A Holistic Weakly Supervised Approach for Liver Tumor Segmentation with Clinical

Knowledge-Informed Label Smoothing

Hairong Wang*, Lingchao Mao*, Zihan Zhang, Jing Li

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of

Technology, Atlanta, GA, USA

*Contributed equally to the manuscript

Corresponding Author: Jing Li (Email: jing.li@isye.gatech.edu)

Abstract

Liver cancer is a leading cause of mortality worldwide, and accurate CT-based tumor

segmentation is essential for diagnosis and treatment. Manual delineation is time-intensive,

prone to variability, and highlights the need for reliable automation. While deep learning

has shown promise for automated liver segmentation, precise liver tumor segmentation

remains challenging due to the heterogeneous nature of tumors, imprecise tumor margins,

and limited labeled data. We present a novel holistic weakly supervised framework that

integrates clinical knowledge to address these challenges with (1) A knowledge-informed

label smoothing technique that leverages clinical data to generate smooth labels, which

regularizes model training reducing the risk of overfitting and enhancing model

performance; (2) A global and local-view segmentation framework, breaking down the task

into two simpler sub-tasks, allowing optimized preprocessing and training for each; and (3)

Pre- and post-processing pipelines customized to the challenges of each subtask, which

enhances tumor visibility and refines tumor boundaries. We evaluated the proposed method

on the HCC-TACE-Seg dataset and showed that these three key components

complementarily contribute to the improved performance. Lastly, we prototyped a tool for

automated liver tumor segmentation and diagnosis summary generation called

1

MedAssistLiver. The app and code are published at https://github.com/lingchm/medassist-liver-cancer.

Keywords: Knowledge-informed machine learning; label smoothing; liver tumor segmentation; deep learning.

1. Introduction

Liver cancer is a significant global health concern, affecting over 800,000 people annually and standing among the leading causes of cancer-related deaths worldwide (American Cancer Society, 2024). Liver is also a common destination for metastatic cancer cells originating from various abdominal organs, including the colon, rectum, pancreas, as well as distant organs such as the breast and lung. Consequently, a thorough examination of the liver and its lesions is critical to comprehensive tumor staging and management strategies. Standard tumor assessment protocols, such as the Response Evaluation Criteria in Solid Tumor (RECIST), require precise measurement of the diameter of the largest target lesion (Eisenhauer et al., 2009). Thus, accurate localization and precise segmentation of liver tumors within CT scans are essential for effective diagnosis, treatment planning, and monitoring of therapeutic response in patients with liver cancer (Shiina et al., 2018; Terranova & Venkatakrishnan, 2024; Virdis et al., 2019).

Manual delineation of target lesions in CT scans is fraught with challenges, being both time-consuming and prone to poor reproducibility and operator-dependent variability (Gul et al., 2022). Automated liver tumor segmentation can provide clinicians with rapid and consistent tumor delineation, thereby improving patient outcomes and reducing healthcare costs. Recently, deep learning algorithms have shown promise for producing automated liver and tumor segmentation (Gul et al., 2022). While many algorithms achieved exceptional performance in liver segmentation, with dice scores ranging from 0.90 to 0.96, enhancing liver tumor segmentation remains a challenge, currently standing at dice scores from 0.41 to 0.67 according to a recent Liver Tumor Segmentation Benchmark (Bilic et al., 2023). Liver tumor segmentation is an inherently challenging task because tumors vary significantly in size, shape, and location across different patients, which leads to a broad spectrum of tumor characteristics and hinders model generalization (Sabir et al., 2022). Moreover, margins of some tumors are imprecise as CT scans exhibit low gentile brightness and

roughness making it difficult to distinguish between tumorous and healthy liver tissue (Sabir et al., 2022). The scarcity of labeled data, coupled with the variations in lesion-to-background contrast, the coexistence of different lesion types, and disease-specific variability introduces substantial technical challenges in developing deep learning algorithms for accurate liver tumor segmentation (Moghbel et al., 2018).

Various studies have emerged to tackle the challenges of tumor segmentation. Recent advancements have primarily focused on optimizing network architecture design to enhance segmentation performance. Some researchers have integrated components like a variational auto-encoder (VAE) branch to regularize shared layers, thereby stabilizing the training process and improving generalization (Myronenko, 2018). Others have explored more complex architectures such as combining ResNet and UNet models, leveraging ResNet's deep feature extraction capabilities with UNet's segmentation proficiency, achieving gains in quality and computational efficiency (Rahman et al., 2022). Similarly, transformers have been employed to capture effective features across varying MRI contrast resolutions, enhancing segmentation accuracy through their self-attention mechanisms (H. Chen et al., 2023; Hatamizadeh et al., 2022). To address the challenge of limited labeled data in medical image segmentation, researchers have increasingly adopted methods that incorporate unlabeled data. For instance, tumor bounding boxes from unlabeled data have been used to guide the student module in a teacher-student learning framework, effectively improving model performance with minimal manual annotation (D. Zhang et al., 2021). Other works have leveraged software-generated pseudo-labels for pixel-wise supervision, allowing models to learn from large datasets with reduced human intervention (Lyu et al., 2022). Fine-tuning models using a combination of unlabeled data and generated pseudo-labels has also been shown to yield continuous improvements in segmentation accuracy (S. Chen et al., 2024). Moreover, some studies have focused on pooling data from multiple sites to enhance model generalization through transfer learning and domain adaptation (You et al., 2022). While current methods make use of increasingly complex architectures and diverse data sources, they often overlook the unique signals and characteristics inherent in the regions of interest within organs and tumors.

To tackle the common challenge of limited labeled data in cancer applications, integrating biological and biomedical domain knowledge—such as biological principles, mathematical models, and knowledge

graphs—presents a promising avenue to alleviate data shortages in training DL models. Depending on the form of domain knowledge, various approaches have been proposed for its integration (Mao et al., 2024). For instance, some researchers have developed customized architectures based on biological knowledge, where nodes represent biological entities and edges encode relationship between them, activating only connections that follow known biological pathways (Fortelny & Bock, 2020; Zhao et al., 2021). Mathematical oncology models, such as tumor growth models, have been used to regularize loss functions (Gaw et al., 2019; L. Wang et al., 2022) or to formulate Physics Informed Neural Networks (PINN) (Kaandorp et al., 2021; Meaney et al., 2023). Additionally, probability atlases containing prior probability of organ locations have been used to guide the segmentation (Huang et al., 2022; Z. Li et al., 2022; Zheng et al., 2019). When knowledge of certain regions of images that are most relevant to the prediction is available, researchers employed attention mechanisms to guide the model in focusing on important areas while suppressing irrelevant regions, thereby emphasizing discriminative parts of the images (He et al., 2021; Tomita et al., 2019). Other researchers have proposed incorporating feature behavior knowledge as attribution priors during DL training. For example, hierarchical relationships among normal, tumorous, and immune gene modules have been used to model the regional distributions of intratumoral heterogeneity (H. Wang et al., 2024). Similarly, ordinal relationships among unlabeled samples, derived based on their proximity to the tumor, have been leveraged to generate spatial predictions from images (Mao et al., 2023; L. Wang et al., 2024). However, there are challenging scenarios where specific domain knowledge is lacking, yet other modalities that contain knowledge about the disease, such as clinical data, are available.

