Automatic Segmentation of Clinical Target Volume and Organs-at-Risk for Breast Conservative Radiotherapy Using a Convolutional Neural Network

Zhikai Liu
Fangjie Liu,*,†
Wanqi Chen,*,†
Yinjie Tao
Xia Liu
Fuquan Zhang
Jing Shen
Hui Guan
Hongnan Zhen
Shaobin Wang
Qi Chen
Xiaorong Hou,†

†Department of Radiation Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, 100730, People’s Republic of China; ‡Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, 510060, People’s Republic of China; †MedMind Technology Co., Ltd., Beijing, 100055, People’s Republic of China

*These authors contributed equally to this work

Objective: Delineation of clinical target volume (CTV) and organs at risk (OARs) is important for radiotherapy but is time-consuming. We trained and evaluated a U-ResNet model to provide fast and consistent auto-segmentation.

Methods: We collected 160 patients’ CT scans with breast cancer who underwent breast-conserving surgery (BCS) and were treated with radiotherapy. CTV and OARs were delineated manually and were used for model training. The dice similarity coefficient (DSC) and 95th percentile Hausdorff distance (95HD) were used to assess the performance of our model. CTV and OARs were randomly selected as ground truth (GT) masks, and artificial intelligence (AI) masks were generated by the proposed model. Two clinicians randomly compared CTV score differences of the contour. The consistency between two clinicians was tested. Time cost for auto-delineation was evaluated.

Results: The mean DSC values of the proposed method were 0.94, 0.95, 0.94, 0.96, 0.96 and 0.93 for breast CTV, contralateral breast, heart, right lung, left lung and spinal cord, respectively. The mean 95HD values were 4.31mm, 3.59mm, 4.86mm, 3.18mm, 2.79mm and 4.37mm for the above structures, respectively. The average CTV scores for AI and GT were 2.89 versus 2.92 when evaluated by oncologist A (P=0.612), and 2.75 versus 2.83 by oncologist B (P=0.213), with no statistically significant differences. The consistency between two clinicians was poor (kappa=0.282). The time for auto-segmentation of CTV and OARs was 10.03 s.

Conclusion: Our proposed model (U-ResNet) can improve the efficiency and accuracy of delineation compared with U-Net, performing equally well with the segmentation generated by oncologists.

Keywords: clinical target volume, organ at risk, auto-segmentation, breast cancer radiotherapy, clinical evaluation

Key Points
1. A U-ResNet model can auto-delineate for breast conservative radiotherapy.
2. CTV and OARs generated by our model can meet the clinical requirements.
3. AI assistance can effectively improve consistency in contouring radiotherapy workflow.

Introduction
Breast cancer (BC) is one of the most common cancers for women throughout the world. Breast radiotherapy after breast-conserving surgery (BCS) is an essential treatment for early breast cancer patients. Radiotherapy of tumors requires
accurate, individualized contouring of clinical target volume (CTV) and organs at risk (OARs) to deliver high radiation doses to the target and to spare healthy tissues.\textsuperscript{4} Therefore, computer-assisted automatic segmentation techniques are highly desired and useful for relieving radiation oncologists from labor-intensive work as well as reducing considerable inter- and intra-observer variability in delineation of the regions of interest (ROIs).\textsuperscript{5,6}

Current automatic approaches can be generally categorized into two groups: atlas-based auto-segmentation (ABAS) and convolutional neural network (CNN) based segmentation. Acceptable results have been reported using ABAS for OARs in head and neck cancer and prostate cancer.\textsuperscript{7–9} However, CTV is not a region with clear boundaries but includes tissues of potential tumor or subclinical diseases that are barely detectable in CT images.\textsuperscript{10} Moreover, the inconsistencies in body shape, organ size, and density of mammary glandular tissue remain large from person to person.\textsuperscript{11,12} Therefore, various kinds of CNN models,\textsuperscript{13–16} have been presented for different cancers,\textsuperscript{8,16–20} showing better performance than ABAS.

A deep dilated residual network (DD-ResNet) was previously proposed by Men et al\textsuperscript{16} to perform automatic breast CTV contouring. A 0.91 DSC was reported for both the right and left breast CTV, but no clinical evaluation was performed. Moreover, this method was focused on CTV contouring; the OARs were not considered.

