scientific reports

Check for updates

OPEN

Deep learning for automatic segmentation of hepatocellular carcinoma in contrast enhanced CT scans

Ilaria Manghi[2]✉, Annarita Pecchi[2], Giuseppe Esposito[4], Maria Macheda[2], Elisa Zanni[1], Fabrizio Di Benedetto[3] & Federica Ferraguti[1]

Liver cancer represents a significant cause of cancer-related mortality, with hepatocellular carcinoma (HCC) being the most prevalent forms. Computed tomography (CT) serves as the principal imaging modality for the diagnosis of liver tumors, particularly HCC. The precise identification of tumor presence and location necessitates highly skilled radiologists. Consequently, automated liver tumor segmentation from CT images offers a valuable tool to support cancer diagnosis and treatment planning. Nevertheless, this task presents considerable challenges due to the inherent variability in tumor shape, dimension, and imaging acquisition techniques. In this paper, we evaluate state-of-the-art segmentation architectures across a range of datasets. These include the publicly accessible Liver Tumor Segmentation (LiTS) dataset, which covers a spectrum of liver lesions, as well as the HCC-TACE-Seg and WAW-TACE datasets, comprising CT scans of HCC patients prior to treatment. In addition, we introduce a novel dataset of contrast-enhanced CT (CECT) scans that are routinely used for HCC diagnosis. The focus of this manuscript is on the segmentation of both the liver and tumors, with specific attention directed toward HCC. This study provides a comparative analysis of segmentation models applied to diverse datasets, encompassing both public and proprietary datasets.

Keywords Liver and tumor segmentation, CT scans, Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) accounts for 75%-85% of all primary liver tumors and ranks as the third leading cause of cancer-related deaths worldwide. Chronic liver disease and inflammation are the predominant risk factors for HCC development, with 80%-90% of cases occurring in the context of cirrhosis[1]. Continuous clinical and radiological surveillance in specialized centers is crucial for early detection and timely treatment. In specific clinical scenarios, such as high-risk patients with cirrhosis, chronic hepatitis B virus, and a current or prior history of HCC, a non-invasive diagnosis of HCC can be made without the need for histopathological confirmation[2]. According to the Liver Imaging Reporting and Data System (LI-RADS) – a standardized system used by radiologists to interpret and report liver imaging findings, particularly in patients at high risk for HCC – contrast-enhanced CT (CECT) or MRI are the recommended imaging modalities for HCC diagnosis, provided appropriate techniques are employed and the studies are evaluated by an expert radiologist[3]. Several international guidelines exist for HCC diagnosis, based on a combination of various imaging criteria; however, most guidelines agree that a lesion size of at least 10 mm, non-rim arterial phase hyperenhancement, and non-peripheral washout are required to diagnose HCC. Hepatic lesions with atypical imaging characteristics may require additional imaging exams, follow-up over time, or histopathological analysis via lesion biopsy[2]. In this context, the application of artificial intelligence (AI) holds the potential to optimize the automatic detection of these typical imaging criteria, enhance the accuracy of non-invasive diagnosis, and facilitate the early detection of HCC, thereby reducing the need for liver biopsies and improving survival outcomes. From a clinical perspective, the integration of AI tools for targeted training of healthcare providers, particularly those in peripheral hospitals, will enhance the overall quality of care for HCC patients. The global shortage of HCC specialists poses a significant challenge, but distributing diagnostic tools can expedite referrals, ensuring patients are transferred

[1]Department of Science and Methods for Engineering, University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy. [2]Department of Medical and Surgical Sciences of Children and Adults, University of Modena and Reggio Emilia, 41224 Modena, Italy. [3]Hepato-Pancreato-Biliary Surgery and Liver Transplantation Unit, University of Modena and Reggio Emilia, 41125 Modena, Italy. [4]Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, 41125 Modena, Italy. ✉email: ilaria.manghi@unimore.it

to specialized centers in a timely manner. The findings of this study are expected to contribute significantly to reducing diagnostic errors and delays, leading to earlier treatment and improved survival outcomes for patients with HCC, while also reducing healthcare costs.

Despite this potential, data collection and labeling remain major challenges in liver and tumor segmentation, as well as medical imaging segmentation in general[4]. Data collection can raise patient privacy concerns and yield heterogeneous data due to variations in acquisition technology and protocols. Data labeling is a time-consuming process requiring expert medical knowledge[5].