In this work, we propose a holistic deep learning-based framework for automated liver and tumor segmentation with a novel way to leverage clinical knowledge. Our approach is designed to address the above-mentioned challenges of model overfitting due to limited labeled data and the difficulty in accurately segmenting tumors with imprecise margins. The main contributions of this work are summarized as follows:

A Knowledge-Informed Label Smoothing Technique: We distill knowledge from clinical data to
derive a high-level clinical indicator of tumor size. This indicator is then used to smoothen the
segmentation labels during model training, allowing the model to fit clinical knowledge-informed,

denoised labels. By softening the segmentation labels, the technique mitigates overfitting and prevents the model from being overly influenced by inherent variabilities in imaging contrast and lesion characteristics.

- A Global and Local-view Segmentation Framework: We propose a sequential framework that begins
 with a global-view liver segmentation model, followed by a local-view tumor segmentation model.
 This two-step approach offers greater flexibility compared to traditional multi-class methods, as it
 enables the optimization of preprocessing, training, and post-processing for each segmentation task.
- Customized pre- and post-processing for liver and tumor segmentation: We designed specialized pre-processing pipelines to enhance the visibility of livers and tumors. One of the key post-processing techniques we employed is an active contour algorithm for refinement of the segmented tumor boundaries, improving the segmentation results for tumors with imprecise margins.

The remainder of this paper is organized as follows. Section 2 describes the details of the proposed framework, structured around the three key contributions. Section 3 presents a case study evaluating the effectiveness of our approach. Section 4 concludes with a summary of findings and future directions.

2. Proposed Global and Local-view Framework with Knowledge-Informed Label Smoothing

We propose an end-to-end framework consisting of three stages: (1) **Image preprocessing**, designed to normalize image intensities and enhance feature visibility for both liver and tumor models; (2) **Liver and tumor segmentation**, utilizing a two-step *global-view* and *local-view* pipeline to separately tailor liver segmentation and tumor segmentation tasks; and (3) **Postprocessing**, where the segmentation outputs are refined, including the application of an active contour adjustment for refining tumor boundaries. Each stage is tailored to address specific challenges in liver tumor segmentation. An overview of the framework is shown in Figure 1. In the following subsections, we detail the proposed knowledge-informed label smoothing technique, the global and local-view segmentation framework, and the customized pre- and post-processing pipeline for liver and tumor segmentation, respectively.

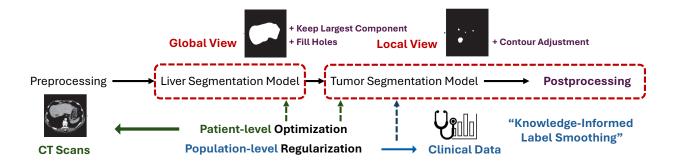


Figure 1. Overview of the proposed framework.

2.1 Clinical Knowledge-Informed Label Smoothing

The significant heterogeneity of tumor shapes, sizes, and intensities across patients is widely observed in liver tumors. A typical training approach that minimizes segmentation loss on a sample-by-sample basis can easily lead to overfitting and limit generalizability, especially when working with small datasets. To address this challenge, we introduced a novel label smoothing technique that leverages clinical knowledge-informed, denoised labels about the tumor to regularize model training and mitigate overfitting. The clinical knowledge related to tumor severity is extracted from clinical data through a regression model, which learns patterns linking clinical variables to tumor size across all patients. This knowledge is then converted into a high-level indicator of Tumor-to-Liver Volume Ratio (TLVR) that is used as a weak label during model training. We refer to this process of optimizing models based on consistency with clinical knowledge extracted across the entire patient cohort as *population-level regularization*. In contrast, we refer *patient-level optimization* as the traditional training process that optimizes models to perform precise segmentation by focusing on minimizing the loss for each individual patients' data.

2.1.1 Knowledge Extraction from Clinical Data

In the liver cancer literature, several well-studied risk factors significantly influence disease progression, including chronic infections with hepatitis B (HBV) and hepatitis C (HCV) viruses, excessive alcohol consumption, smoking, metabolic conditions such as obesity and diabetes, and a family history of liver cancer (El-Serag & Mason, 2000; Mohammadian et al., 2018). These factors not only drive the initiation and progression of tumors but also contribute to its heterogeneity. On the other hand, clinical prognostic

measurements such as the Cancer of the Liver Italian Program (CLIP) score and Alpha-fetoprotein (AFP) levels can be informative indicators of disease severity (Morshid et al., 2019). Incorporating insights from these risk factors to guide the training of tumor segmentation models can enhance model accuracy and reliability, ultimately improving diagnostic value and therapeutic decision-making.

Building on existing research regarding the risk factors for liver cancer (Mohammadian et al., 2018; Morshid et al., 2019; Shah et al., 2023), we began by identifying the relevant risk factors present in the clinical dataset that are known to potentially influence the progression of liver cancer. We then used regression models to extract clinical knowledge about the relationship between these risk factors and the tumor size for the patient cohort in the dataset. Specifically, we define the Tumor-to-Liver Volume Ratio (TLVR) to quantify the overall tumor size relative to the liver size, denoted as R_i for patient i as:

$$R_{i} = \frac{number\ of\ voxels\ in\ tumor\ mask_{i}}{(number\ of\ voxels\ in\ liver\ mask_{i} + number\ of\ voxels\ in\ tumor\ mask_{i})}.$$
 (1)

We employed linear and non-linear models to learn the relationship between the clinical risk factors and the TLVR extracted from ground truth segmentation masks. This model synthesizes clinical insights about the tumor size for this patient cohort. This intrinsic relationship established through this knowledge extraction stage is then used to regularize model training, as described in the next section.

2.1.2 Weakly Supervised Learning with Smooth Labels

To mitigate the risk of overfitting associated with *patient-level optimization* on small datasets, we introduce *population-level regularization* to encourage model's consistency with clinical knowledge extracted across the entire patient cohort. Specifically, the clinical model described in Section 2.1.1 is used to generate predictions of TLVR, denoted as \hat{R}_i , which serves as a clinically informed, denoised weak label. This weak label provides a high-level, imprecise representation of tumor size, while the ground truth tumor segmentation mask offers pixel-level, precise measurement of tumor presence.

We introduce a weakly supervised loss, L_{weak} , that measures the model's predicted tumor mask's deviation from the clinically informed TLVR:

$$L_{weak} = \frac{1}{n} \sum_{i} \left(R_i - \hat{R}_i \right)^2. \tag{2}$$

This loss term is added to the overall training objective, as described in Section 2.4. By training the model using both ground truth and weak labels, the model can consider common clinical trends across patients while optimizing for individual cases, thereby enhancing the learning process.

The weak label generation process is akin to label smoothing, a regularization technique in deep learning commonly used to cope with label noise. In computer vision applications, one-hot training labels are blended with uniform or Gaussian vectors to account for the fact that datasets may have noise in their labels, so directly maximizing the likelihood can sometimes be harmful. In other words, the output distribution is regularized in an attempt to penalize overconfident outputs. This approach has been empirically shown to improve both predictive performance and model calibration (W. Li et al., 2020; Lukasik et al., 2020). If domain knowledge about the distribution of output labels is available, incorporating this knowledge into the label smoothing process can further boost model performance. Thus, we distilled the clinical knowledge from clinical data to generate smooth weak labels of tumor size.

2.2 A Global and Local-view Segmentation Framework

While deep learning models succeeded in liver segmentation, liver tumor segmentation remains a challenging task (Ayalew et al., 2021; Bilic et al., 2023; Hänsch et al., 2022). The difficulty is exacerbated when the lesion comprises a mix of small and large tumors. To address this complexity, we adopt a two-step strategy to break down the original task of "segmenting liver tumors from abdomen" into two simpler sub-tasks: (i) "segmenting liver from abdomen" and (ii) "segmenting tumor from liver". This design assumes that the subtasks involve distinct modeling challenges and require different informative signals.

