

Deep Learning-Based Auto-Segmentation for Liver Yttrium-90 Selective Internal Radiation Therapy

Jun Li, PhD¹ , Wookjin Choi, PhD¹, and Rani Anne, MD¹

Abstract

The aim was to evaluate a deep learning-based auto-segmentation method for liver delineation in Y-90 selective internal radiation therapy (SIRT). A deep learning (DL)-based liver segmentation model using the U-Net3D architecture was built. Auto-segmentation of the liver was tested in CT images of SIRT patients. DL auto-segmented liver contours were evaluated against physician manually-delineated contours. Dice similarity coefficient (DSC) and mean distance to agreement (MDA) were calculated. The DL-model-generated contours were compared with the contours generated using an Atlas-based method. Ratio of volume (RV, the ratio of DL-model auto-segmented liver volume to manually-delineated liver volume), and ratio of activity (RA, the ratio of Y-90 activity calculated using a DL-model auto-segmented liver volume to Y-90 activity calculated using a manually-delineated liver volume), were assessed. Compared with the contours generated with the Atlas method, the contours generated with the DL model had better agreement with the manually-delineated contours, which had larger DSCs (average: 0.94 ± 0.01 vs 0.83 ± 0.10) and smaller MDAs (average: 1.8 ± 0.4 mm vs 7.1 ± 5.1 mm). The average RV and average RA calculated using the DL-model-generated volumes are 0.99 ± 0.03 and 1.00 ± 0.00 , respectively. The DL segmentation model was able to identify and segment livers in the CT images and provide reliable results. It outperformed the Atlas method. The model can be applied for SIRT procedures.

Keywords

auto-segmentation, liver delineation, deep learning, atlas, resin yttrium-90

Received: 6 November 2024; revised: 10 February 2025; accepted: 17 February 2025

Introduction

Radioembolization using Yttrium-90 (Y-90) microspheres, a selective internal radiation therapy (SIRT), is a promising procedure to treat non-resectable primary and metastatic liver cancers.¹ Resin-based Y-90 procedures are used to treat metastatic liver cancers. In these procedures, the body-surface-area method is commonly used as the dosimetry method for activity calculation.² Liver volume size and tumor volume size are needed in the activity calculation, which are obtained from contours delineated by physicians in CT or MR images. It is desired that an auto-segmentation method can be applied for liver delineation in SIRT procedures to improve the efficiency of the activity calculation process.

In recent years, auto-segmentation methods for organ delineations have been studied extensively.³⁻¹⁶ For liver delineation, several methods have been investigated.¹⁰⁻¹⁶ However, there are very few studies on liver auto-segmentation for SIRT applications, which were based on Atlas-based auto-segmentation methods.^{15,16} Studies on Artificial Intelligence (AI) for biomedical image segmentation have shown that

AI-based segmentations are advanced auto-segmentation methods for organ delineations.¹⁷⁻²⁰ The study here was aimed to build a deep learning (DL)-based auto-segmentation model and explore the feasibility of applying the model for liver delineation in resin Y-90 SIRT.

Material and Method

Deep Learning Segmentation Model and Data

Liver segmentation has evolved significantly over the years, transitioning from traditional methods such as thresholding and region growing, which relied on handcrafted rules and

¹Department of Radiation Oncology, Thomas Jefferson University, Philadelphia, PA, USA

Corresponding Author:

Jun Li, PhD, Department of Radiation Oncology, Thomas Jefferson University, Philadelphia, PA 19107, USA.