Here, we constructed a new CNN model based on the 2D U-Net model to solve the large inconsistencies between source and target image, even with a scarce amount of labelled training data. The proposed model was trained and then compared against U-Net. The accuracy and effectiveness were evaluated by both performance metrics and qualified radiation oncologists.

**Materials and Methods**

**Data Acquisition**

CT scans of patients with early-stage BC who underwent BCS in Peking Union Medical College Hospital were collected from January 2019 to December 2019. This study was approved by the Institutional Review Board of Peking Union Medical College Hospital. Informed consent/assent from the patient and/or parent/guardian, as appropriate, was obtained before enrollment. This study was conducted in accordance with the Declaration of Helsinki. The inclusion criteria are as follows: (1) Patients who were diagnosed with early-stage BC and underwent breast conservative surgery. (2) All the patients met the indication for radiotherapy and received whole-breast irradiation. Patients who underwent axilla or supraclavicular lymph nodes radiotherapy were excluded.

In total, 12,640 CT slices were collected from 160 patients; 79 patients had left-sided BC and the remainder had right-sided BC. All the CT scans followed the digital imaging and communications in medicine (DICOM) protocol and were scanned using a Philips Brilliance Big Bore CT scanner. CT images were reconstructed using a matrix size of 512×512 and a thickness of 5 mm. The pixel spacing of the data was 1.1543 mm × 1.1543 mm.

Contouring of the CTV and OARs (contralateral breast, lungs, heart, and spinal cord) were delineated manually by trained radiation oncologists following the European Society for Radiotherapy and Oncology (ESTRO)\textsuperscript{21} and the Radiation Therapy Oncology Group (RTOG)\textsuperscript{22} protocols. The specific sketching standards for CTV are shown in Table 1. All the contours were reviewed and approved by two professional radiation oncologists with more than 10 years’ experience in our center.

### Network Architecture

Our model, called U-ResNet, is originated from the 2D U-Net model, which is composed of encoder and decoder paths. To conduct the segmentation task for BC radiotherapy, especially for the CTV segmentation, a deep network should be added to the U-Net to extract features as different abstraction levels. At the same time, the vanishing gradients of deep convolutional networks should be avoided. Therefore, ResNet is used as the encoder part. It encodes low-, middle- and high-level features and passes these features to the decoder part via four shortcut paths.

<table>
<thead>
<tr>
<th>Borders</th>
<th>Cranial</th>
<th>Caudal</th>
<th>Ventral</th>
<th>Dorsal</th>
<th>Lateral</th>
<th>Medial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual breast</td>
<td>Upper border of palpable/visible breast tissue; maximally up to the inferior edge of the sternoclavicular joint</td>
<td>Most caudal CT slice with visible breast</td>
<td>5 mm below the skin surface</td>
<td>Major pectoral muscle or costae and intercostal muscles where no muscle</td>
<td>Lateral breast fold; anterior to the lateral thoracic artery</td>
<td>The edge of sternum</td>
</tr>
</tbody>
</table>

**Table 1** The Standard Delineation of CTV After BCS
connections. In the decoder part, the upscaling is achieved using nearest neighbour interpolation, followed by a convolutional layer and a residual block. In this way, multiple-level features in the encoder and decoder parts are concatenated. The overall architectures of DD-ResNet and our proposed method are shown in Figure 1. DD-ResNet has no shortcut connections between the encoder and the decoder. The output of the sum layer was interpolated to the original size with a factor of 8, which may result in information loss.

The breast is a continuous and smooth surface. A 2D architecture may result in a rough segmentation result in a 3D view. To obtain the 3D information of CT scans, the network is designed as a 2.5D architecture by assigning three adjacent slices into three channels as the input.

Implementation Details
The dataset, composed of 160 patients, was randomly assigned in 8:1:1 to three cohorts: 1) a training set of 128 patients was used to construct the segmentation model, 2) a validation set of 16 patients was used to optimize the parameters and 3) a testing set of 16 patients was used to obtain artificial intelligence-generated contouring for performance assessment. During the testing phase, all the CT slices of the 16 testing cases were tested individually.