In the first sub-task, a segmentation model focuses on delineating the liver from other structures within the whole abdominal CT image (*global view*). This task requires comprehensive understanding of anatomical structures. Since the location and shape of livers compared to other organs is generally consistent across patients, existing deep learning models have solved this problem with high accuracy.

Utilizing the liver masks obtained from the first task, the CT image is cropped to a liver-focused bounding box from which a second model segments the tumor (*local view*). The model's focus narrows to the liver area, demanding the capture of precise, fine-grained feature information to distinguish between healthy and tumorous tissues within the organ. Unlike liver segmentation where every liver's location and shapes are relatively similar, tumor segmentation presents unique challenges of tumor heterogeneity in number, sizes, shapes, and intensities. We tackle this challenge by introducing a knowledge-informed label smoothing technique, as described in Section 2.1. Additionally, some tumors have imprecise boundaries that are not easily distinguishable. To tackle this issue, we employed an active contour adjustment algorithm to refine tumor contours, as detailed in Section 2.3.

This two-step approach not only trains models specialized for simpler sub-tasks but also offers the flexibility to apply different pre-processing, model architectures, and post-processing techniques tailored to the specific requirements of various segmentation tasks.

2.3 Customized pre- and post-processing for liver tumor segmentation

Effective preprocessing and post-processing steps can significantly boost model performance. Preprocessing addresses imaging heterogeneity by standardizing inputs and enhancing the visibility of relevant anatomical structures. Following segmentation, post-processing techniques refine the outputs by improving coherence in the liver class and accurately delineating complex tumor boundaries. These steps are essential to achieving precise segmentation results, as detailed in the following subsections.

2.3.1 Image preprocessing

Image preprocessing is a critical step to reduce imaging differences across CT scans and ensure that the deep learning models can learn effectively. We designed preprocessing steps to optimize the input data for both liver and tumor segmentation models, addressing challenges such as image heterogeneity and enhancing the focus on relevant anatomical structures (Figure 2).

The preprocessing pipeline for the segmentation model begins with loading raw medical imaging data and standardizing the image orientation. Hounsfield units (HU) windowing was applied to enhance the visibility of liver and tumor by adjusting intensity values to ranges recommended in the literature (Kim et al., 2020). Intensity normalization is then performed across the dataset.

For the tumor segmentation model, additional preprocessing steps were implemented after standardizing the image orientation. Assuming that liver tumors are located inside the liver (Tummala et al., 2017), the inputs to the tumor segmentation model are simplified by masking out regions outside the liver as background. This approach narrows the model's focus to distinguishing healthy and tumorous tissues within the liver, eliminating the need to navigate uninformative and potentially distractive features from surrounding structures. Given that each patient's liver varies in size, the bounding box containing the liver is randomly cropped to a fixed size. If a liver bonding box is smaller than this size, it was padded with background to ensure a consistent input size for the model.

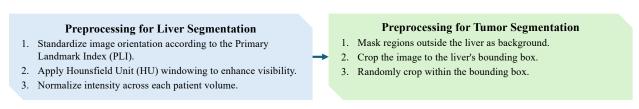


Figure 2. Customized pre-processing for liver and tumor segmentation.

2.3.2 Post-processing

In the postprocessing phase of the proposed segmentation pipeline, the segmentation results for both the liver and tumor classes are refined, with a specific focus on enhancing the coherence of the liver segmentation while cautiously handling tumor regions due to their complex characteristics. For the liver class, a strategy to retain only the largest connected components and then apply a fill holes technique was employed. This approach is motivated by the anatomical consistency of the liver, which typically forms a single, large volume within the abdominal cavity. By keeping the largest connected component, we remove any small, spurious regions that were incorrectly labeled as liver tissue during the segmentation process. Filling holes within this largest component helps to ensure that the liver's representation is solid and continuous, correcting for any internal inaccuracies that might have arisen during segmentation.

The post-processing assumptions for liver segmentation does not apply to the case of tumor. Given the inherent variability and complexity of tumor appearances, tumors can exhibit a wide range of sizes, shapes,

and may sometimes be composed of multiple disjointed parts, especially in cases of metastatic or multifocal liver disease. Applying the keep largest component and fill holes strategy might inadvertently eliminate smaller, yet clinically significant tumor regions. Instead, we employed an active contours algorithm for detecting contour without borders for refining tumor segmentation outputs (Chan & Vese, 2001; Márquez-Neila et al., 2014). In classical active contour algorithms, an edge-based energy function determines when to stop the curve evolution based on gradient of the image. However, these algorithms suffer when the object lacks clear edges or when the image is noisy. Thus, we employed an active contour algorithm that is capable of detecting smooth boundaries without relying on image gradients (Chan & Vese, 2001). This algorithm iteratively identifies the boundary curve that differentiates energy inside and outside the object. We provided the tumor segmentation outputs as the initial curve and run the algorithm for two iterations to refine the boundary. This method is adept at capturing the irregular tumor boundaries in noisy CT images, which is crucial for achieving high fidelity in the segmentation of complex tumor shapes.

2.4 Model Training and Validation Procedure

The training process for the proposed framework consists of two sequential steps. First, a liver segmentation model is trained taking entire CT scans as input. Then, a tumor segmentation model is trained on cropped liver regions delineated based on ground truth liver masks. Each segmentation stage involves its own pre- and post-processing steps, as previously described. Our framework is model agnostic, allowing researchers to choose the backbone model that best suits their dataset or to adapt state-of-the-art architectures accordingly.

The following loss was used to train the tumor segmentation model:

$$L_{tumor} = \lambda_f L_{focal} + \lambda_d L_{dice} + \lambda_w L_{weak}, \tag{3}$$

where λ_f , λ_d , λ_w are balancing terms. L_{dice} is Dice loss (Hänsch et al., 2022) commonly used for segmentation models,

$$L_{dice} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}, \tag{4}$$

where the sums run over the N voxels, of the predicted segmentation volume $p_i \in P$ and the ground truth binary volume $g_i \in G$. L_{focal} is focal loss (Lin et al., 2017) that accounts for imbalanced segmentation,

$$L_{focal} = -\sum_{i=1}^{N} (i - p_i)^{\gamma} \log(p_i),$$
 (5)

where $\gamma \geq 0$ is the focusing parameter. This combined loss function ensures that the model not only learns to accurately segment tumors based on individual patient data but also encourages the model to be consistent with clinical knowledge about tumor size extracted across the patient cohort, effectively balancing *patient-specific optimization* with *population-level regularization*. Based on the experiment results, we observed that this label smoothing technique effectively reduced overfitting, leading to a model that is more accurate and robust, as shown in Section 3.2.5.

During evaluation, each test sample follows the same two-step segmentation process, where predicted liver masks are input for tumor segmentation. No clinical features are required at inference, making the model more practical and versatile for deployment. All models were implemented in PyTorch and trained on a V100 16GB GPU. Parameter settings details are provided in Appendix A.

3. Application in Hepatocellular Carcinoma

Hepatocellular carcinoma (HCC) is the most common liver cancer, with rising incidence and limited treatment options like transarterial chemoembolization (TACE), which fails in up to 60% of cases (S. Zhang et al., 2022). We utilized the HCC-TACE-Seg dataset from the Cancer Imaging Archive (TCIA), which contains retrospectively collected pre- and post-procedural CT images of 105 confirmed HCC patients who underwent TACE between 2002 and 2012 (Moawad et al., 2021). After excluding patients with missing or misaligned segmentation files, we extracted pre-procedural CT scans of liver tumors from 98 patients, along with corresponding segmentation labels for five classes of objects: background, liver, tumor mass, portal vein, and abdominal aorta (Figure 3). Segmentations were created using AMIRA (Stalling et al., 2005) and manually curated by clinicians. To focus on the liver and tumor classes, we masked the portal vein as liver and the abdominal aorta as background. Each CT scan has spatial dimensions 512 x 512 x N, where the

number of slices N varies by patient. The dataset was randomly split into 80% training and 20% test sets, with the train-test split repeated for three replications.