Email: Jun.Li@jefferson.edu



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>).

were challenged by variations in image contrast and anatomical structures, to more advanced machine learning and Atlas-based techniques. While these methods improved segmentation accuracy, they required extensive feature engineering and struggled with inter-patient variability.²¹ The advent of deep learning revolutionized segmentation, particularly with the introduction of convolutional neural network (CNN)-based architectures. U-Net became the de facto standard due to its encoder-decoder structure and skip connections, enabling effective feature extraction across different spatial scales. Variants like V-Net extended these capabilities to 3D medical imaging, further enhancing volumetric segmentation. Recent advancements, including Attention U-Net and Residual U-Net, have incorporated attention mechanisms and residual connections to refine feature representation and improve segmentation precision.²²

We employed a Unet3D architecture¹⁸ for liver segmentation, leveraging its encoder-decoder structure with skip connections to reserve spatial details while capturing high-level contextual information. The network consists of four down sampling and up sampling levels. Each level incorporates convolutional layers and leaky ReLU activation. The encoding path extracts rich feature representations from the input image using strided convolutions for down sampling, and the decoding path generates the segmented output using transposed convolution for up sampling. The segmentation model generates the final segmentation using a softmax activation function to allow multi-class segmentation of liver region. To mitigate overfitting, a dropout rate of 0.1 was applied to deeper layers.

Machine learning library PyTorch and DL framework MONAI were used to build the liver segmentation model. The segmentation model was trained on the Liver Tumor Segmentation (LiTS) dataset,²³ which is a large and diverse dataset of contrast-enhanced abdominal CT scans containing a variety of liver shapes and sizes. The training set of LiTS contains 131 CT scans, and the test set contains 70 CT scans. This dataset comprises 3D CT scans with corresponding ground-truth contours. Preprocessing steps include resampling images to isovoxel spacing of $1.0 \times 1.0 \times 1.0$ mm, clipping CT intensities to the range of $[-200, 250]$ Hounsfield units (HU), and normalizing intensity values to $[0, 1]$. In order to minimize computational overhead and focus on the liver region, bounding box cropping was applied. Data augmentations, including random flipping, random cropping, random rotations along the three base axes, and random intensity shifting, were applied. A patch-based training strategy was adopted, where $128 \times 128 \times 128$ voxel patches were randomly extracted from the preprocessed images.

The network utilized residual units with two convolutional layers per unit to enhance feature learning. The AdamW optimizer was used with an initial learning rate of $1e-4$, adjusted dynamically using a reduce-on-plateau scheduler with a decay factor of 0.5. The loss function combined Dice loss and focal loss to ensure accurate segmentation of both large and small structures. The model was trained with a batch size of 2. Training was conducted on an NVIDIA A6000 graphics processing unit (GPU) with 48GB VRAM, and early stopping was applied based on validation performance.

For inference, a sliding window approach with a patch size of $128 \times 128 \times 128$ and 50% overlap was used to generate smooth segmentation maps for large volumetric images. Post-processing techniques, including softmax activation followed by argmax to obtain class labels and the application of morphological operations (hole filling and small object removal), were used to refine the final segmentations.

The trained segmentation model was deployed in the clinic using DICOM communication, as depicted in Figure 1. The segmentation workflow begins by exporting a CT scan in DICOM format from the clinical database. This scan is then automatically transferred to the AI segmentation pipeline, which can be hosted by the Computing Server either on-premises or in the cloud. Upon receiving the DICOM files, the automated segmentation pipeline is triggered. The pipeline coordinates a series of procedural stages. The process begins with pre-processing, where DICOM files are converted into a single 3D volume to optimize input for the subsequent AI segmentation model. The inference stage, powered by a GPU, applies the AI segmentation model to delineate the liver volume. Next, the post-processing stage transforms the results in a 3D mask volume into the DICOM RT Structure Set (DICOM-RTst) format. Upon completion, the pipeline exports the segmented results in the DICOM-RTst format to the initial DICOM storage location where the CT scans reside, or to a designated DICOM location. The automatically generated liver contours can be accessed by clinic users using any DICOM-compatible software.

The model was tested with the CT images of 18 SIRT patients who were treated for metastatic liver adenocarcinoma at our institution in recent years. The images were contrast-enhanced diagnostic CT images. No image processing was applied. Table 1 lists the patients' characteristics.

