advantages of non-ionizing radiation, portability, accessibility, and cost effectiveness.<sup>1,2</sup> With the emergency of radiomics, extracting qualitative and quantitative information from medical images and modeling with clinical data have become a hot translation field for precision medicine to support evidence-based clinical decision making.<sup>3</sup> Radiomics features extracted from US images had been proved to be strongly associated with breast biologic characteristics,<sup>4</sup> gestational age,<sup>5</sup> neonatal respiratory morbidity,<sup>6</sup> etc. Studies also reported that US-based radiomics was able to predict the lymph node metastasis (LNM) for patients with papillary thyroid carcinoma and cervical cancer.<sup>7, 8</sup>

One critical step of the radiomics process is to segment regions of interest (ROIs) from which radiomics features are extracted for further modeling. However, manual delineation is very labor-intensive and tedious. It is also not always feasible since radiomics usually requires very large data sets.<sup>9</sup> On the other hand, manual delineation of US images is a subjective and error prone task highly dependent on the image quality and the expertise of the observer. Studies demonstrated that the analysis and modeling of radiomics could be significantly affected by the inherent variability of manual segmentation.<sup>10, 11</sup> Therefore, automatic segmentation is preferred to address these problems.

Recently, there has been an extensive development of detection and segmentation algorithms for medical US images.<sup>1, 12</sup> However, few studies concerned the effects of automatic segmentations on the reproducibility and stability of quantitative features, as well as on the models using these features obtained from US in oncology.<sup>13</sup> Previously, automatic segmentations using deep learning algorithm on US and magnetic resonance (MR) images were investigated and reported a generally high accuracy for patients with ovarian cancer and cervical cancer.<sup>14–16</sup> The purpose of this study is to further investigate the effects of these automatic segmentation algorithms on the accuracy of radiomics modeling in predicting the status of LNM for patients with early stage cervical cancer preoperatively.