Consequently, only a limited number of datasets for liver and tumor segmentation are publicly available. The 3D-IRCADb-01[6] contains 20 anonymized abdominal CT scans, but its small size is not well-suited for training deep learning networks. The Liver Tumor Segmentation (LiTS) dataset[7] and the LLD-MMRI-Dataset[8] include scans with different liver lesions; however, they do not specifically focus on HCC. Moreover, the LiTS dataset does not provide contrast-enhanced images, while the LLD-MMRI-Dataset is MRI-based rather than CT. For HCC specifically, the HCC-TACE-Seg[9] and the WAW-TACE[10] datasets contain CT images of HCC patients, while the LiverHccSeg[11] and ATLAS[12] datasets offer contrast-enhanced images, although both are MRI-based rather than CT. Verification of LI-RADS criteria[3] necessitates optimal multi-phasic imaging, involving CT or MRI acquisitions in the codified arterial, portal-venous, and delayed phases. Achieving the highest diagnostic accuracy requires a combined evaluation of all post-contrast acquisitions. However, none of the aforementioned datasets provide all three codified phases with corresponding HCC labeling. Therefore, a dataset containing all phases, labeled and verified by expert radiologists, is essential for precise classification of HCC tumors following lesion segmentation, as HCC characteristics evolve across these phases.

Several deep learning architectures exist for liver tumor segmentation. NnU-Net[13] is an automatic self-configuring framework for medical image segmentation adaptable to various datasets and tasks. Hameedur Rahman et al.[14] segmented the liver and tumors in CT images using ResUNet, a U-Net variant incorporating residual connections. Graph convolutional networks have also been employed to capture spatial information by modeling key regions as graphs, as seen in[15]. In the same year, Jakob Wasserthal et al.[16] presented TotalSegmentator, a nnU-Net based model demonstrating strong performance across diverse medical segmentation tasks. Furthermore, attention mechanisms have been introduced to enhance liver tumor segmentation, as shown by Kasun Hettihewa et al.[17]. In[18], the authors present a 3D segmentation model that combines U-Net and transformer architectures. Although not specifically designed for liver tumor segmentation, it proved to be one of the top-performing architectures for 3D medical image segmentation on public test datasets[19]. Similarly, Jun Ma et al.[20] introduce the U-Mamba hybrid architecture, which outperforms state-of-the-art models in biomedical image segmentation.

To provide the best solution for liver tumor segmentation, in this paper we selected nnU-Net, SwinUNETR, and U-Mamba, commonly recognized as the most suitable for this task, and analyzed and compared their performance on the aforementioned datasets.

The main contributions of this paper are:

- Creation of a new dataset, namely the HCC-ARSeg dataset (Section 3.1), consisting of contrast-enhanced CT scans including arterial, portal-venous, and delayed phases, with each phase meticulously labeled by an expert radiologist. We evaluated this dataset using the aforementioned networks and benchmarked its performance against existing datasets. This labeled multi-phasic dataset can be effectively utilized for HCC classification following lesion segmentation.
- Evaluation and comparison of the performance of different architectures trained and tested on both publicly available datasets and on the HCC-ARSeg dataset.

Methods

This section presents the segmentation models selected for comparison. We selected three well-established and widely referenced approaches as benchmarks[21], considered among the leading architectures for this task, to provide a valuable comparison. These models, each based on a different underlying architecture, are nnU-Net, a self-adapting convolutional network; Swin UNETR, a transformer-based segmentation model; and U-Mamba, which employs an innovative architecture combining convolutional neural networks (CNNs) with state space sequence models (SSMs). All of the previously mentioned architectures were implemented in 3D, as the study is based on CT scans.

nnU-net

The first selected architecture is an instance of nnU-Net ("no-new-Net")[22], a robust, self-configuring deep learning framework tailored for medical image segmentation. Recognized as state-of-the-art in this domain, nnU-Net has consistently demonstrated superior performance across various biomedical challenges, including brain tumor segmentation[13], lung lesion segmentation[23], and liver parenchyma segmentation[24]. A key feature of nnU-Net is its inherent ability to manage all dataset conversions and adaptations required for individual datasets. It can process both 2D and 3D data, negating the need for task-specific optimizations while accommodating the inclusion of custom rules. The framework is built upon 2D and 3D vanilla U-Nets and follows an encoder-decoder architecture with skip connections linking corresponding levels of the two pathways. For processing 3D CT scans, in this paper the network was trained using a 3D full-resolution configuration with five-fold cross-validation, as implemented within nnU-Net. Furthermore, the Residual Encoder[21] available in the nnU-Net framework was employed to enhance segmentation performance. Training utilized a composite loss function of Dice loss and binary cross-entropy.