Evaluation Metrics

The auto-segmented liver contours were compared with the liver contours used in the SIRT procedures, which were manually delineated by radiation oncologists. The latter were taken as the ground truth. Dice similarity coefficient (DSC) and mean distance to agreement (MDA) were calculated.

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

where A and B represent auto-segmented contour and manually-delineated contour, respectively. The DSC quantifies the overlap between the two contours: 1 represents a perfect overlap and 0 represents no overlap. MDA represents the average distance between the two contours: the smaller the MDA, the better the contour agreement is.

The DL auto-segmentations were compared with Atlas-based auto-segmentations, which were performed with MiM Maestro (version 6.67) using a similar method as that in the literature.¹⁵ DSC and MDA were compared.

Further, the liver volumes obtained in the DL auto-segmentations were compared to the liver volumes obtained from the manual delineations, using the ratio of volume (RV).

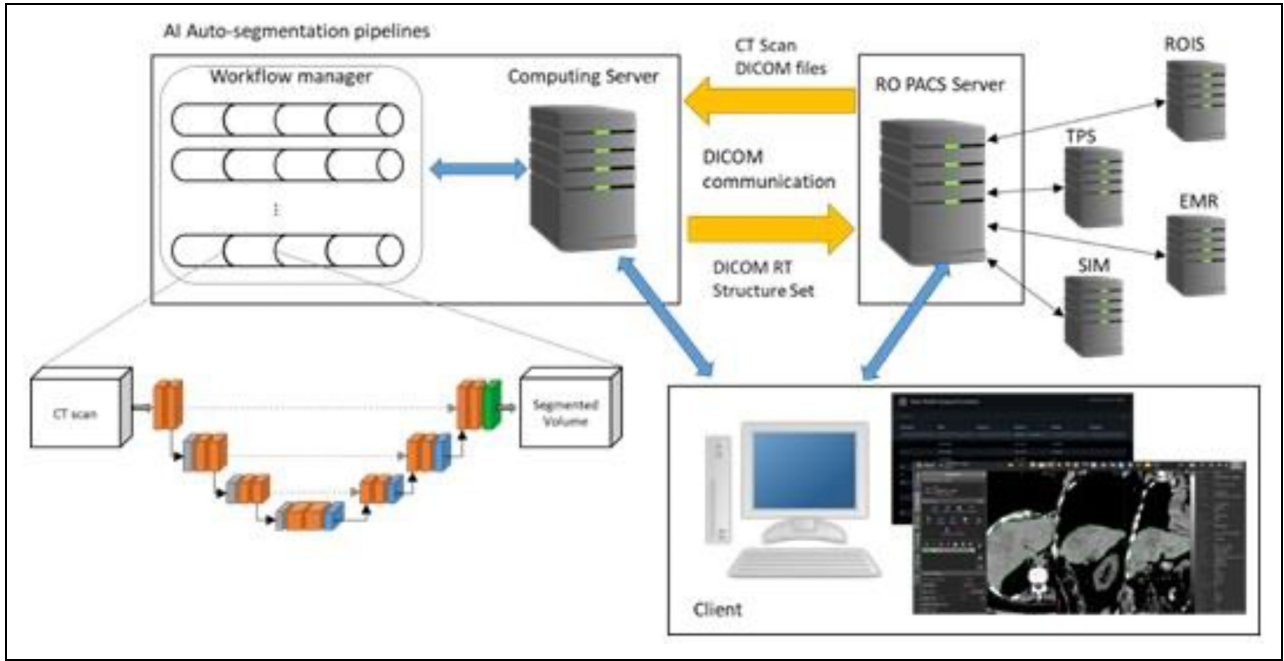


Figure 1. Schematic diagram of deep learning-based auto segmentation implementation for clinical use.

Table 1. Characteristics of Patients (N = 18).