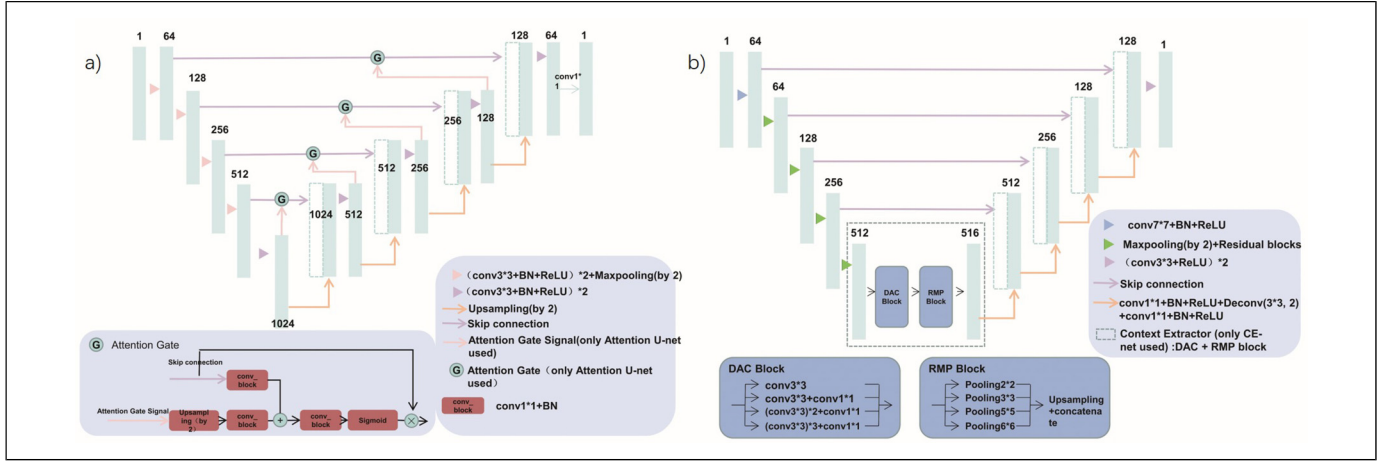
## Materials and Methods

### Patients and Images

By retrieving electronic medical records, cervical cancer patients who underwent radical hysterectomy and pelvic lymphadenectomy in the authors' hospital from January 2014 to September 2018 were retrospectively reviewed. All patients met the following inclusion criteria: (1) patients who underwent radical hysterectomy and systematic pelvic lymph node dissection; (2) cervical cancer and pelvic LNM status were confirmed by histopathology; (3) patients should have standard ultrasonic examination for 2 weeks before the hysterectomy; and (4) clinical characteristics were available. Patients were excluded from this study for the following reasons: (1) without complete clinical data; (2) with other malignancies or combined malignancies; (3) had preoperative chemotherapy or radiotherapy; and (4) tumor's size and shape were incomplete in US images were excluded. Clinicopathologic features of all patients were obtained from medical records including age, histological subtype, International Federation of Gynecology and Obstetrics (FIGO) stage, etc. To reduce the effects of different US machines, only patients with US images acquired using Philips IU22 (Philips Medical Systems, Best, The Netherlands) with 5 to 14 MHz linear probes were included in this study. This retrospective study was conducted following the Declaration of Helsinki and approved by the Ethics Committee in Clinical Research (ECCR) of Wenzhou Medical University First Affiliated Hospital (ECCR No. 2019059). The requirement of written informed consent was waived by the ECCR due to the retrospective nature of this study with confirmation of patient data confidentiality.

### Manual and Automatic Segmentation

There were 8 to 15 standard US images collected for each patient. Manual segmentation on US images was conducted and confirmed by two senior radiologists with several years of experience in gynecological imaging using the LIFEx package (<http://www.lifexsoft.org>).<sup>17</sup> Manual contours were regarded as the ground truth for automatic segmentation models' training and validation. Automatic segmentation was performed with the



**Figure 1.** (a) The architecture of attention U-net: U-net with added attention gate (AG); (b) U-net with Resnet: the backbone of U-net is replaced by Resnet; context encoder network (CE-net): context extractor consisted of dense atrous convolution (DAC) and residual multi-kernel pooling (RMP) block was added into U-net with Resnet.

U-net scheme and its multiple variations, U-net with Resnet, context encoder network (CE-net), and attention U-net.

As shown in Figure 1a, U-net is a symmetrical U-shaped model consisting of an encoder—decoder architecture. The left side encoder is a down-sampling used to get feature map, similar to a compression operation, while the right side decoder is an up-sampling used to restore the encoded features to the original image size and to output the results. Skip connection was added to encoder–decoder networks to concatenate the features of high- and low levels together.<sup>18</sup> Compared with U-net, attention U-net adds an attention gate (AG) which filters the features propagated through the skip connections. Feature selectivity in AG is achieved using contextual information (gating) extracted in coarser scales,<sup>19</sup> as shown in Figure 1a. U-net with Resnet uses Resnet as a fixed feature backbone encoder to deepen the layers of the network and to solve the vanishing gradient,<sup>20</sup> a new technique called skip connections has been used to skip training from a few layers and connects directly to the output. CE-net adds a context extractor consisting of dense atrous convolution (DAC) block and residual multi-kernel pooling (RMP) block into U-net with Resnet,<sup>21</sup> which consists of three major parts: a feature encoder, a feature decoder module, and a context extractor, as shown in Figure 1b. Image clipping was performed for all US images before network training which make sure the robustness of the training model. The clipping box should not exceed the image edge and covers the tumor. In the training phase, Adam algorithm was used as a optimizer with its learning rate setting to  $2e-4$  because of its straightforward implementation and computational efficiency. The input image was four batch sizes and the number of epoch was 120. The loss function is

$$Loss = \lambda BCE + (1 - \lambda) * Loss_{DSC} \quad (1)$$

where BCE is the Binary Cross Entropy,  $Loss_{DSC}$  is the loss of dice similarity coefficient (DSC) and  $\lambda$  was set to 0.8 in this study. A typical segmentation results from manual delineation,

U-net, attention-net, CE-net, and Resnet models were shown in Figure 2. The construction and accuracy of these automatic segmentation models for US images had been reported in previous studies.<sup>14, 15</sup>