We constructed our model using Python 3.6 and PyTorch 1.0. An Adam optimization algorithm was used for the optimization. The learning rate was 0.001. We trained and evaluated our model with a GTX 1080 GPU. The proposed model was trained over 50 circles to select the best model according to the lowest validation loss score. The convolutional layers are initialized using Xavier Uniform Initialization, and batch normalization layers were added after convolution layers to improve the training speed and to prevent overfitting.

Performance Measurement
Performance of the proposed method was evaluated using the dice similarity coefficient (DSC) and the 95th percentile Hausdorff distance (95HD) to quantify the results. The mean and standard deviation were also calculated.

The DSC was used to measure the spatial overlap between AI and GT contours, which is defined in Equation (1).

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

where $A$ represents the volume of the human-generated contour; $B$ is the volume of an AI contour; and $A \cap B$ is the intersect volume that $A$ and $B$ have in common. The DSC value was between 0 and 1 (0 = no overlap, 1 = complete overlap).

The 95HD is defined as follows:

![Figure 1 Architecture of (A) deep dilated convolutional neural network (DDCNN), (B) our proposed network, and (C) the residual block used in decoder part of our network.](https://doi.org/10.2147/CMAR.S330249)
where \( A \) represents the human-generated contour; \( B \) is the AI contour. \( || . || \) means the Euclidean norm of the points of \( A \) and \( B \). The 95HD means the 95 percentile maximum mismatch between \( A \) and \( B \). When the 95HD value decreases, the similarity between \( A \) and \( B \) increases.

### Oncologist Evaluation

#### OARs Evaluation

Considering that evaluation metrics cannot provide a comprehensive insight into whether the contours need to be modified in clinical practice, another 20 cases in clinical practice from our center were randomly collected. Each case was delineated with GT and AI contours for OARs and then distributed to two radiation oncologists with more than 10 years of clinical experience for further evaluation. Each slice was carefully evaluated, and the results were graded in four levels: 3 points (no need to be edited), 2 points (the number of layers need to be edited \( \leq 4 \)), 1 point (the number of layers need to be edited \( \geq 4 \)) and 0 point (not acceptable).

#### CTV Evaluation

CTV segmentations generated by AI and GT were also evaluated blindly slice by slice. The test data contained 10 patients and 650 slices in total (AI: 327 slices vs GT: 323 slices). The representative results were also graded on four levels: 3 points (acceptable for subsequent treatment), 2 points (Minor Revision), 1 point (Major Revision) and 0 point (Not Acceptable for treatment). When the score \( \geq 2 \), it was defined as suitable for clinical application.

Furthermore, to verify the consistency of the judgment of two oncologists, we collected the CTV score of each slice evaluated by two oncologists, constituting a total of 650 slices of data sets. The data were classified into the same group if the slice was evaluated by two oncologists with the same CTV score. We calculated the weighted kappa coefficient to analyze for consistency.

### Time Cost

Processing time was measured for the AI tool and pre- and post-AI assistance in the delineation of CTV and OARs for BC radiotherapy.

### Statistical Analysis

The Wilcoxon matched-pairs signed-rank test was used to compare DSC and 95HD between our proposed model and U-Net and the differences between the two oncologists during the evaluation of CTV and OARs segmentation. McNemar test and kappa test were used to assess the consistency of the two oncologists. Statistical significance was set at two-tailed \( P < 0.05 \).

### Results

#### Performance of U-ResNet and Comparison with U-Net

The median age for all the 160 patients in dataset was 49 [42, 58]. The average CTV volume was 494.41 \( \pm \) 198.51 cm\(^3\).

For CTV segmentation, the average DSC values of U-ResNet and U-Net were 0.94 vs. 0.93 \((P=0.001)\), and the average 95HD value was 4.31 mm vs 4.88 mm separately \((P=0.030)\). Both differences were statistically significant, implying better accuracy of CTV contouring by U-ResNet.

Among all OARs, significant differences between U-ResNet and U-Net were achieved for the spinal cord (DSC: 0.93 vs 0.92 \((P=0.015)\), 95HD: 4.37 mm vs 5.07 mm \((P=0.003)\)) and the contralateral breast (DSC: 0.95 vs 0.93 \((P<0.001)\), 95HD: 3.59 mm vs 4.15 mm \((P=0.010)\)). The right lung contouring also displayed a statistically significant difference in 95HD (3.18 mm vs 2.98 mm \((P=0.041)\)).