	Range (average $\pm$ std)
Age	47-84 (63 $\pm$ 11)
Liver volume size (cm ³)	1120-4284 (2021 $\pm$ 780)

$$RV = \frac{|A|}{|B|} \quad (2)$$

Ratio of activity (RA), ie, ratio of the Y-90 activity TA_{auto} calculated with the body-surface-area method² using the DL auto-segmented liver volume to the Y-90 activity TA_{manual} calculated using the manually-delineated liver volume, was calculated and used to evaluate activity deviations from the accurate values. A value of 1 represents no deviations.

$$RA = \frac{TA_{auto}}{TA_{manual}} \quad (3)$$

Because the time that physicians spent on manual delineation in each SIRT procedure was not recorded, the contouring time comparison between manual delineation and auto-segmentation was not conducted in this retrospective study. The study focused on evaluating the segmentation accuracy and exploring the feasibility of applying auto-segmentation for SIRT.

Statistical Analysis

Wilcoxon signed rank test was applied to measure the difference in DSC and MDA between DL auto-segmented contours and Atlas auto-segmented contours. A *P*-value less than .05 was considered statistically significant.

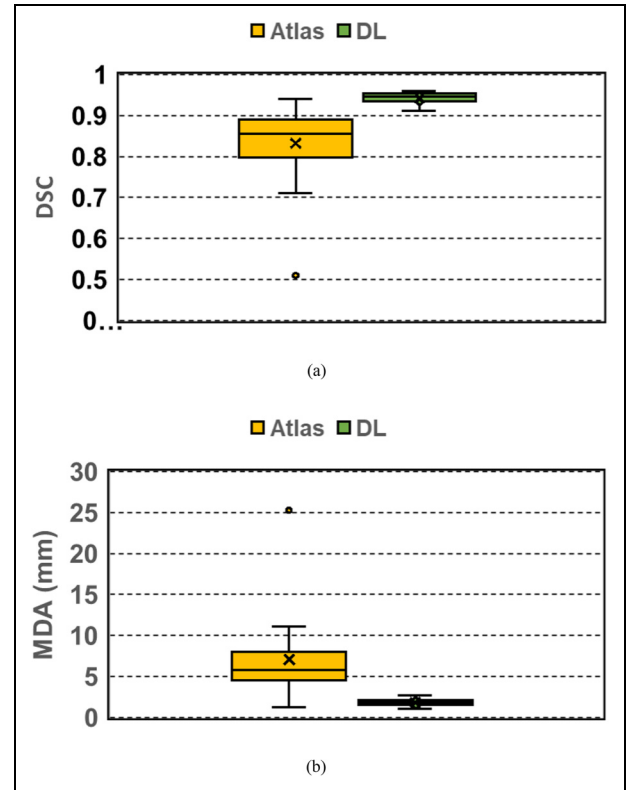


Figure 2. (a) Dice similarity coefficient and (b) mean distance agreement, calculated between auto-segmented liver contours and manually-delineated liver contours. The results of Atlas-based auto-segmentations (in orange) are provided as a comparison to the results of deep learning (DL)-based auto-segmentations (in green).