### Feature Extraction and Model Building

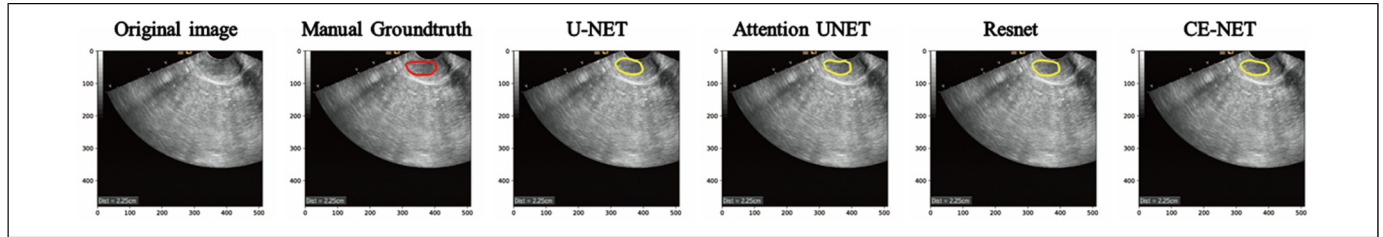
Intensity normalization was performed in US images before feature extraction to transform arbitrary gray intensity values into a standardized intensity range. A total of 449 radiomics features were extracted using Pyradiomics (version 3.0.1; <https://pypi.org/project/pyradiomics>).<sup>22</sup> Intraclass correlation coefficients (ICC) (two-way random effects, absolute agreement, and single rater/measurement) were calculated to select the reliable radiomics features (ICC > 0.9) that resulted from automatic segmentations by comparing them with those resulted from manual segmentation.<sup>23</sup>

The dimension of radiomics features extracted from manual and automatic segmentations were further reduced by Mann–Whitney U tests with  $P < .05$  as potentially informative features to classify positive and negative LNM. Then, the least absolute shrinkage and selection operator (LASSO) and the ridge regression were applied to select the optimal features in the training cohort using a 10-fold cross-validation.<sup>24</sup> The coefficient  $\lambda$  of the elastic net was tuned to achieve minimum  $SD$  and maximum area under the curve (AUC) of receiver operating characteristic (ROC) curves during feature selection.

Radiomics signature was built for each patient by combining the selected features linearly using the logistic regression model to predict the lymph node status. AUC was quantified to further assess the performance of radiomics signature that resulted from different segmentations in both the training and validation cohorts.

### Statistical Analysis

General statistical analyses were performed in SPSS Statistics (version 20.0.0, SPSS Inc.) and R analysis platform (version



**Figure 2.** A typical segmentation results from manual delineation, U-net, attention-net, context encoder network (CE-net), and Resnet models

**Table 1.** The Characteristics of Enrolled Patients in the Training and Validation Data Sets.

Characteristics	Training set			Validation set		
	LNM + (n = 18)	LNM -(n = 93)	P	LNM + (n = 11)	LNM -(n = 26)	P
Age						
Average	48.72	51.59	.39	49.45	53.23	.40
Median	50.5	51		50	52	
Range	33 to 60	33 to 75		30 to 65	34 to 78	
SD	8.14	10.52		11.69	11.18	
Histological type						
Squamous cell carcinoma	14	86	.08	8	26	.02
Adenocarcinoma	3	4		2	0	
Adenosquamous cell carcinoma	1	2		1	0	
Tumor stage						
I	8	43	.89	2	13	.15
II	10	50		9	13	

Note: (1) *P*-value is calculated from the univariate association test between subgroups. (2) Fisher's exact test and chi-square test were used for categorized variables. (3) Two-sample *t*-test was used for continuous variables. LNM = lymph node metastasis; - = negative; + = positive.

3.6.0, MathSoft, Auckland, New Zealand). Key features selection and logistic regression model building were performed with the “glmnet” package. For continuous clinical variables, a two-sample *t*-test was used to assess the equality of variances between positive and negative LNM groups. For categorical variables, Fisher's exact test and chi-square test were used to test the difference between groups. For all tests,  $p < .05$  was considered as statically significant.

## Results

A total of 148 patients with early stage cervical cancer was enrolled in this study with a median age of 51 years (range from 30 to 78 years). Most of the patients were squamous cell carcinoma (90.5%). Patients were randomly divided into training (111) and testing set (37) at a ratio of 3:1. The patients with positive and negative LNM in the training and testing cohorts were 18 versus 93, and 11 versus 26, respectively. Detailed characteristics of enrolled patients was shown in

Table 1. The DSC of U-net, attention U-net, CE-net, and U-net with Resnet were 0.88, 0.89, 0.88, and 0.90, respectively.