The results of the comparison are summarized in Table 2 and Figure 2.

Figure 3 shows the visualization segmentation samples in GT, U-Net and U-ResNet, respectively. The auto-segmented contours with U-ResNet were in good concordance with the GT contours.

### Oncologist Evaluation

Tables 3 and 4 show the oncologist evaluation results of OAR and CTV contours. Scores \( \geq 2 \) were defined as suitable for clinical application. When using our grading criteria for contour evaluation, the majority of AI- and GT-generated OAR contours were deemed acceptable by the experts. Only one contour (5\%) of the heart was assessed to require major revision by oncologist A.

Regarding CTV contours, 99.4\% of those generated by AI were clinically acceptable by oncologist A, compared with 98.1\% of GT segmentations. For oncologist B, the
results were 99.4% for both methods. The average CTV scores for AI and GT were 2.89 vs 2.92 when evaluated by oncologist A ($P=0.612$) and 2.75 vs 2.83 by oncologist B ($P=0.213$), with no statistical differences.

Wilcoxon matched-pairs test was performed for the evaluation of the two oncologists for AI and GT contours separately. The results indicated that the average score of oncologist A was higher than that of oncologist B in AI contours, with a significantly statistical difference ($P=0.009$ for AI contours and $P=0.314$ for GT contours).

The comparison of the average CTV scores evaluated by two oncologists is shown in Figure 4.

The evaluation results of the two oncologists were further analyzed for consistency. We independently collected the CTV scores of all the 650 slices generated by both AI and GT segmentations, and used the data to calculate weighted kappa coefficient. The results are shown in Table 5. Of all the 650 slices, 532 (81.8%) were evaluated with the same CTV score. The consistency between two oncologists was not good (kappa=0.282).

### Table 2 DSC and 95HD for CTV and All OARs

<table>
<thead>
<tr>
<th>ROI</th>
<th>CTV</th>
<th>Contra-Lateral Breast</th>
<th>Spinal Cord</th>
<th>Heart</th>
<th>Right Lung</th>
<th>Left Lung</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSC±STD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U-ResNet</td>
<td>0.94±0.019</td>
<td>0.94±0.016</td>
<td>0.93±0.024</td>
<td>0.94±0.023</td>
<td>0.96±0.017</td>
<td>0.96±0.025</td>
</tr>
<tr>
<td>U-Net</td>
<td>0.93±0.022</td>
<td>0.93±0.020</td>
<td>0.92±0.029</td>
<td>0.93±0.026</td>
<td>0.96±0.018</td>
<td>0.96±0.024</td>
</tr>
<tr>
<td>P value</td>
<td>0.001</td>
<td>&lt; 0.001</td>
<td>0.015</td>
<td>0.066</td>
<td>0.257</td>
<td>0.942</td>
</tr>
<tr>
<td>95HD±STD (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U-ResNet</td>
<td>4.31±1.76</td>
<td>3.59±1.56</td>
<td>4.37±2.13</td>
<td>4.86±1.48</td>
<td>3.18±1.64</td>
<td>2.79±1.62</td>
</tr>
<tr>
<td>U-Net</td>
<td>4.88±1.64</td>
<td>4.15±1.28</td>
<td>5.07±2.26</td>
<td>5.17±2.32</td>
<td>2.89±1.26</td>
<td>2.76±1.52</td>
</tr>
<tr>
<td>P value</td>
<td>0.030</td>
<td>0.010</td>
<td>0.003</td>
<td>0.559</td>
<td>0.041</td>
<td>0.781</td>
</tr>
</tbody>
</table>

Figure 2 Boxplots obtained for DSC and 95HD analyses of U-ResNet and U-Net. (A) DSC analyses, (B) 95HD analyses.

Figure 3 CTV and OAR contours generated by (A) GT, (B) U-ResNet, and (C) U-Net after breast conservative surgery.
Time Cost

Time for auto-segmentation of CTV and OARs with U-ResNet was 10.03 s, compared with 20 minutes and 30 minutes by experienced oncologists. With AI assistance, the delineation time can be reduced to 10 minutes and 5 minutes for CTV contouring and OAR contouring, respectively.