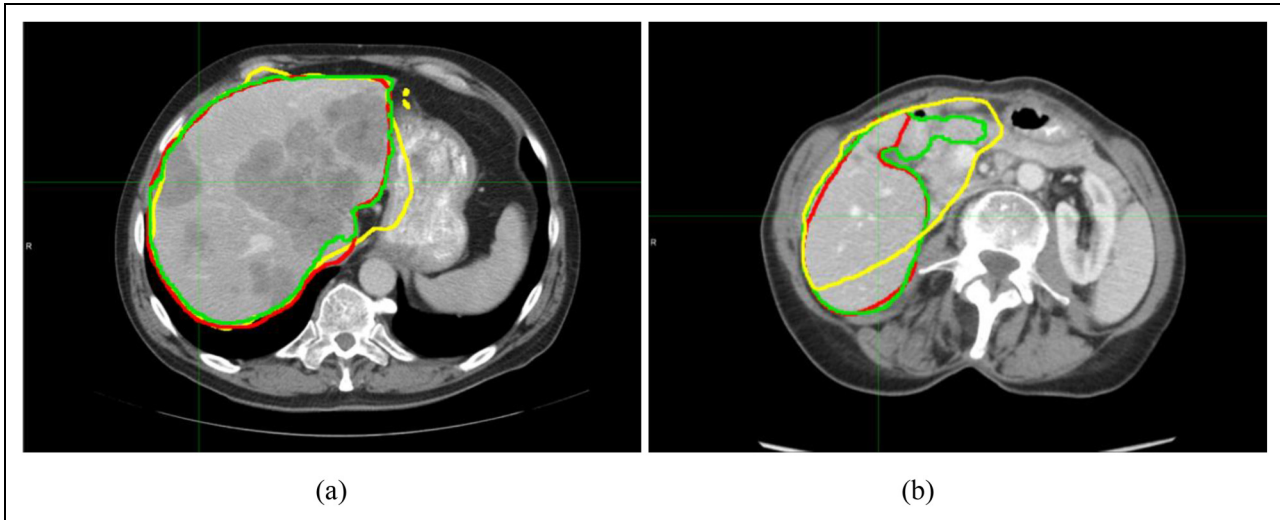


Figure 3. Auto-segmented liver contours generated in two cases: (a) $DSC_{Atlas} = 0.94$ and $DSC_{DL} = 0.95$; and (b) $DSC_{Atlas} = 0.79$ and $DSC_{DL} = 0.91$. The DL-based auto-segmented contour is in green, the Atlas-based auto-segmented contour is in yellow, and the manually-delineated contour (ground-truth) is in red.

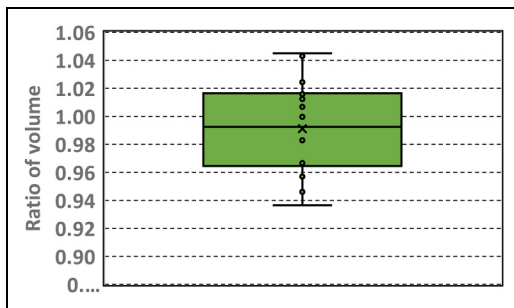


Figure 4. Ratio of DL auto-segmented liver volumes to manually-delineated liver volumes.

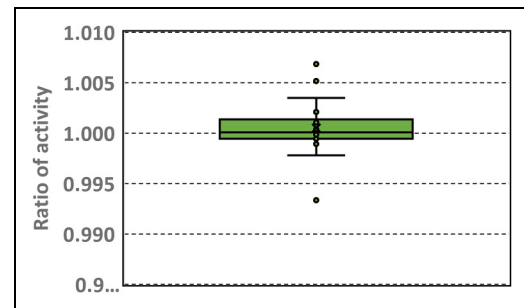


Figure 5. Ratio of Y-90 activities calculated using DL auto-segmented liver volumes to Y-90 activities calculated using manually-delineated liver volumes.

Results

The DSC and MDA are shown in Figure 2(a) and (b), respectively. The DSC of DL auto-segmented contours ranges from 0.91 to 0.96 (average: 0.94 ± 0.01), which indicates good agreement between the DL auto-segmented contours and the manually-delineated contours. The MDA ranges from 1.0 to 2.7 mm (average: 1.8 ± 0.4 mm). The DSC of Atlas auto-segmented contours ranges from 0.51 to 0.94 (average: 0.83 ± 0.10), and the MDA ranges from 1.2 to 25.3 mm (average: 7.1 ± 5.1 mm). The DL auto-segmented contours have a larger average DSC and smaller average MDA than the Atlas auto-segmented contours.