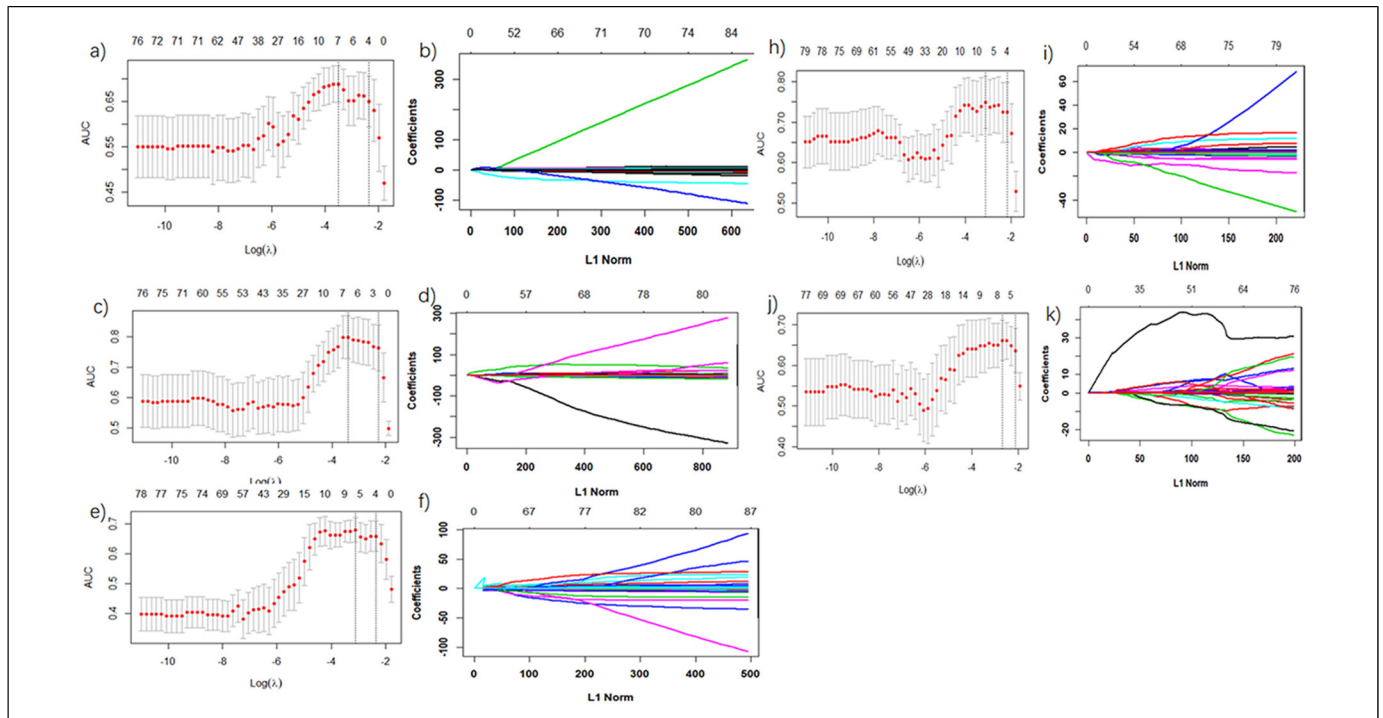
A total of 449 radiomics features were extracted from manual and automatic segmented ROIs with Pyradiomics. Features with an ICC > 0.9 were all 257 (57.2%) from manual contours and automatic segmented contours with four deep learning models. Detailed radiomics features selected after ICC were shown in Supplementary Excel S1. The number of radiomics features selected by Mann-Whitney U test is 86, 85, 90, 92, 87, and after LASSO is 7, 7, 8, 9, and 7 for manual, attention U-net, CE-Net, Resnet, and U-net ROIs, respectively. Figure 3 shows the selection of LNM-associated radiomics features using the elastic net method for manual, U-net, CE-Net, Resnet, and attention U-net segmentations, respectively. Detailed radiomics features selected after each analysis were shown in Supplementary Excel S2 and S3.