SwinUNETR

SwinUNETR[18] is an encoder-decoder architecture where the encoder combines a U-shaped network with a Swin Transformer, and the decoder employs convolutional neural networks (CNNs). Operating at multiple resolutions, the encoder and decoder are interconnected through skip connections. Initially developed for brain tumor semantic segmentation, SwinUNETR has been successfully adapted for diverse tasks, including multi-organ segmentation[19]. This architecture, specifically designed for 3D volumetric data, is well-suited for application to 3D CT scans. In this paper, SwinUNETR was implemented using MONAI[1] and the dataset was partitioned into training, validation, and testing sets. Training utilized Dice cross-entropy loss, with the encoder initialized using self-supervised pretraining weights from the BTCV challenge[19].

U-Mamba

U-Mamba[20] is presented as a general-purpose encoder-decoder network for biomedical image segmentation. It employs an innovative approach by introducing a hybrid CNN-SSM block, enabling the capture of both local features through convolutional layers and long-range dependencies via SSMs. The network's convolutional component is based on the U-Net architecture. Notably, U-Mamba also features a self-configuring mechanism, allowing it to automatically adapt to diverse datasets without manual intervention, similar to nnU-Net.

Two U-Mamba variants exist: one implementing the U-Mamba blocks solely in the bottleneck, and another integrating them throughout the entire encoder. In this study, the network was trained using the U-Mamba bottleneck configuration and a five-fold cross-validation strategy. Training employed a composite loss function comprising Dice loss and binary cross-entropy.

Results

This section evaluates the performance of the aforementioned architectures on various liver and tumor segmentation datasets. Specifically, three publicly available datasets were analyzed and expanded upon with a novel dataset to address the multi-phasic requirement detailed in Section 1. The evaluation across these datasets focuses solely on the portal venous phase, as it is widely recognized as the most suitable for clear differentiation between liver and tumor[25] and is consistently available in all datasets. However, the availability of all three phases (arterial, portal-venous, and delayed) would enable precise identification and classification of HCC following segmentation.

Investigated datasets

A concise overview of the liver and tumor segmentation datasets used in this study is presented here. The description starts with the general LiTS dataset and then proceeds to more specialized datasets focusing on HCC tumors. Figure 1 illustrates an example of an original image (one slice from the CT scan) with the corresponding overlaid segmentation mask for each dataset.

LiTS

The LiTS benchmark dataset[7] comprises abdominal CT scans from seven distinct clinical sites. Not all scans in this dataset contain liver lesions. The dataset features a range of primary and secondary liver tumors exhibiting varying lesion-to-background ratios. Acquisition protocols and imaging equipment differ across the sites. The number of tumors per scan varies from 0 to 12. A radiologist with over three years of experience manually annotated all images, and three additional radiologists validated these annotations.

HCC-TACE-Seg

The HCC-TACE-Seg collection[9] comprises CT scans from 105 patients with confirmed HCC, all treated at the University of Texas MD Anderson Cancer Center. This dataset includes multi-phasic contrast-enhanced CT images acquired both before and within 14 weeks following transarterial chemoembolization (TACE). Notably, all scans were obtained using the same scanner and acquisition protocols. For each patient, arterial and delayed phases were acquired and subsequently resampled to the portal-venous phase. The dataset provides segmentations for both tumors and the liver, annotated by three radiology residents and reviewed by an expert radiologist. The study specifically included only patients with a single tumor.

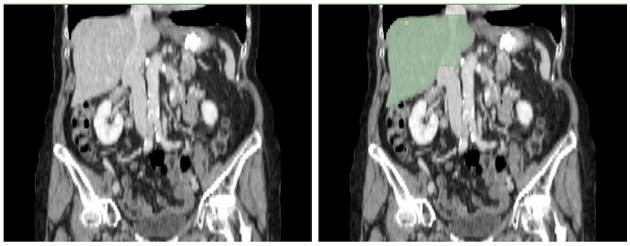
WAW-TACE

The WAW-TACE dataset[10] comprises multi-phase abdominal CT images from 233 treatment-naive patients with HCC, all from the Medical University of Warsaw. These patients underwent multi-phasic CT imaging within 90 days prior to their initial TACE cycle. The dataset includes segmentation masks for HCC lesions and various anatomical structures. While organ segmentations were automatically generated using the TotalSegmentator deep learning model[26], only the liver segmentations were retained to be considered in the evaluation. Tumor annotations were performed manually by radiologists on specific contrast phases, not across all phases.