Figure 3 shows contour comparison in two cases. The DL auto-segmented contour is in green, the Atlas auto-segmented contour is in yellow, and the manually-delineated contour is in red. Figure 3(a) is the case where the Atlas auto-segmented contour has the highest DSC (0.94) among all the Atlas auto-segmented contours. The Atlas auto-segmented contour shows good agreement with the manually-delineated contour. In this case, the DL auto-segmented contour ($DSC: 0.95$) shows even better agreement with the manually-delineated contour. Figure 3(b) is the case where the DL auto-segmented contour

has the lowest DSC (0.91) among all the DL auto-segmented contours. The DL auto-segmented contour is still better than the Atlas auto-segmented contour, which has a DSC of 0.79.

Statistical analysis shows that the differences in both DSC and MDA are significant between the DL auto-segmented contours and Atlas auto-segmented contours ($P < .01$). The DL auto-segmented contours have better agreement with the manually-delineated contours (the ground truth) than the Atlas auto-segmented contours.

Figure 4 and Figure 5 show RV and RA, respectively, which were calculated using the DL auto-segmented liver volumes. The RV ranges from 0.94 to 1.05 (average: 0.99 ± 0.03), and RA ranges from 0.99 to 1.01 (average: 1.00 ± 0.00). Table 2 lists the result summary.

Discussion

To the best of our knowledge, published studies on auto-segmentation in SIRT were based on Atlas-based segmentation methods.^{15,16} There were no publications on applying

Table 2. DSC, MDA, RV, and RA of DL Auto-Segmented Contours (DSC and MDA of Atlas Auto-Segmented Contours are in Brackets) (N=18).

	DSC	MDA (mm)	RV	RA
Min	0.91 (0.51)	1.0 (1.2)	0.94	0.99
Max	0.96 (0.94)	2.7 (25.30)	1.05	1.01
Median	0.95 (0.86)	1.8 (5.8)	0.99	1.00
Mean $\pm$ Stdev	0.94 $\pm$ 0.01 (0.83 $\pm$ 0.10)	1.8 $\pm$ 0.4 (7.1 $\pm$ 5.1)	0.99 $\pm$ 0.03	1.00 $\pm$ 0.00
P value (DL vs Atlas)	<.01	<.01		

DL-based auto-segmentation in SIRT. In this study, we built a DL model and explored its application to SIRT. The results show that the liver contours generated with the DL model have better agreement with the manual delineations than those generated with the Atlas-based method. The study demonstrates the application of DL based auto-segmentation in SIRT and shows that DL-based auto-segmentations are superior to Atlas-based auto-segmentations in the SIRT applications.

A recent publication that evaluated five commercial AI software for organ delineation in radiotherapy, reported DSCs ≥ 0.96 for liver delineation in breast cancer patients and lung cancer patients,²⁴ which are slightly higher than the DSCs (≥ 0.91) in our study. It is noted that the test data are different between that study and our study. The test data in our study were of liver cancer patients, and the images were contrast-enhanced images. The heterogeneity of the liver (due to tumors) and the fact that adjacent tissues might have similar image intensities as the liver or higher image intensities (due to contrast agent) (see Figure 3(b)), brought challenges to the auto-segmentation of the liver in our study.

The DSCs of liver contours generated with the DL model in this study have similar magnitudes to those achieved with the best segmentation algorithm in the competition studies using test data of liver cancer patients.²¹ Although the test data and the segmentation models are different between our study and that study, the DSCs indicate that our DL auto-segmentation model performs well. The results also indicate that the LiTS dataset can be used as training data for DL models used in SIRT.

RA results show that Y-90 activities calculated using the DL auto-segmented liver volumes are close to the accurate activities calculated using the manually-delineated liver volumes: the differences are within 1%. The results indicate that the DL model can be applied for SIRT procedures. With physicians' review or slight edits, the contours can be used for activity calculations.