The AUCs of training models with radiomics features extracted from manual, attention U-net, CE-net, Resnet, and U-net were 0.763, 0.782, 0.761, 0.774, and 0.794, respectively, as shown in Figure 4a. The AUCs of validation models with radiomics features extracted from manual, attention U-net, CE-net, Resnet, and U-net were 0.692, 0.755, 0.696, 0.689, and 0.710, respectively, as shown in Figure 4b. The AUC numbers of models with automatic segmentation features were 9.10%, 0.58%, -0.43%, and 2.60% higher compared with models with manual contoured features for attention U-net, CE-net, Resnet, and U-net, respectively.

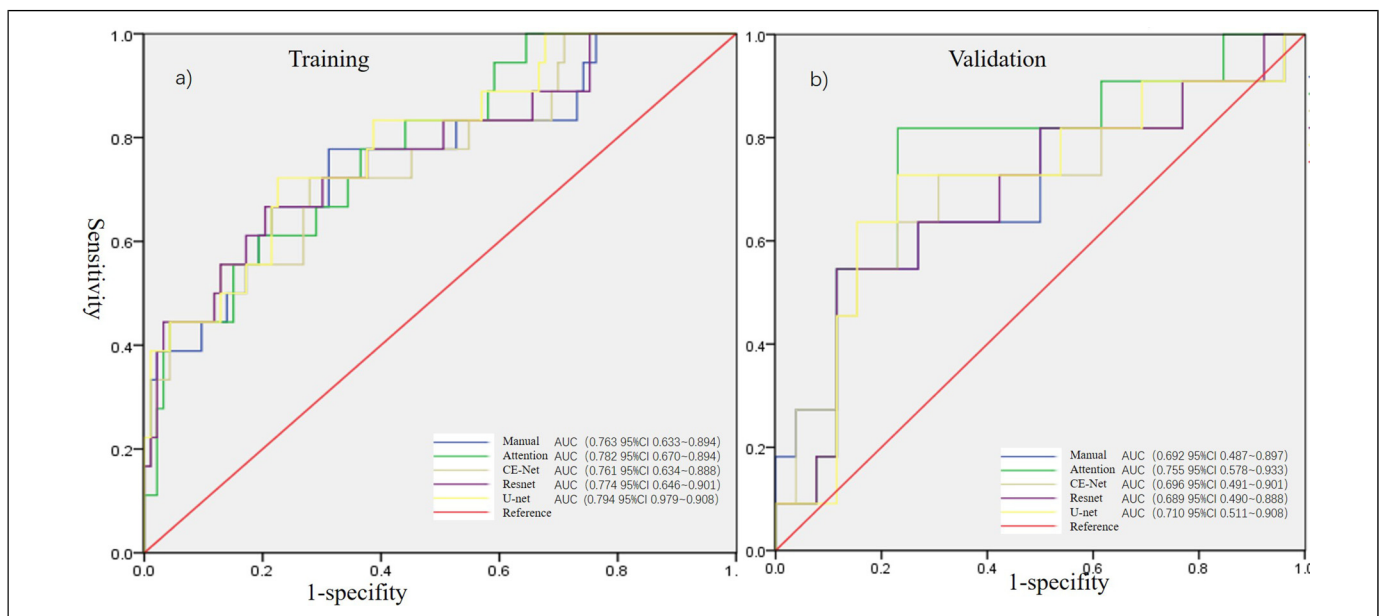
## Discussion

The effects of different deep learning-based automatic segmentations on the radiomics models for the prediction of LNM for patients with early stage cervical cancer were investigated in this study. An AUC number of 9.10% differences was observed for models with automatic segmentation features in comparison with models with manual segmentation features.

In radiomics analysis, tumor segmentation is a fundamental step as it determines which region will be analyzed. However, since there is no gold standard for tumor segmentation, tumor segmentation is challenging, no matter whether manual, semi-automatic, or automatic segmentation methods are applied.<sup>9</sup> Especially, studies reported that radiomics features can be changed by different delineation. van Velden et al demonstrated that some features were sensitive to the change of delineation



**Figure 3.** Selection of lymph node metastasis (LNM)-associated radiomics features using the elastic net method: (a, c, e, h, j) Tuning parameter ( $\lambda$ ) in the elastic net used 10-fold cross-validation via maximum area under the curve (AUC); (b, d, f, i, k) The coefficient profiles of radiomics features against the L1 norm (inverse proportional to  $\log \lambda$ ) for manual, U-net, CE-Net, Resnet, and attention U-net segmentations, respectively.