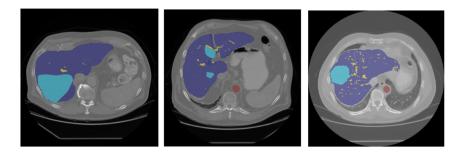


Figure 3. Sample CT scan slices from the HCC-TACE-Seg dataset with overlaid segmentation labels.

Purple=liver, blue=tumor, yellow=portal vein, red=abdominal aorta.

Additionally, a clinical data file containing related demographics, medical history, diagnosis, and treatment response is provided in HCC-TACE-Seg. From these clinical variables, we selected variables that are known risk factors of liver cancer (Mohammadian et al., 2018; Morshid et al., 2019; Shah et al., 2023). These clinical variables are detailed in Appendix B.

3.1. Clinical indicator of TLVR

Based on the experimental results, linear regression demonstrated relatively strong performance in predicting TLVR among various linear and non-linear models. Due to its interpretability, we selected linear regression for further analysis. The outcomes of the model, including the significance of each feature, are shown in Figure 4, offering insights into their respective contributions to predicting R_i . The mean squared error (MSE) of the predicted TLVR \hat{R}_i is 0.012, and the mean absolute error (MAE) is 0.087. Notably, the Pearson correlation coefficient between R_i and \hat{R}_i is 0.820, indicating a strong correlation and highlighting the predictive power of the clinical features in relation to TLVR. To minimize the risk of overfitting and enhance model interpretability, we selected the features with the highest contributions for TLVR prediction. The detailed feature selection process is outlined in Appendix B.

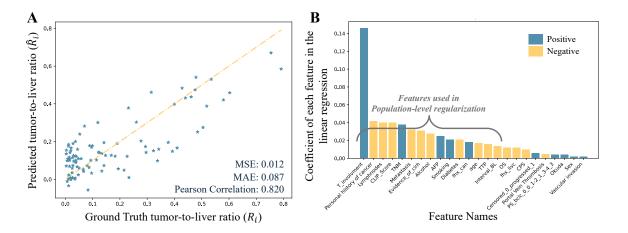


Figure 4. (A) The results of the linear regression, showcasing the relationship between clinical features and the tumor-to-liver volume ratio (TLVR). (B) The coefficient plot indicating feature contributions.

3.2. Liver and tumor segmentations

In this section, we present the segmentation results for both liver and tumor.

3.2.1. Segmentation Evaluation Metrics

We evaluated segmentation performance class-wise for liver and tumor using the following metrics:

- Accuracy: This metric measures the proportion of true results (both true positives and true negatives) within the total number of pixels in each 3D CT image. It is calculated as: $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$, where TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively.
- **Dice Coefficients (Dice)**: The Dice coefficient is a statistic used to gauge the similarity of two samples. It is particularly useful in the field of image segmentation. The Dice coefficient can be defined as: $Dice = \frac{2 \times TP}{2 \times TP + FP + FN}$. This metric highlights the importance of true positives by considering the size of the intersection over the average size of two samples.
- Intersection over Union (IoU): The IoU is an evaluation metric that measures the overlap between the predicted segmentation and the ground truth. It is defined as: $IoU = \frac{TP}{TP + FP + FN}$.

• Sensitivity: Sensitivity measures the proportion of actual positives that are correctly identified. It is particularly crucial in medical image analysis for identifying conditions (e.g. tumors), where missing a positive case can have serious consequences. Sensitivity is calculated as: $sensitivity = \frac{TP}{TP+FN}$. This metric indicates the model's ability to correctly detect positives out of all actual positive cases.

3.2.2. Backbone Model Search

We benchmarked various state-of-the art networks for medical image segmentation on this dataset to select the best backbone model. The competing models included:

- UNet (Kerfoot et al., 2019): an improved version of U-Net with residual units;
- UNet++ (Zhou et al., 2018): an improved U-Net with nested skip pathways between encoder and decoder blocks at multiple resolutions;
- SegResNet and SegResNetVAE (Myronenko, 2018): an asymmetric network featured by a large encoder and a compact decoder, and an optional variational auto-encoder to regularize reconstruction;
- MA-Net (Fan et al., 2020): a variant of U-Net with multi-scale attention network;
- Swin UNETR (Hatamizadeh et al., 2022): a transformer-based network that extracts features at five different resolutions.

We trained two versions of the UNet model: a four-layer UNet with 32, 64, 128, and 256 channels in each upsampling layer, respectively, and two residual units; and a five-layer UNet with an additional 16-channel layer and four residual units, doubling the number of parameters. Other models followed the default architecture in their original papers. We applied instance normalization and a dropout probability of 0.2 across all models. While both 2D and 3D models were trained, we primarily focused on 3D models for tumor segmentation, as they are better suited to capturing the tumor's varying shape in the 3D space. Knowledge-informed label smoothing and special post-processing for tumors were not applied at this stage to select the backbone model with the best off-shelf performance.

The comparison of different backbone models for liver and tumor segmentation is summarized in Table 1 and Table 2. Swin UNETR struggled to converge on this small dataset, so its results were not reported. The best performing backbone for liver segmentation was SegResVAE, achieving Dice of 0.942, which is aligned with the state of the art performance reported in other datasets in the literature (Bilic et al., 2023). For tumor segmentation, SegResNet and SegResNetVAE significantly outperformed other models in terms of Dice score and sensitivity, demonstrating their ability to accurately detect tumors and minimize false negatives. We observed that more complex models, such as U-Net Large and MA-Net, did not outperform smaller models, likely due to the challenge of learning to segment heterogeneous tumors from a small dataset. Given these results, we chose SegResNetVAE as the backbone because it performs best overall.

SegResNetVAE is built upon an asymmetrical encoder-decoder based Convolutional Neural Networks (CNNs) architecture (Myronenko, 2018). It features an encoder designed for comprehensive image feature extraction and a compact decoder to reconstruct segmentation masks. To augment this structure, an additional variational auto-encoder (VAE) branch is integrated at the encoder's endpoint. This branch is designed for reconstructing the original image, thereby providing extra guidance and regularization to the encoder, a feature particularly beneficial in scenarios with limited training data. This model has not only been widely adopted and demonstrated to perform effectively in a variety of segmentation tasks (Myronenko, 2018; Pattisapu et al., 2021; Yang et al., 2021) but it has also been proven to consistently achieve high training accuracy regardless of the initial random weights (Myronenko, 2018). This robustness underscores its reliability and effectiveness across diverse medical imaging applications.

Table 1. Comparison of backbone models for liver segmentation on the test set. (2D: Training and testing data are slice-based 2D CT scans; 3D: Training and testing data are 3D CT scans.)

Model	Num. parameters	Accuracy	Dice	Sensitivity
U-Net (2D)	1,627,934	0.994	0.920	0.935
U-Net (3D)	4,809,920	0.993	0.932	0.940
SegResNet (3D)	4,697,554	0.993	0.931	0.954
SegResNetVAE (3D)	9,258,756	0.995	0.942	0.955

Table 2. Comparison of backbone models for tumor segmentation on the test set. (2D: Training and testing data are slice-based 2D CT scans; 3D: Training and testing data are 3D CT scans.)