The DL auto-segmentation approach implemented in the study can process large amounts of data efficiently. The DICOM communication makes the approach easy to deploy in a clinical environment. In the study, CT images of all 18 cases were sent to the server at the same time to generate contours. The entire process of all the cases, from sending CT images to receiving contours in the clinical database, took about 30 min. For a single case, the process takes about 1-2 min. The quick process is important for clinical applications where a large volume of CT scans can be segmented. Except for the step of selecting image data to send, other steps of the process (ie, generating contours and returning contours back to the clinical database) are fully automatic.

The number of test cases in the study was limited by the number of SIRT patients that were treated at our institution. We expect to test more cases in the future.

In the study, lower DSCs occurred in the cases where the tissues adjacent to the liver had similar image contrast as the liver or higher image contrast. The current model had difficulty in dealing with such situations. We anticipate further studies to improve the model to overcome these challenges.

Conclusions

A DL-based segmentation model was built, which was able to successfully identify and segment livers in the CT images of SIRT patients and provide reliable results. It outperformed the Atlas method. The model can be easily deployed in a clinical environment using DICOM communication, which can process large amounts of data efficiently. The application of the model will improve the efficiency of liver segmentation in SIRT clinical practice.

Abbreviations

SIRT	selective internal radiation therapy
DL	deep learning
DSC	Dice similarity coefficient
MDA	mean distance to agreement
RV	Ratio of volume
RA	ratio of activity
AI	artificial intelligence
LiTS	Liver Tumor Segmentation

ORCID iD

Jun Li  <https://orcid.org/0000-0003-1430-8148>

Statements and Declarations

Ethical Considerations

IRB office determined that the study did not require IRB approval because the retrospective study used anonymized data.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Publication made possible in part by support from the Thomas Jefferson University Open Access Fund.