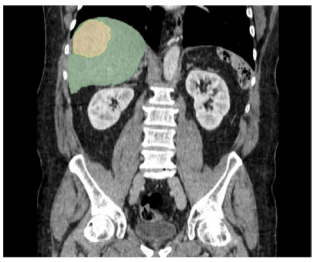
HCC-ARSeg

The creation of a dedicated dataset stems from the absence in existing literature of a well-defined resource specifically focused on HCC that utilizes contrast agents, features high-precision acquisition, and provides accurately validated segmentations for the tumor and liver. Furthermore, a dataset encompassing both the images and annotations across the three key contrast phases is currently lacking in the literature, thus hindering the exploitation of multi-phasic algorithms for segmentation and classification. The dataset comprises multi-

[1] Medical Open Network for Artificial Intelligence: https://monai.io/,

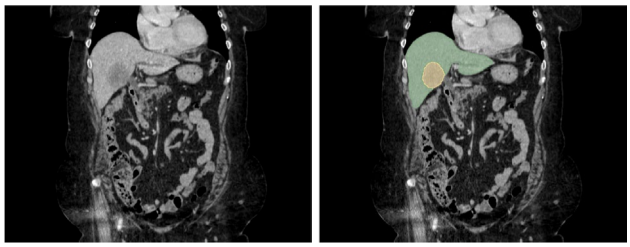


(a) LiTS dataset.



(b) WAW-TACE dataset.

(c) HCC-TACE dataset.



(d) HCC-ARSeg dataset.

Fig. 1. Datasets used for evaluation. Left image: original image; right image: overlaid segmentation mask.

phase abdominal CT scans. All the scans were acquired before and after the injection of iodinated contrast; however, for the purposes of this work, scans without contrast were not included. The scans were collected in a cohort of patients diagnosed with HCC who underwent hepatic resection and/or liver transplantation between January 1, 2014, and December 31, 2023. All procedures were performed at the Hepato-Pancreato-Biliary Surgery and Liver Transplantation Unit, Policlino Modena Hospital, Italy. Contrast-enhanced CT (CECT) scans were acquired according to established literature guidelines, including post-contrastographic phases at the following time points after intravenous injection of iodinated contrast: late hepatic arterial phase (approximately 35–45 s), portal venous phase (60–75 s), and delayed phase (3-5 min).

The annotation process began with an initial automatic segmentation using pre-trained weights from the publicly available Medical Segmentation Decathlon (Medical Segmentation Decathlon: http://medicaldecathlon.com/), thus reducing the burden of manual annotation. Subsequently, an expert radiologist specializing in liver diseases at AOU Modena performed a radiological evaluation based on the LI-RADS v2018 criteria[27]. Ultimately, the final histopathological report from an expert liver pathologist at AOU Modena served as the basis. A radiology resident then meticulously performed manual annotation of both the liver and the tumor in all three phases (arterial, portal-venous, and delayed), using the automatic pre-annotations as a starting point. Finally, the expert radiologist reviewed and validated these annotations. The study cohort includes 94 patients aged 18 years or older with a confirmed radiological and histopathological diagnosis of HCC.

All datasets are annotated with three classes: background, liver, and tumor. Table 1 shows how the pixels are distributed among the different classes in the examined datasets. The values indicate the proportion of pixels belonging to each class, reflecting the class imbalance commonly encountered in liver tumor segmentation tasks.

Datasets	Background (%)	Liver (%)	Tumor (%)
LiTS	97.78	2.10	0.12
HCC-TACE-Seg	94.60	4.18	1.22
WAW-TACE	97.27	2.57	0.16
HCC-ARSeg	98.00	1.94	0.06

Table 1. Percentage distribution of pixel classes—background, liver, and tumor—in each dataset.

	CT scans			Slices		
Datasets	Train	Valid	Test	Train	Valid	Test
LiTS	103	14	14	45227	7258	6126
HCC-TACE-Seg	85	10	10	7666	813	860
WAW-TACE	59	7	7	12542	1506	2258
HCC-ARSeg	76	9	9	24919	2875	3095

Table 2. Distribution of scans across training, validation, and test sets.

Moreover, HCC-ARSeg dataset contains the lowest percentage of tumor pixels among the datasets. The values are calculated based on expert annotations, as the ratio between the number of pixels belonging to a specific class and the total number of pixels in the image.

Evaluation on different datasets

The architectures selected and described in Section 2 were trained and tested on each dataset. To ensure a fair comparison across models, each dataset was consistently partitioned into training and test sets for all training runs. For nnU-Net and U-Mamba, the validation set was automatically determined during pre-processing and varied per fold. In contrast, a specific validation set was created for training Swin-UNETR.