Model	Num. parameters	Accuracy	Dice	Sensitivity
U-Net (3D)	4,809,920	0.962	0.443	0.612
U-Net (3D) Large	9,442,862	0.968	0.341	0.451
U-Net ++ (3D)	6,976,140	0.960	0.397	0.562
MA-Net (2D)	22,453,723	0.988	0.486	0.576
SegResNet (3D)	4,697,571	0.976	0.539	0.715
SegResNetVAE (3D)	9,258,756	0.972	0.501	0.737

3.2.3. Liver Tumor Segmentations

Using the selected backbone from the previous experiment, we evaluate the proposed segmentation framework on tumor segmentation performance. Note that the liver segmentation performance were the same as those reported in the previous section. Table 3 summarizes the segmentation performance for all tumors, as well as results separated by large and small tumors separately. As a baseline comparison, previous work by Morshid et al. (2019) achieved a Dice score of 0.67 for large tumors (≥10 cm) and 0.37 for small tumors (<10 cm) on the same dataset. Our model outperformed these results, achieving higher Dice scores for both large and small tumor segmentations. Figure 5 show sample segmentation outputs.

Table 3. Segmentation performance for small and large tumors across three replications.

	Accuracy	Dice	Sensitivity	IoU
All tumors	0.993±0.003	0.552±0.042	0.785 <u>+</u> 0.115	0.426±0.052
Large tumors (≥10cm)*	0.989±0.004	0.841±0.022	0.827±0.060	0.727±0.032
Small tumors (<10cm)*	0.998±0.001	0.406±0.045	0.698±0.076	0.287±0.033
Large tumors (≥500cm ³)**	0.987±0.002	0.799±0.037	0.759±0.038	0.675±0.040
Small tumors (<500cm ³)**	0.997±0.001	0.444±0.072	0.658±0.026	0.322±0.064

^{*}Categorization based on provided clinical data, which is unavailable for a subset of patients

^{**}Categorization based on ground truth segmentation masks

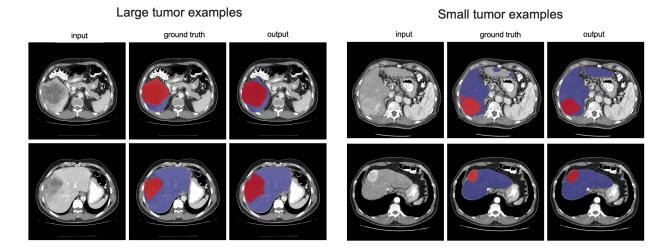


Figure 5. Sample tumor segmentation outputs.

3.2.4. Ablation Study

Next, we conduct ablation study to understand the effect of the three key components of the proposed model. To assess the value of global-local two-step approach, we compare with the standard multi-class approach in the literature that simultaneously segments both liver and tumor in a single model. Additionally, we compare with two simpler versions of the proposed model without the clinical knowledge-informed label smoothing and without the active contour algorithm for tumor segmentation boundary refinement.

Table 4 summarizes the ablation study results. The multi-class segmentation model, which simultaneously optimizes both tasks, resulted in worse liver segmentation performance than the two-step model, in which liver segmentation was optimized individually. At the same time, the multi-class model has high tumor sensitivity with low Dice suggesting elevated false positive rate rates. Notably, the two-step model without contour adjustment post-processing and knowledge-informed label smoothing underperformed the multi-class model in tumor segmentation, suggesting that tumor segmentation task is challenging and prone to overfitting. Incorporating knowledge-informed label smoothing improved tumor segmentation results, demonstrating that this regularization technique was effective in improving model generalizability. Adding contour adjustment for tumor segmentation outputs further improved tumor segmentation, proving that this technique was effective in tackling tumors with imprecise margins. The

proposed framework that incorporates all three designs achieved the best liver and tumor segmentation performance.

Table 4. Liver and Tumor segmentation performance. "smoothing": clinical knowledge-informed label smoothing; "contour adjust": active contour algorithm for tumor boundary refinement.

	Methods	Accuracy	Dice	Sensitivity	IoU
Liver	Multi-class	0.992	0.894	0.900	0.814
	Two-step	0.994	0.942	0.955	0.895
Tumor	Multi-class	0.994	0.534	0.842	0.439
	Two-step	0.993	0.472	0.513	0.347
	Two-step + smoothing	0.994	0.500	0.589	0.389
	Two-step + contour adjust	0.995	0.563	0.741	0.454
	Two-step + contour adjust + smoothing	0.995	0.570	0.801	0.460

It is worth noting that, SegResNet outperformed SegResNetVAE in tumor segmentation in our backbone model search experiments (Section 3.1.2), likely due to the latter's susceptibility to overfitting in the context of small datasets. Nevertheless, SegResNetVAE outperformed SegResNet under our framework. By integrating knowledge-informed label smoothing and special post-processing for tumor contours, we provided SegResNetVAE –the more complex model— with a robust learning framework to demonstrate its advanced pattern recognition capabilities.

3.2.5. Knowledge-informed Label Smoothing Reduced Overfitting

From our experiments, we observe that the *population-level* regularization from knowledge-informed smoothed labels reduced overfit-ting, leading to a model that is more accurate and capable of generalizing well to unseen data. To further understand the effect of knowledge-informed label smoothing in model training, we conducted a comparative analysis of the training loss and validation Dice scores using the same model under identical settings. We compared two versions of the model in Figure 6: one trained with weak supervision (in yellow), and another without knowledge-informed label smoothing (in orange). The model incorporating label smoothing exhibited a more gradual decrease in training loss and a more rapid increase in the validation dice score, whereas the model without label smoothing's training loss quickly converged

and its validation metric saturated early. Thus, label smoothing effectively reduced overfitting on the training data and improved the model's generalizability to unseen data.

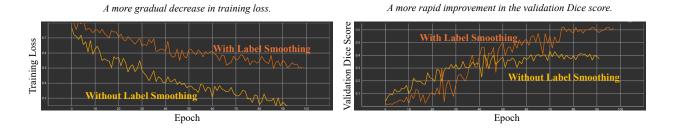


Figure 6. Training and validation curves with vs. without knowledge-informed label smoothing.

3.3. MedAssist-Liver: An Prototype for AI-Powered Clinical Tool

We developed a publicly available web-based application called *MedAssist-Liver* to illustrate how the proposed model could be used to support liver tumor diagnosis workflows. The tool can be accessed at https://lingchmao-medassist-liver-cancer.hf.space/. This user-friendly interface could serve as a supplementary tool for clinicians and researchers to analyze liver CT scans. An overview of the workflow shown in Figure 7. The process begins with users uploading a 3D CT scan, which is then displayed in an integrated image viewer. The user can choose to generate liver and/or tumor segmentations, then the tool automatically generates segmentation maps overlaid on the image. Users have the option to export segmentations outputs locally for manual review. Once segmentations are generated, the tool generates an automated diagnostic summary of the tumor, simplifying the interpretation of complex medical data. This is accomplished by extracting characteristics about the tumor as inputs into a Large Language Model (LLM) and prompting the model to generate a report commenting on diagnosis of the disease and treatment recommendations. We used the publicly available Llama 2 for the diagnosis report generation for demonstration in our prototype. More powerful LLM could be used for diagnosis report generation, as this is an ongoing area of research. Assessing the generative capability is beyond the scope of this work and we refer interested readers to this comprehensive survey (Hartsock & Rasool, 2024).