Conflicting Interests

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

References

- Kennedy A, Nag S, Salem R, et al. Recommendations for radioembolization of hepatic malignancies using Yttrium-90 microsphere brachytherapy: A consensus panel report from the radioembolization brachytherapy oncology consortium. *Int J Radiat Oncol Biol Phys*. 2007;68(1):13–23.
- Bernardini M, Smadja C, Faraggi M, et al. Liver selective internal radiation therapy with (90)Y resin microspheres: Comparison between pre-treatment activity calculation methods. *Phys Med*. 2014 Nov;30(7):752–764. doi: 10.1016/j.ejmp.2014.05.004. Epub 2014 Jun 9. PMID: 24923844.
- Huyskens DP, Maingon P, Vanuytsel L, et al. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. *Radiation Oncol*. 2009 Mar;90(3):337–345. doi: 10.1016/j.radonc.2008.08.007. Epub 2008 Sep 21. PMID: 18812252.
- Delpont G, Escande A, Ruef T, et al. Comparison of automated Atlas-based segmentation software for postoperative prostate cancer radiotherapy. *Front Oncol*. 2016 Aug 3;6:178. doi: 10.3389/fonc.2016.00178. PMID: 27536556; PMCID: PMC4971890.
- Lee H, Lee E, Kim N, et al. Clinical evaluation of commercial Atlas-based auto-segmentation in the head and neck region. *Front Oncol*. 2019 Apr 9;9:239. doi: 10.3389/fonc.2019.00239. PMID: 31024843; PMCID: PMC6465886.
- Casati M, Piffer S, Calusi S, et al. Methodological approach to create an atlas using a commercial auto-contouring software. *J Appl Clin Med Phys*. 2020 Dec;21(12):219–230. doi: 10.1002/acm2.13093. Epub 2020 Nov 25. PMID: 33236827; PMCID: PMC7769405.
- Meillan N, Bibault JE, Vautier J, et al. Automatic intracranial segmentation: Is the clinician still needed? *Technol Cancer Res Treat*. 2018 Jan 1;17:1533034617748839. doi: 10.1177/1533034617748839. PMID: 29343204; PMCID: PMC5784565.
- Kim N, Chang JS, Kim YB, Kim JS. Atlas-based auto-segmentation for postoperative radiotherapy planning in endometrial and cervical cancers. *Radiat Oncol*. 2020 May 13;15(1):106. doi: 10.1186/s13014-020-01562-y. PMID: 32404123; PMCID: PMC7218589.
- Nemoto T, Futakami N, Yagi M, et al. Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi. *J Radiat Res*. 2020 Mar 23;61(2):257–264. doi: 10.1093/jrr/rrz086. PMID: 32043528; PMCID: PMC7246058.
- La Macchia M, Fellin F, Amichetti M, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiat Oncol*. 2012 Sep 18;7:160. doi: 10.1186/1748-717X-7-160. PMID: 22989046; PMCID: PMC3493511.
- Heimann T, van Ginneken B, Styner M, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging*. 2009;28(8):1251–1265.
- Lu X, Xie Q, Zha Y, Wang D. Fully automatic liver segmentation combining multi-dimensional graph cut with shape information in 3D CT images. *Sci Rep*. 2018 Jul 16;8(1):10700. doi: 10.1038/s41598-018-28787-y. PMID: 30013150; PMCID: PMC6048104.
- Yan Z, Zhang S, Tan C, et al. Atlas-based liver segmentation and hepatic fat-fraction assessment for clinical trials. *Comput Med Imaging Graph*. 2015 Apr;41:80–92. doi: 10.1016/j.compmedimag.2014.05.012. Epub 2014 Jun 9. PMID: 24962337.
- Bousabarah K, Letzen B, Tefera J, et al. Automated detection and delineation of hepatocellular carcinoma on multiphasic contrast-enhanced MRI using deep learning. *Abdom Radiol (NY)*. 2021 Jan;46(1):216–225. doi: 10.1007/s00261-020-02604-5. Epub 2020 Jun 4. PMID: 32500237; PMCID: PMC7714704.
- Li J, Anne R. Comparison of Eclipse Smart Segmentation and MIM Atlas Segment for liver delineation for Yttrium-90 selective internal radiation therapy. *J Appl Clin Med Phys*. 2022 Jun 15: e13668. Doi: 10.1002/acm2.13668
- Li J, Anne R. Evaluation of Atlas-based auto-segmentation of liver in MR images for Yttrium-90 selective internal radiation therapy. *J Appl Clin Med Phys*. 2023 May;24(5):e13979. doi: 10.1002/acm2.13979. Epub 2023 Apr 17. PMID: 37070130; PMCID: PMC10161143.
- Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2020:1–9.
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. 2015; MICCAI, vol. 9351:234–241.
- Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision. 3DV; 2016; IEEE:565–571.
- Samarasinghe G, Jameson M, Vinod S, et al. Deep learning for segmentation in radiation therapy planning: A review. *J Med Imaging Radiat Oncol*. 2021 Aug;65(5):578–595. doi: 10.1111/1754-9485.13286. Epub 2021 Jul 26. PMID: 34313006.
- Gul S, Khan MS, Bibi A, Khandakar A, Ayari MA, Chowdhury MEH. Deep learning techniques for liver and liver tumor segmentation: A review. *Comput Biol Med*. 2022;146:105620. <https://doi.org/10.1016/j.compbiomed.2022.105620>
- Kerfoot E, Clough J, Oksuz I, Lee J, King AP, Schnabel JA. Left-Ventricle quantification using residual U-net. In: *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges (STACOM 2018)*. Springer; 2019:371–380. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-030-12029-0_40
- Bilic P, Christ P, Li HB, et al The liver tumor segmentation benchmark (LiTS). *Med Image Anal*. 2023 Feb;84:102680. doi: 10.1016/j.media.2022.102680. Epub 2022 Nov 17. PMID: 36481607.
- Doolan PJ, Charalambous S, Roussakis Y, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. *Front Oncol*. 2023 Aug 4;13: 1213068. doi: 10.3389/fonc.2023.1213068. PMID: 37601695; PMCID: PMC10436522.