The sets have been created as follows:

- The LiTS dataset, containing 201 CT scans, had annotations for only 131. Therefore, we utilized only these annotated scans, allocating 103 to the training set and evenly distributing the remaining 28 into validation and test sets (14 each).
- For the HCC-TACE dataset, comprising 105 images, we focused solely on the portal-venous phase. The training set consisted of 85 scans, with the remaining 20 scans equally divided between the validation and test sets.
- The WAW-TACE dataset was limited to the 74 CT scans with available tumor annotations in the portal-venous phase. Of these, 59 scans were assigned to the training set, with the remaining scans equally split into validation and test sets.
- The HCC-ARSeg dataset included 94 scans from distinct patients, each with three imaging phases (arterial, portal-venous, and delayed). To maintain comparability with the other datasets, we used only the portal-venous phase, assigning 76 patients to the training set and evenly distributing the remaining 18 between the validation and test sets.

Table 2 details the partitioning of the three datasets into training, validation, and test sets. The first three columns specify the number of scans in each split, while the final three columns present the corresponding number of slices.

Table 3 presents the performance of the three examined architectures. Each architecture was trained on a single dataset and evaluated on a comprehensive test set comprising all test samples from the four datasets. The first column indicates the training dataset. Performance is reported using the Dice Score and Symmetric Average Surface Distance (SASD), with each entry in the table providing these two metrics as the mean value between the liver and tumor classes. The last row in Table 3 shows the performance on a modified iteration of the HCC-ARSeg dataset. For this version, we utilized TotalSegmentator[16] for liver segmentation. The resulting liver masks were then used to define a region of interest (ROI), and the CT scans were cropped accordingly. This cropping step aimed to reduce the input size for the tumor segmentation network, concentrating on the anatomically relevant liver region. We employed the nnU-Net framework for only tumor segmentation on the cropped CT scans, given its superior performance. This procedure resulted in two separate segmentation masks, one for the liver and one for the tumor. Ultimately, the segmented liver and tumor segmentation masks were merged into a single multi-class mask for performance metric calculation.

Discussion

Table 3 shows the evaluation results for nnU-Net, SwinUNetr, and U-Mamba architectures on various training datasets. The evaluation is based on two key metrics: the Dice similarity score, which measures segmentation overlap, and the Symmetric Average Surface Distance (SASD), indicating the average distance between the predicted and ground truth surfaces. Table 3 demonstrates that the nnU-Net architecture outperformed U-Mamba and SwinUNetr in terms of Dice score and SASD in all training scenarios evaluated. U-Mamba shows

Training datasets	nnU-Net	SwinUNetr	U-Mamba
	Dice	Dice	Dice
LiTS	0.688	0.476	0.652
HCC-TACE-Seg	0.761	0.427	0.757
WAW-TACE	0.782	0.423	0.770
HCC-ARSeg	0.752	0.384	0.734
HCC-ARSeg (mod)	0.712	-	-
Training datasets	**nnU-Net**	**SwinUNetr**	**U-Mamba**
	SASD	**SASD**	**SASD**
LiTS	5.401	65.052	9.420
HCC-TACE-Seg	6.836	81.982	9.452
WAW-TACE	2.720	90.126	7.097
HCC-ARSeg	4.680	94.916	13.601
HCC-ARSeg (mod)	3.903	-	-

Table 3. Performance of the investigated architectures evaluated on the presented datasets.

comparable Dice performance to nnU-Net, while its SASD values are slightly lower which suggests that it detects the correct tumor regions with less precise contours. SwinUNETR yields significantly lower performance with respect to the other two architectures. The lowest performance of SwinUNetr in our experiments can be mainly attributed to extensive training data required by transformer-based architectures. The datasets used in this study are relatively limited in size and variability. This is proven by the fact that LiTS dataset achieves the best performance with SwinUNETR. This dataset is indeed the largest in size and the more varied since it includes different kinds of liver tumors.

Although overall performance remains consistent across different training datasets, the following major observations can be made. The WAW-TACE dataset produced the best results for both nnU-Net and U-Mamba. It achieved the highest Dice score and the lowest SASD. In contrast, training with the LiTS dataset led to the lowest performance for both nnU-Net and U-Mamba, in terms of both Dice score and SASD. This suggests that LiTS, as a training dataset, does not allow these models to generalize effectively. On the other side, as previously mentioned, it is the dataset that produces the best results for SwinUNETR. The HCC-TACE-Seg and HCC-ARSeg datasets yielded comparable Dice scores for nnU-Net and U-Mamba. However, HCC-ARSeg provided superior SASD results for nnU-Net, indicating its better balance between segmentation accuracy and boundary precision, while HCC-TACE-Seg produced less precise boundary contours.