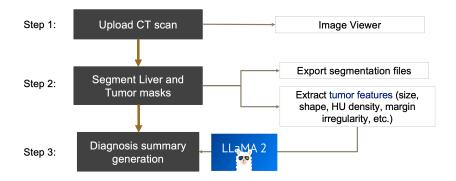


Figure 7. Workflow of MedAssist-Liver: an automated tool for liver tumor segmentation and diagnosis report generation.

4. Conclusion

This research introduces a weakly supervised framework with for liver and tumor segmentation. Each of the key components of this framework are designed to tackle existing challenges of tumor segmentation, effectively improving data efficiency, addressing model overfitting, and achieving accurate delineation of tumor boundaries. We propose a novel clinical knowledge-informed label smoothing technique that allows our model to leverage broader insights for effective learning regularization and overfitting mitigation. Our approach is also distinctive from standard multi-class segmentation approaches in the literature in its two-step global and local view segmentation process, which simplifies the original challenging task into two subtasks, thereby enhancing flexibility and enabling task-specific optimization across the preprocessing, training, and post-processing phases. We customized the pre- and post-processing pipeline separately for each subtask, including the usage of an active contour algorithm to refine tumor contours designed to tackle the imprecise margins of some tumors. These three design components synergistically contribute to the overall framework achieving state-of-the-art segmentation performance for liver tumors. Lastly, we prototyped a web application MedAssistLiver with automated liver and tumor segmentation and diagnosis summary generation capabilities to inspire potential ways these AI tools can support clinical practice.

This study has several limitations. One limitation of this study is the constrained sample size of labeled data. While the proposed model is designed to mitigate this issue by employing knowledge-informed label smoothing, the importance of a sufficient number of labeled samples for training an accurate and robust

model cannot be overstated. Expanding the number of labeled samples not only has the potential to enhance segmentation accuracy for the current backbone models but also enables training of deeper backbone models, such as Swin UNETR, which may further improve the current segmentation performance. Additionally, given demographic variations in liver cancer, the development of demographic-specific models, such as age-specific versions, may further refine segmentation outcomes. Beyond demographics, tumor size is another critical factor influencing model training. Future work could involve developing specialized models for large and small tumors to further enhance segmentation accuracy. For small tumors, a larger dataset and more precise training guidance are essential, including details on the number of lesion areas, tumor sizes, and boundaries. A second limitation is the restricted set of clinical features in this dataset. Despite this limitation, the proposed model has demonstrated the effectiveness of knowledge informed label smoothing. However, if a more comprehensive set of clinical features were available, it could provide more accurate and comprehensive knowledge guidance for model training. Although liver tumor segmentation demands further research to improve segmentation accuracy, this work highlight the potential of these AI tools to assist with tumor segmentation, paving the way for improved diagnostic and treatment strategies.

Appendix

Appendix A. Hyperparameter Settings

1. SegResNetVAE – Liver Segmentation

Parameters	Description	Value
input_image_size	the size of images to input into the network	(512, 512, 16)
vae_estimate_std	whether to estimate the standard deviations in VAE	False
vae_default_std	if not to estimate the std, use the default value	0.3
vae_nz	number of latent variables in VAE	256
spatial_dims	spatial dimension of the input data	3
blocks_down	number of down sample blocks in each layer	[1, 2, 2, 4]
blocks_up	number of up sample blocks in each layer	[1, 1, 1]
init_filters	number of output channels for initial convolution layer	16
in_channels	number of input channels for the network	1
norm	feature normalization type	'instance'
out_channels	number of output channels for the network	2
dropout_prob	probability of an element to be zero-ed	0.2

hu_range	an interval of intensity values on the Hounsfield scale to	(-150, 250)
	enhance the contrast of target tissues	

2. SegResNetVAE – Tumor Segmentation

Parameters	Description	Value
input_image_size	the size of images to input into the network	(256, 256, 32)
vae_estimate_std	whether to estimate the standard deviations in VAE	False
vae_default_std	if not to estimate the std, use the default value	0.3
vae_nz	number of latent variables in VAE	256
spatial_dims	spatial dimension of the input data	3
blocks_down	number of down sample blocks in each layer	[1, 2, 2, 4]
blocks_up	number of up sample blocks in each layer	[1, 1, 1]
init_filters	number of output channels for initial convolution layer	16
in_channels	number of input channels for the network	1
norm	feature normalization type	'instance'
out_channels	number of output channels for the network	3
dropout_prob	probability of an element to be zero-ed	0.2
hu_range	an interval of intensity values on the Hounsfield scale to	(-200, 250)
	enhance the contrast of target tissues	
lambda_weak	weight of population-level regularization	0.5 or 0.6

3. Training Setting

Parameters	Description	Value
max_epochs	max number of epochs in training	150
batch_size	batch size	1
learning_rate	learning rate	1e-4
optimizer	algorithm used for optimizing the neural network weights	Adam
weight_decay	weight decay (L2 penalty) of the optimizer	1e-5
lr scheduler	strategy to adjust the learning rate during training	CosineAnnealingLR

Appendix B. Feature Selection for TLVR

According to feature contributions in linear regression, we selected the top n features most predictive for R_i using 5-fold cross-validation (CV). The number of features n was determined as 15 by the Pearson Correlation coefficient between R_i and \hat{R}_i based on 5-fold CV. The results are shown in Figure B1. Table B1 displays selected top 15 most predictive features in descending order of importance, along with their descriptions.

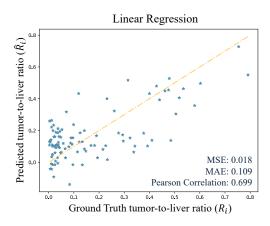


Figure B1. The results of the linear regression.

In our analysis, we aimed to identify the clinical features most predictive of the TLVR R_i . To achieve this, we employed a rigorous selection process based on the contributions of features in linear regression models. Specifically, we selected the top n features that demonstrated the highest predictive power for R_i , utilizing a 5-fold cross-validation (CV). The optimal number of features, n, was determined to be 15, based on achieving the highest Pearson Correlation coefficient between R_i and \hat{R}_i across the 5-fold CV.

The effectiveness of our feature selection process is underscored by the predictive accuracy of the linear regression model. By focusing on the most predictive features identified through our analysis, we were able to achieve a high level of accuracy in estimating TLVR. This precision is particularly important for our subsequent application of the predicted ratio in population-level regularization, which aims to enhance the generalizability of the model for segmentations.

Table B1. Clinical variables included for knowledge extraction.

Clinical Feature	Description
T_involvement	Liver involvement by the tumor, either less than 50% involvement of
	liver or more.
Personal history of cancer	Whether the patient has cancer before.
Lymphnodes	Whether the patient has lymphnodes.
CLIP_Score	CLIP score.
TNM	TNM staging.
Metastasis	Whether cancer cells spread from the original tumor site to distant parts
	of the body, forming new tumors in organs or tissues.
Evidence_of_cirh	Whether there is evidence of late stage of scarring (fibrosis) of the liver.
Alcohol	Whether the patient drinks alcohol or not.
AFP	Alpha fetoprotein level (ng/ml) obtained from blood test.

Smoking	Whether the patient smokes or not.
Diabetes	Whether the patient has diabetes.
fhx_can	Whether the patient's family has history of cancer.
age	The age of patient.
TTP	Number of weeks between the procedure and the first evidence of
	tumor progression.
Interval_BL	Number of Days between HCC diagnosis (either by previous imaging
	or biopsy) and pre- procedural (baseline) CT.