Discussion

Accurate and consistent delineation of CTV and OARs is a basic requirement for contemporary radiotherapy planning, while it is also the most burdensome step in the radiotherapy workflow.\textsuperscript{25,26} Manual delineation is a time-consuming process and has considerable inter- and intra-observer variability in anatomical contouring.\textsuperscript{27,28} In recent years, computer-assisted automatic segmentation techniques have made great breakthroughs in increasing reliability and accuracy as well as in relieving radiation oncologists from time-intensive contouring.

Among these automatic methods, CNNs are the most advanced method available for medical images and have shown better performance than atlas-based methods.\textsuperscript{29} As a completely end-to-end network, only a limited amount of data is required. Among all the CNN-based models, U-Net has been a remarkable and the most popular deep network. However, it may lose some abstract information with relatively lower level convolutional layers. To alleviate the disparity between the encoder-decoder features, we have trained and evaluated a new model based on the 2D U-Net model.

Our model was based on U-Net architecture. Therefore, the performance of the proposed model was also compared with U-Net. During the training stage, U-Net had the same training configuration as that of the proposed model. This was the first study assessing the performance of both CTV and OARs segmentation for breast-conservative radiotherapy. The analysis revealed that U-ResNet algorithm outperformed the U-Net algorithms. Moreover, U-ResNet performed well, with good agreement to the segmentations contoured manually by oncologists.

In our previous work, we have demonstrated that CNN architecture could facilitate the delineation of CTV for radical mastectomy radiotherapy.\textsuperscript{30} In this piece of work, we focused on both CTV and OARs for breast conservative radiotherapy. We also made some modifications in CNN architecture. Firstly, the whole encoder part was completely replaced with the ResNet34 instead of a residual block. Moreover, we also proposed a self-adaptive weighted cross entropy loss function to tackle our multi-class structures segmentation problem, so that

| Table 3: Evaluation for CTV and OARs by Oncologist A |
|---|---|---|---|---|---|---|
| Scores | CTV | Contra-Lateral Breast | Spinal Cord | Heart | Right Lung | Left Lung |
| | AI | GT | AI | GT | AI | GT | AI | GT | AI | GT | AI | GT |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0.6% | 1.9% | 0 | 0 | 0 | 0 | 0 | 5% | 0 | 0 | 0 | 0 |
| 3 | 7% | 7.4% | 30% | 35% | 5% | 20% | 20% | 20% | 10% | 0 | 0 | 0 |
| 4 | 92.4% | 90.7% | 70% | 65% | 95% | 80% | 80% | 80% | 90% | 100% | 0 | 0 |
| P value | 0.612 | 0.414 | 0.083 | 0.527 | 1.00 | 0.157 |

| Table 4: Evaluation for CTV and OARs by Oncologist B |
|---|---|---|---|---|---|---|
| Scores | CTV | Contra-Lateral Breast | Spinal Cord | Heart | Right Lung | Left Lung |
| | AI | GT | AI | GT | AI | GT | AI | GT | AI | GT | AI | GT |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0.6% | 0.6% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 23.5% | 14.9% | 10% | 15% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 75.9% | 85.5% | 90% | 85% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| P value | 0.612 | 0.655 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
we could train and predict all the CTV and OARs in one time, introducing greater efficiency. Notably, however, an automatic segmentation model of CTV may not take into account particular anatomical variations and clinical needs related to the multiple decision-making criticalities to be considered in the context of breast cancer treatment.31

The DSC and 95HD are used as quantitative evaluation metrics to assess the proposed method. DSC values are used more commonly than 95HD values in the literature. Numerous studies have reported DSC values of CTV ranges from 0.88 to 0.93,16,32–34 The average DSC of our proposed model was 0.94, which was higher than those in historical reports. The result indicated a strong concordance between our automatic model and human experts for CTV contouring. The corresponding average DSC of U-Net was 0.93. For our model, we also used the 95HD to exclude the unreasonable distances caused by outliers, and the value was 4.31 mm for CTV. According to 95HD values, our model performed better than DDCNN (15.6 mm), DDNN (14.1 mm), and DD-ResNet (10.7 mm) in left breast CTV, as reported by Men et al.16 Moreover, we also evaluated the 95HD for OARs. Among all OARs, the significant results were achieved for spinal cord (95HD=4.37 mm, DSC=0.93). This is mainly because the spinal cord has good low-contrast visibility and its shape is regular. Both of the above metrics are statistically significant. Based on the results stated above, we can demonstrate that U-ResNet is superior to the U-Net architecture for this task. However, U-Net seems to be superior to U-ResNet in contouring right lung with a lower 95HD value. Considering the fact that the DSC for right lung with two architecture both reached 0.96, it remained to be discussed whether the significant difference resulted from the performance between two architectures or from the intrinsic nature of 95HD itself. In our following study, we will improve the delineation of right lung.