The modified HCC-ARSeg dataset resulted in one of the lowest SASD (3.903), demonstrating improved boundary precision. However, this setup slightly reduced the Dice score, suggesting that the ROI-based approach might discard some broader contextual information necessary for accurate tumor segmentation. Nevertheless, this configuration shows promise for enhancing computational efficiency and localization accuracy.

In conclusion, the Table 3 underscores nnU-Net's consistent superiority across the different datasets. Notably, it demonstrates strong and comparable performance on the datasets and showcase that results achieved on HCC-ARSeg are in line with those obtained from other well-established datasets. While SwinUNetr struggles, U-Mamba emerges as a promising alternative.

The class distribution shown in Table 1 reveals a significant class imbalance in the HCC-ARSeg dataset, with tumor pixels constituting only 0.06% of all pixels and liver pixels only 1.94%. This is the lowest tumor proportion among all datasets analyzed and is 20 times smaller than the tumor presence in the HCC-TACE-Seg dataset. This imbalance present a serious challenge for deep learning models, which often struggles to detect under-represented classes. To the best of our knowledge, this explains the drop in performance on HCC-ARSeg. However, the performance remains in line with the results achieved on the other more balanced datasets and, in terms of SASD and the best performing architecture, our dataset achieve the second best result. Additionally, we analyzed the results with expert radiologists and observed that the lowest segmentation performance was more strongly influenced by tumor shape and variability in scan acquisition. In particular, smoother tumor shapes allowed for more precise segmentation and better contour definition.

Given comparable performance, the HCC-ARSeg dataset's primary strength lies in its inclusion of all three contrast phases — arterial, portal-venous, and delayed — each meticulously annotated by a resident and subsequently evaluated and refined by an expert radiologist. This comprehensive approach is particularly crucial for HCC, as it is a neoplasm in which, unlike other tumors, a non-invasive diagnosis is possible by imaging in certain clinical scenarios, such as in high-risk patients with cirrhosis, chronic hepatitis B virus, or a history of previous HCC[2]. This is due to specific radiological signs, like non-rim arterial phase hyperenhancement, non-peripheral washout, enhancing capsule, and threshold growth, that have demonstrated high sensitivity and specificity for HCC diagnosis. Verifying these criteria demands optimal multi-phasic imaging, involving CT or MR acquisitions across distinct, codified phases: arterial, portal-venous, and delayed[3]. Detecting these signs and achieving the highest diagnostic accuracy necessitates a combined evaluation of all post-contrast acquisitions. Based on this established literature, an efficient dataset should clearly incorporate all the different, precisely annotated post-contrast phases. The HCC-ARSeg dataset represents a valuable step in this direction, strengthening the training potential of software designed to automatically recognize HCC lesions.

Future work will focus on developing a multi-phasic model. This model will simultaneously process the three contrast phases to generate more precise segmentation masks. The HCC-ARSeg dataset is particularly well-suited for this development. Furthermore, we intend to expand this dataset by adding new cases encompassing a wider variety of tumor types and other liver diseases. Each new case will include multiple contrast-enhanced phases, accompanied by expert annotations for each phase. This expansion will significantly enhance the dataset's clinical relevance and diversity. Consequently, a primary objective of future work is to develop a model capable of accurately classifying various hepatic pathologies, with a strong emphasis on differentiating HCC from other hepatic malignancies and non-malignant conditions. Moreover, recent advances in foundation models for medical image segmentation, such as VISTA3D[28], have demonstrated superior performance across a variety of tasks. While this study focused on traditional architectures to ensure a fair comparison with established state-of-the-art methods, future investigations aim at incorporating such models into the framework to verify their effectiveness in HCC segmentation and directly compare their performance with conventional approaches.

Data availability

The LiTS dataset analyzed during the current study is publicly available on the Medical Segmentation Decathlon website at http://medicaldecathlon.com/. The HCC-TACE-Seg dataset is publicly available through The Cancer Imaging Archive at https://www.cancerimagingarchive.net/collection/hcc-tace-seg/. The WAW-TACE dataset is available on Zenodo at https://zenodo.org/records/12741586. The HCCARSeg dataset used and analysed during the current study is available from the corresponding author upon reasonable request.