References:

- American Cancer Society. (2024). Key Statistics About Liver Cancer. *All about Cancer*. https://www.cancer.org/cancer/types/liver-cancer/about/what-is-key-statistics.html
- Ayalew, Y. A., Fante, K. A., & Mohammed, M. A. (2021). Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method. *BMC Biomedical Engineering*, 3(1), 4. https://doi.org/10.1186/s42490-021-00050-y
- Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G. E. H., Chartrand, G., Lohöfer, F., Holch, J. W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M. M., Vivanti, R., ... Menze, B. (2023). The Liver Tumor Segmentation Benchmark (LiTS). *Medical Image Analysis*, 84, 102680. https://doi.org/10.1016/j.media.2022.102680
- Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2), 266–277. IEEE Transactions on Image Processing. https://doi.org/10.1109/83.902291
- Chen, H., An, J., Jiang, B., Xia, L., Bai, Y., & Gao, Z. (2023). WS-MTST: Weakly Supervised Multi-Label Brain Tumor Segmentation With Transformers. *IEEE Journal of Biomedical and Health Informatics*, 27(12), 5914–5925. IEEE Journal of Biomedical and Health Informatics. https://doi.org/10.1109/JBHI.2023.3321602
- Chen, S., Lin, L., Cheng, P., & Tang, X. (2024). ASLSEG: Adapting Sam in the Loop for Semi-Supervised Liver Tumor Segmentation. 2024 IEEE International Symposium on Biomedical Imaging (ISBI), 1–5. https://doi.org/10.1109/ISBI56570.2024.10635501
- Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., Shankar, L., Dodd, L., Kaplan, R., Lacombe, D., & Verweij, J. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2), 228–247. https://doi.org/10.1016/j.ejca.2008.10.026

- El-Serag, H. B., & Mason, A. C. (2000). Risk Factors for the Rising Rates of Primary Liver Cancer in the United States. *Archives of Internal Medicine*, 160(21), 3227–3230. https://doi.org/10.1001/archinte.160.21.3227
- Fan, T., Wang, G., Li, Y., & Wang, H. (2020). MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. *IEEE Access*, 8, 179656–179665. https://doi.org/10.1109/ACCESS.2020.3025372
- Fortelny, N., & Bock, C. (2020). Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biology*, 21(1), 190. https://doi.org/10.1186/s13059-020-02100-5
- Gaw, N., Hawkins-Daarud, A., Hu, L. S., Yoon, H., Wang, L., Xu, Y., Jackson, P. R., Singleton, K. W., Baxter, L. C., Eschbacher, J., Gonzales, A., Nespodzany, A., Smith, K., Nakaji, P., Mitchell, J. R., Wu, T., Swanson, K. R., & Li, J. (2019). Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI. *Scientific Reports*, 9(1), 10063. https://doi.org/10.1038/s41598-019-46296-4
- Gul, S., Khan, M. S., Bibi, A., Khandakar, A., Ayari, M. A., & Chowdhury, M. E. H. (2022). Deep learning techniques for liver and liver tumor segmentation: A review. *Computers in Biology and Medicine*, 147, 105620. https://doi.org/10.1016/j.compbiomed.2022.105620
- Hänsch, A., Chlebus, G., Meine, H., Thielke, F., Kock, F., Paulus, T., Abolmaali, N., & Schenk, A. (2022). Improving automatic liver tumor segmentation in late-phase MRI using multi-model training and 3D convolutional neural networks. *Scientific Reports*, 12(1), 12262. https://doi.org/10.1038/s41598-022-16388-9
- Hartsock, I., & Rasool, G. (2024). Vision-Language Models for Medical Report Generation and Visual Question Answering: A Review (arXiv:2403.02469). arXiv. https://doi.org/10.48550/arXiv.2403.02469
- Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., & Xu, D. (2022). Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images (arXiv:2201.01266). arXiv. https://doi.org/10.48550/arXiv.2201.01266
- He, K., Ji, W., Zhou, T., Li, Z., Huo, J., Zhang, X., Gao, Y., Shen, D., Zhang, B., & Zhang, J. (2021). Cross-Modality Brain Tumor Segmentation via Bidirectional Global-to-Local Unsupervised Domain Adaptation (arXiv:2105.07715). arXiv. http://arxiv.org/abs/2105.07715
- Huang, H., Chen, Q., Lin, L., Cai, M., Zhang, Q., Iwamoto, Y., Han, X., Furukawa, A., Kanasaki, S., Chen, Y.-W., Tong, R., & Hu, H. (2022). MTL-ABS ³ Net: Atlas-Based Semi-Supervised Organ Segmentation Network With Multi-Task Learning for Medical Images. *IEEE Journal of*

- *Biomedical and Health Informatics*, *26*(8), 3988–3998. https://doi.org/10.1109/JBHI.2022.3153406
- Kaandorp, M. P. T., Barbieri, S., Klaassen, R., Van Laarhoven, H. W. M., Crezee, H., While, P. T., Nederveen, A. J., & Gurney-Champion, O. J. (2021). Improved unsupervised physics-informed deep learning for intravoxel incoherent motion modeling and evaluation in pancreatic cancer patients. *Magnetic Resonance in Medicine*, 86(4), 2250–2265. https://doi.org/10.1002/mrm.28852
- Kerfoot, E., Clough, J., Oksuz, I., Lee, J., King, A. P., & Schnabel, J. A. (2019). Left-Ventricle Quantification Using Residual U-Net. In M. Pop, M. Sermesant, J. Zhao, S. Li, K. McLeod, A. Young, K. Rhode, & T. Mansi (Eds.), Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges (pp. 371–380). Springer International Publishing. https://doi.org/10.1007/978-3-030-12029-0 40
- Kim, K.-S., Shim, M.-C., Park, M.-H., Yun, J.-Y., & Lee, K.-A. (2020). Asymmetry in the Mechanical Properties of Block Ni-Cr-Al Superalloy Foam Fabricated by the Combination of Powder Alloying and Hot Rolling Processes. *Korean Journal of Metals and Materials*, 58(2), 103–111.
- Li, W., Dasarathy, G., & Berisha, V. (2020). Regularization via Structural Label Smoothing. *Proceedings* of the Twenty Third International Conference on Artificial Intelligence and Statistics, 1453–1463. https://proceedings.mlr.press/v108/li20e.html
- Li, Z., Li, Z., Liu, R., Luo, Z., & Fan, X. (2022). Coupling Deep Deformable Registration with Contextual Refinement for Semi-Supervised Medical Image Segmentation. 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 1–5. https://doi.org/10.1109/ISBI52829.2022.9761683
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal Loss for Dense Object Detection.
- Lukasik, M., Bhojanapalli, S., Menon, A. K., & Kumar, S. (2020, March 5). *Does label smoothing mitigate label noise?* arXiv.Org. https://arxiv.org/abs/2003.02819v1
- Lyu, F., Ma, A. J., Yip, T. C.-F., Wong, G. L.-H., & Yuen, P. C. (2022). Weakly Supervised Liver Tumor Segmentation Using Couinaud Segment Annotation. *IEEE Transactions on Medical Imaging*, 41(5), 1138–1149. IEEE Transactions on Medical Imaging. https://doi.org/10.1109/TMI.2021.3132905
- Mao, L., Wang, H., Hu, L. S., Tran, N. L., Canoll, P. D., Swanson, K. R., & Li, J. (2024). Knowledge-Informed Machine Learning for Cancer Diagnosis and Prognosis: A review (arXiv:2401.06406). arXiv. https://doi.org/10.48550/arXiv.2401.06406
- Mao, L., Wang, L., Hu, L. S., Eschbacher, J. M., Leon, G. D., Singleton, K. W., Curtin, L. A., Urcuyo, J.,
 Sereduk, C., Tran, N. L., Hawkins-Daarud, A., Swanson, K. R., & Li, J. (2023). Weakly-Supervised
 Transfer Learning With Application in Precision Medicine. *IEEE Transactions on Automation Science and Engineering*, 1–15. https://doi.org/10.1109/TASE.2023.3323773