Except for performance metrics, we randomly distributed the AI and GT contours to experienced oncologists for further evaluation to verify their significance in clinical practice. The results showed that the majority of the AI contours can be accepted for subsequent treatment. There are no significant differences in scores grading between AI and GT contours.

Table 5 The Consistency Test Between Two Oncologists

<table>
<thead>
<tr>
<th>CTV Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>5</td>
<td>26</td>
<td>75</td>
<td>106</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>504</td>
<td>521</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>7</td>
<td>47</td>
<td>596</td>
<td>650</td>
</tr>
</tbody>
</table>

![Figure 4](https://doi.org/10.2147/CMAR.S330249)

Figure 4 The average CTV scores evaluated by two oncologists.
for CTV and all OARs, meaning that the U-ResNet model performed well, with good agreement for the manual contours. However, a Wilcoxon matched-pairs test indicated the significant difference between the evaluation of two oncologists. The inter-observer variability of the delineation standard is also one of the limitations of our study.

In addition, the consistency between the two radiation oncologists did not seem to be good. Among all the 650 slices of CTV marked with AI and GT contours, 532 slices (81.8%) were evaluated with the same CTV score, and the weighted kappa test (kappa = 0.282) showed the poor consistency between the oncologists. The results proved the considerable inter- and intra-observer variability, resulting in different judgment criteria. This may illustrate the fact that the acceptance of AI contouring methods is still impacted by the opinion and expertise of the treating radiation oncologists. We also evaluated the time needed for segmentation. With an approximately the same time cost, U-ResNet enjoyed an advantage of auto-contouring CTV and OARs spontaneously, saving time for clinical work. With U-ResNet assistance, contouring could be much easier and more efficient.

Several limitations of our study should be noted. First, the study is a single-center study, and the ground truth contours were approved by only two oncologists. Therefore, the proposed model cannot meet the contouring preferences of other centers and all clinicians. In the future, multi-center research should be conducted to obtain larger datasets to improve the generalization ability of the auto-segmentation model. Second, studies have suggested that 3D architecture showed better performance on segmentation compared with 2D.\(^{35}\) We may consider extending our model to 3D U-Net in the future. Third, CT images with artifacts caused by pacemakers as well as contrast injected CT images were not included in our training set. We were unable to deal with these conditions so far. Moreover, the CT slice thickness used in our study was 5 mm for model training. The auto-contouring will be less accurate if it is used for images with some other slice thickness.

**Conclusion**

Accurate and consistent segmentation is important for improving radiotherapy outcomes. Our study implemented a U-ResNet model to auto-delineate the CTV and OARs for breast conservative radiotherapy on planning CT images. The results showed that AI assistance can effectively improve consistency in contouring and streamlining radiotherapy workflows. In the future, with the assistance of AI, we may be able to obtain an initial standard segmentation to reduce the bias caused by observer preferences. However, for different patients and diseases, the contouring details should still be delineated specifically according to the principle of individualization.

**Abbreviations**

BC, breast cancer; BCS, breast conserving surgery; WBI, whole-breast irradiation; CTV, clinical target volume; OARs, organs at risk; ROIs, regions of interest; ABAS, atlas-based auto-segmentation; CNNs, convolutional neural networks; DIR, deformable image registration; GT, ground truth; AI, artificial Intelligence; DICOM, digital imaging and communications in medicine; ESTRO, European Society for Radiotherapy and Oncology; RTOG, Radiation Therapy Oncology Group.

**Funding**

The paper was funded by the National Foundation for Education Sciences Planning (grant number BLA200216).

**Disclosure**

Ms Qi Chen, Mr Shaobin Wang, and Mr Yu Chen are employees of MedMind Technology Co. The authors report no other conflicts of interest in this work.

**References**