Received: 15 July 2025; Accepted: 27 October 2025
Published online: 25 November 2025

References

1. Vogel, A., Meyer, T., Sapisochin, G., Salem, R. & Saborowski, A. Hepatocellular carcinoma. *Lancet* **400**(10360), 1345–1362. https://doi.org/10.1016/S0140-6736(22)01200-4 (2022) arXiv:3608.4663.
2. Cannella, R., Zins, M. & Brancatelli, G. ESR essentials: Diagnosis of hepatocellular carcinoma-practice recommendations by ESGAR. *Eur. Radiol.* **34**. https://doi.org/10.1007/s00330-024-10606-w (2024).
3. Kambadakone, A., Fung, A., Gupta, R., Hope, T., Fowler, K., Lyshchik, A., Ganesan, K., Yaghmai, V., Guimaraes, A., Sahani, D. & Miller, F. Li-RADS technical requirements for CT, MRI, and contrast-enhanced ultrasound. *Abdomin. Radiol.* **43**. https://doi.org/10.1007/s00261-017-1325-y (2018).
4. Hettihewa, K., Kobchaisawat, T., Tanpowpong, N. & Chalidabhongse, T. Manet: A multi-attention network for automatic liver tumor segmentation in computed tomography (CT) imaging. *Sci. Rep.* **13**. https://doi.org/10.1038/s41598-023-46580-4 (2023).
5. Hu, Q., Chen, Y., Xiao, J., Sun, S., Chen, J., Yuille, A. & Zhou, Z. Label-Free Liver Tumor Segmentation. arXiv:2303.14869 (2023).
6. Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J., Moreau, J., Osswald, A., Bouhadjar, M. & Marescaux, J. 3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. In Technical Report, IRCAD. Vol. 1(1) (2010).
7. ...Bilic, P. et al. The liver tumor segmentation benchmark (LITS). *Med. Image Anal.* **84**, 102680. https://doi.org/10.1016/j.media.2022.102680 (2023).
8. Lou, M., Ying, H., Liu, X., Zhou, H.-Y., Zhang, Y. & Yu, Y. SDR-Former: A Siamese Dual-Resolution Transformer for Liver Lesion Classification Using 3D Multi-Phase Imaging. arXiv:2402.17246 (2024).
9. Moawad, A., Morshid, A., Khalaf, A., Elmohr, M., Hazle, J., Fuentes, D., Badawy, M., Kaseb, A., Hassan, M., Mahvash, A., Szklaruk, J., Qayyum, A., Abusaif, A., Bennett, W., Nolan, T., Camp, B. & Elsayes, K. Multimodality annotated hepatocellular carcinoma data set including pre- and post-tace with imaging segmentation. *Sci. Data* **10**(1) (2023). https://doi.org/10.1038/s41597-023-01928-3.
10. Bartnik, K. et al. WAW-TACE: A hepatocellular carcinoma multiphase CT dataset with segmentations, radiomics features, and clinical data. *Radiol. Artif. Intel.* **6**(6), 240296. https://doi.org/10.1148/ryai.240296 (2024) (**PMID: 39441110**).
11. Gross, M., Arora, S., Huber, S., Kuecuekkaya, A. & Onofrey, J. LiverHccSeg: A publicly available multiphasic MRI dataset with liver and HCC tumor segmentations and inter-rater agreement analysis. *Data Brief* **51**, 109662. https://doi.org/10.1016/j.dib.2023.109662 (2023).
12. Quinton, F., Popoff, R., Presles, B., Leclerc, S., Meriaudeau, F., Nodari, G., Lopez, O., Pellegrinelli, J., Chevallier, O., Ginhac, D., Vrigneaud, J.-M. & Alberini, J.-L. A tumour and liver automatic segmentation (atlas) dataset on contrast-enhanced magnetic resonance imaging for hepatocellular carcinoma. *Data* **8**(5). https://doi.org/10.3390/data8050079 (2023).
13. Isensee, F., Jaeger, P. F., Full, P. M., Vollmuth, P. & Maier-Hein, K. H. nnU-Net for Brain Tumor Segmentation. *arXiv* arXiv:2011.00848 (2020).
14. Rahman, H., Bukht, T. F. N., Imran, A., Tariq, J., Tu, S. & Alzahrani, A. A deep learning approach for liver and tumor segmentation in CT images using ResUNet . *Bioengineering* **9**(8). https://doi.org/10.3390/bioengineering9080368 (2022).
15. Khoshkhabar, M., Meshgini, S. & Afrouzian, R. Automatic liver tumor segmentation from CT images using graph convolutional network. *Sensors* **23**, 7561. https://doi.org/10.3390/s23177561 (2023).
16. Wasserthal, J., Breit, H.-C., Meyer, M. T., Pradella, M., Hinck, D., Sauter, A. W., Heye, T., Boll, D. T., Cyriac, J., Yang, S., Bach, M. & Segeroth, M. Totalsegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiol. Artif. Intell.* **5**(5). https://doi.org/10.1148/ryai.230024 (2023).
17. Hettihewa, K., Kobchaisawat, T., Tanpowpong, N. & Chalidabhongse, T. Manet: A multi-attention network for automatic liver tumor segmentation in computed tomography (CT) imaging. *Sci. Rep.* **13**. https://doi.org/10.1038/s41598-023-46580-4 (2023).
18. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. & Xu, D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. arXiv:2201.01266 (2022).
19. Tang, Y., Yang, D., Li, W., Roth, H., Landman, B., Xu, D., Nath, V. & Hatamizadeh, A. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. arXiv:2111.14791 (2022).
20. Ma, J., Li, F. & Wang, B. U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation. arXiv:2401.04722 (2024).
21. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K. & Jaeger, P. F. nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. arXiv:2404.09556 (2024).
22. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211. https://doi.org/10.1038/s41592-020-01008-z (2021).
23. Ferrante, M. et al. Application of nnU-Net for automatic segmentation of lung lesions on CT images and its implication for radiomic models. *J. Clin. Med.* **11**(24), 7334. https://doi.org/10.3390/jcm11247334 (2022).