**Figure 4.** The area under the curve (AUCs) of models with radiomics features extracted from manual, attention U-net, CE-net, Resnet, and U-net in (a) training cohort and in (b) validation cohort.

with two baseline whole-body PET/CT scans from 11 patients with non-small cell lung cancer (NSCLC).<sup>25</sup> Leijenaar et al reported that the percent stable radiomics features using semiautomatic and manual segmentation were 71% and 91% in a test–

retest cohort of 11 NSCLC patients and in an inter-observer cohort of 23 NSCLC patients, respectively.<sup>26</sup> However, only 57.2% of radiomics features were reproducible ( $ICC > 0.9$ ) in this study from manual and automatic segmentations. This



may be due to the low quality of US image data. There are many characteristic artifacts in the US images, such as attenuation, speckle, shadows, signal dropout that make the segmentation in US images complicated. The advances in transducer design, spatial/temporal resolution, digital systems, etc., will certainly improve the quality of US image data and improve the outcome radiomics features.

One of the main motivations for using automatic segmentation is to reduce the labor intensity of physician and to increase the reproducibility and reliability of segmentation by decreasing the intra-reader and inter-reader variability. Previous studies using python and Pyradiomics for automatic segmentation and feature extraction demonstrated that U-net models are able to achieve a DSC around 0.88 to 0.90 in US images for cervical cancer with an average ICC around 0.99.<sup>15</sup>

In this study, radiomics features from the attention U-net showed best performance in the LNM prediction model with a lowest discrepancy between training and validation AUCs. The reported AUC of 0.755 was close to the result of 0.77 in a previous study for preoperative LNM prediction for patients with early stage cervical cancer.<sup>8</sup> However, the AUCs of models among manual and different automatic segmentations were different with a largest difference of 9.11% (ranges from 0.689 to 0.755). This indicated that automatic segmentation is an important source of variability for radiomics modeling.

This study demonstrated that the variability of radiomics features and models could be caused by manual segmentation, automatic segmentation, as well as by different feature extracting methods. Similar, studies reported that human-derived radiomics methods introduced human bias into radiomics process.<sup>27</sup> Variation in imaging and preprocessing techniques for feature extraction also affected the reproducibility.<sup>28</sup> With the advent of deep learning methods, more and more effects had been focused on using deep learning methods to increase the generalizability and accuracy while reducing potential bias for radiomics models. The full tumor segmentation was replaced with approximate localization with the change of ROI definition based on a single point within the tumor volume in deep learning radiomics applications so as to minimize the need for human input,<sup>29, 30</sup> Additionally, relevant radiographic features could be learned automatically using deep learning methods without the need of previous definition by researchers.<sup>31</sup>

There are a few limitations to our study. First, we found that the repeatability and reproducibility of US images radiomics features extracted from different machines were unsatisfactory so only one kind of US machine was analyzed in this work. In addition, external validation studies that use a multi-institutional data are necessary to confirm our findings.

## Conclusions

The performance of LNM prediction models with US radiomics features extracted from manual segmentation, and deep learning-based automatic segmentations were invested in this study. The reliability and reproducibility of radiomics features,

as well as the performance of radiomics models, were affected by manual segmentation and automatic segmentations.

## Acknowledgments

This work was supported in part by Wenzhou Municipal Science and Technology Bureau (Y20190183), Zhejiang Engineering Research Center of Intelligent Medicine (2016E10011), and the National Natural Science Foundation of China (No. 11675122).

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by Wenzhou Municipal Science and Technology Bureau (Y20190183), Zhejiang Engineering Research Center of Intelligent Medicine (2016E10011), and Wenzhou Key Lab of Radiation Oncology and Translational Research (2021100848).

## Ethical Approval

This retrospective study was conducted following the Declaration of Helsinki and approved by the Ethics Committee in Clinical Research (ECCR) of the authors' hospital (ECCR No. 2019059). The requirement of written informed consent was waived by the ECCR due to the retrospective nature of this study with confirmation of patient data confidentiality.