24. Pettit, R. W., Marlatt, B. B., Corr, S. J., Havelka, J. & Rana, A. nnU-Net deep learning method for segmenting parenchyma and determining liver volume from computed tomography images. *Ann. Surg. Open* **3**(2), 155. https://doi.org/10.1097/AS9.0000000000000155 (2022).
25. Xu, Y. et al. PA-ResSeg: A phase attention residual network for liver tumor segmentation from multiphase CT images. *Med. Phys.* **48**(7), 3752–3766. https://doi.org/10.1002/mp.14922 (2021).
26. Wasserthal, J. et al. Totalsegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiol. Artif. Intell.* **5**(5), 230024. https://doi.org/10.1148/ryai.230024 (2023).
27. Chernyak, V. et al. Liver imaging reporting and data system (LI-RADS) version 2018: Imaging of hepatocellular carcinoma in at-risk patients. *Radiology* **289**(3), 816–830. https://doi.org/10.1148/radiol.2018181494 (2018) arXiv:3025.1931.
28. He, Y., Guo, P., Tang, Y., Myronenko, A., Nath, V., Xu, Z., Yang, D., Zhao, C., Simon, B., Belue, M., Harmon, S., Turkbey, B., Xu, D. & Li, W. Vista3d: A unified segmentation foundation model for 3D medical imaging. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20863–20873. https://doi.org/10.1109/CVPR52734.2025.01943 (2025).

Author contributions

I.M., E.Z., and F.F. are computer science and robotics engineers. I.M. implemented the network architectures, conducted the technical evaluation, and analyzed the results. She also wrote and reviewed the manuscript. E.Z. contributed to data analysis and reviewed the manuscript. F.F. conceived the study design and wrote and reviewed the manuscript. A.P. and M.M. are radiologists. A.P. formulated the problem from a radiologist's perspective and contributed to the dataset creation by evaluating the annotation quality and results. She also reviewed the manuscript. M.M. contributed to data analysis and meticulously annotated the data. G.E., general surgeon, conceived the study design and provided data. He defined the requirements and specifications and guided the overall study. F.DB., general surgeon, reviewed the manuscript.

Funding

The research leading to these results has received funding from the Italian Ministry of University and Research under the D.D. 1236 (August, 1 2023), Fondo italiano per la scienza - Bando FIS 2 (TRAMIS, CUP E53C24003820001) to Federica Ferraguti. Moreover, it has received funding from the European Union - NextGenerationEU through the Italian Ministry of University and Research under PNRR - M4C2-I1.3 Project PE_00000019 "HEAL ITALIA" to Fabrizio di Benedetto and Ilaria Manghi, CUP E93C22001860006. The views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Declarations

Competing interests

The authors declare no competing interests.

Ethics committee evaluation

The Area Vasta Emilia Nord Ethics Committee, at its meeting convened on July 23, 2024, provided a favorable decision concerning the implementation of this study (283/2024/DISP/AOUMO SIRER ID 7555-AI for HCC).

Additional information

Correspondence and requests for materials should be addressed to I.M.