## ORCID iD

Xiance Jin  <https://orcid.org/0000-0002-4117-5953>

## Supplemental Material

Supplemental material for this article is available online.

## References

1. Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging*. 2006;25(8):987-1010.
2. Bridal SL, Correas JM, Saied A, Laugier P. Milestones on the road to higher resolution, quantitative, and functional ultrasonic imaging. *Proc IEEE*. 2003;91(10):1543-1561.
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577.
4. Guo Y, Hu Y, Qiao M, et al. Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma. *Clin Breast Cancer*. 2018;18(3):e335-e344.
5. Cobo T, Bonet-Carne E, Martínez-Ferrón M, et al. Feasibility and reproducibility of fetal lung texture analysis by automatic quantitative ultrasound analysis and correlation with gestational age. *Fetal Diagn Ther*. 2012;31(4):230-236.
6. Bonet-Carne E, Palacio M, Cobo T, et al. Quantitative ultrasound texture analysis of fetal lungs to predict neonatal respiratory morbidity. *Ultrasound Obstet Gynecol*. 2015;45(4):427-433.
7. Liu T, Zhou S, Yu J, et al. Prediction of lymph node metastasis in patients with papillary thyroid carcinoma: a radiomics method

- based on preoperative ultrasound images. *Technol Cancer Res Treat*. 2019;18:1533033819831713.
8. Jin X, Ai Y, Zhang J, et al. Noninvasive prediction of lymph node status for patients with early-stage cervical cancer based on radiomics features from ultrasound images. *Eur Radiol*. 2020;30(7):4117-4124.
  9. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30(9):1234-1248.
  10. Owens CA, Peterson CB, Tang C, et al. Lung tumor segmentation methods: impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLoS One*. 2018;13(10):e0205003.
  11. Lee M, Woo B, Kuo MD, et al. Quality of radiomic features in glioblastoma multiforme: impact of semi-automated tumor segmentation software. *Korean J Radiol*. 2017;18(3):498-509.
  12. Jalalian A, Mashohor S, Mahmud HR, Saripan M, Ramli A, Karasfi B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clin Imaging*. 2013;37(3):420-426.
  13. Liu Z, Wang S, Dong D, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*. 2019;9(5):1303-1322.
  14. Jin J, Zhu H, Zhang J, et al. Multiple U-net-based automatic segmentations and radiomics feature stability on ultrasound images for patients with ovarian cancer. *Front Oncol*. 2021;10:614201.
  15. Jin X, Jin J, Xie C. The Accuracy and Radiomics Feature Effects of Multiple U-Net Based Automatic Segmentation Models for Transvaginal Ultrasound Images of Cervical Cancer. AAPM annual meeting, 2021.
  16. Kurata Y, Nishio M, Moribata Y, et al. Automatic segmentation of uterine endometrial cancer on multi-sequence MRI using a convolutional neural network. *Sci Rep*. 2021;11(1):14440. Published 2021 Jul 14. doi:10.1038/s41598-021-93792-7.
  17. Christophe N, Fanny O, Sarah B, et al. Lifex: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res*. 2018;78(16):4786-4789.
  18. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *MICCAI*. 2015. DOI: 10.1007/978-3-319-24574-4\_28
  19. Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: learning where to look for the pancreas. 2018; arXiv:1804.03999.
  20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2016. DOI: 10.1109/CVPR.2016.90
  21. Gu Z, Cheng J, Fu H, et al. Ce-net: context encoder network for 2d medical image segmentation. *IEEE Trans Med Imaging*. 2019;38(10):2281-2292.
  22. Fedorov A, Parmar C, Hosny A, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104-e107.
  23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
  24. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.
  25. Velden FHPV, Kramer GM, Frings V, et al. Repeatability of radiomic features in non-small-cell lung cancer [18 F]FDG-PET/CT studies: impact of reconstruction and delineation. *Mol Imag Biol*. 2016;18(5):788-795.
  26. Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol (Madr)*. 2013;52(7):1391-1397.
  27. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One*. 2015;10(5):e0124165.
  28. Traverso A, Wee L, Dekker A. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys*. 2018;102(4):1143-1158.
  29. Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res*. 2019;25(11):3266-3275.
  30. Hosny A, Parmar C, Coroller TP, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Med*. 2018;15(11):e1002711.
  31. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.