- Márquez-Neila, P., Baumela, L., & Alvarez, L. (2014). A Morphological Approach to Curvature-Based Evolution of Curves and Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1), 2–17. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2013.106
- Meaney, C., Das, S., Colak, E., & Kohandel, M. (2023). Deep learning characterization of brain tumours with diffusion weighted imaging. *Journal of Theoretical Biology*, 557, 111342. https://doi.org/10.1016/j.jtbi.2022.111342
- Moawad, A. W., Fuentes, D., Morshid, A., Khalaf, A. M., Elmohr, M. M., Abusaif, A., Hazle, J. D., Kaseb, A. O., Hassan, M., Mahvash, A., Szklaruk, J., Qayyom, A., & Elsayes, K. (2021). *Multimodality annotated HCC cases with and without advanced imaging segmentation* (Version 1) [Dataset]. The Cancer Imaging Archive. https://doi.org/10.7937/TCIA.5FNA-0924
- Moghbel, M., Mashohor, S., Mahmud, R., & Saripan, M. I. B. (2018). Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artificial Intelligence Review*, *50*(4), 497–537. https://doi.org/10.1007/s10462-017-9550-x
- Mohammadian, M., Mahdavifar, N., Mohammadian-Hafshejani, A., & Salehiniya, H. (2018). *Liver cancer in the world—Epidemiology, incidence, mortality and risk factors*.
- Morshid, A., Elsayes, K. M., Khalaf, A. M., Elmohr, M. M., Yu, J., Kaseb, A. O., Hassan, M., Mahvash, A., Wang, Z., Hazle, J. D., & Fuentes, D. (2019). A Machine Learning Model to Predict Hepatocellular Carcinoma Response to Transcatheter Arterial Chemoembolization. *Radiology: Artificial Intelligence*, *I*(5), e180021. https://doi.org/10.1148/ryai.2019180021
- Myronenko, A. (2018). 3D MRI brain tumor segmentation using autoencoder regularization (arXiv:1810.11654). arXiv. http://arxiv.org/abs/1810.11654
- Pattisapu, V. K., Daunhawer, I., Weikert, T., Sauter, A., Stieltjes, B., & Vogt, J. E. (2021). PET-Guided Attention Network for Segmentation of Lung Tumors from PET/CT Images. In Z. Akata, A. Geiger, & T. Sattler (Eds.), *Pattern Recognition* (pp. 445–458). Springer International Publishing. https://doi.org/10.1007/978-3-030-71278-5_32
- Rahman, H., Bukht, T. F. N., Imran, A., Tariq, J., Tu, S., & Alzahrani, A. (2022). A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet. *Bioengineering*, 9(8), Article 8. https://doi.org/10.3390/bioengineering9080368
- Sabir, M. W., Khan, Z., Saad, N. M., Khan, D. M., Al-Khasawneh, M. A., Perveen, K., Qayyum, A., & Azhar Ali, S. S. (2022). Segmentation of Liver Tumor in CT Scan Using ResU-Net. *Applied Sciences*, 12(17), 8650. https://doi.org/10.3390/app12178650
- Shah, P. A., Patil, R., & Harrison, S. A. (2023). NAFLD-related hepatocellular carcinoma: The growing challenge. *Hepatology*, 77(1), 323. https://doi.org/10.1002/hep.32542

- Shiina, S., Sato, K., Tateishi, R., Shimizu, M., Ohama, H., Hatanaka, T., Takawa, M., Nagamatsu, H., & Imai, Y. (2018). Percutaneous Ablation for Hepatocellular Carcinoma: Comparison of Various Ablation Techniques and Surgery. *Canadian Journal of Gastroenterology and Hepatology*, 2018, e4756147. https://doi.org/10.1155/2018/4756147
- Stalling, D., Westerhoff, M., & Hege, H.-C. (2005). amira: A Highly Interactive System for Visual Data Analysis. In *Visualization Handbook* (pp. 749–767). Elsevier. https://doi.org/10.1016/B978-012387582-2/50040-X
- Terranova, N., & Venkatakrishnan, K. (2024). Machine Learning in Modeling Disease Trajectory and Treatment Outcomes: An Emerging Enabler for Model-Informed Precision Medicine. *Clinical Pharmacology & Therapeutics*, 115(4), 720–726. https://doi.org/10.1002/cpt.3153
- Tomita, N., Abdollahi, B., Wei, J., Ren, B., Suriawinata, A., & Hassanpour, S. (2019). Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. *JAMA Network Open*, 2(11), e1914645. https://doi.org/10.1001/jamanetworkopen.2019.14645
- Tummala, K. S., Brandt, M., Teijeiro, A., Graña, O., Schwabe, R. F., Perna, C., & Djouder, N. (2017).
 Hepatocellular Carcinomas Originate Predominantly from Hepatocytes and Benign Lesions from Hepatic Progenitor Cells. *Cell Reports*, 19(3), 584–600.
 https://doi.org/10.1016/j.celrep.2017.03.059
- Virdis, F., Reccia, I., Di Saverio, S., Tugnoli, G., Kwan, S. H., Kumar, J., Atzeni, J., & Podda, M. (2019). Clinical outcomes of primary arterial embolization in severe hepatic trauma: A systematic review. *Diagnostic and Interventional Imaging*, 100(2), 65–75. https://doi.org/10.1016/j.diii.2018.10.004
- Wang, H., Argenziano, M. G., Yoon, H., Boyett, D., Save, A., Petridis, P., Savage, W., Jackson, P., Hawkins-Daarud, A., Tran, N., Hu, L., Al Dalahmah, O., Bruce, J. N., Grinband, J., Swanson, K. R., Canoll, P., & Li, J. (2024). Biologically-informed deep neural networks provide quantitative assessment of intratumoral heterogeneity in post-treatment glioblastoma. *Research Square*, rs.3.rs-3891425. https://doi.org/10.21203/rs.3.rs-3891425/v1
- Wang, L., Hawkins-Daarud, A., Swanson, K. R., Hu, L. S., & Li, J. (2022). Knowledge-Infused Global-Local Data Fusion for Spatial Predictive Modeling in Precision Medicine. *IEEE Transactions on Automation Science and Engineering*, 19(3), 2203–2215. https://doi.org/10.1109/TASE.2021.3076117
- Wang, L., Wang, H., D'Angelo, F., Curtin, L., Sereduk, C. P., Leon, G. D., Singleton, K. W., Urcuyo, J.,
 Hawkins-Daarud, A., Jackson, P. R., Krishna, C., Zimmerman, R. S., Patra, D. P., Bendok, B. R.,
 Smith, K. A., Nakaji, P., Donev, K., Baxter, L. C., Mrugała, M. M., ... Li, J. (2024). Quantifying
 intra-tumoral genetic heterogeneity of glioblastoma toward precision medicine using MRI and a

- data-inclusive machine learning algorithm. *PLOS ONE*, *19*(4), e0299267. https://doi.org/10.1371/journal.pone.0299267
- Yang, D., Xu, Z., Li, W., Myronenko, A., Roth, H. R., Harmon, S., Xu, S., Turkbey, B., Turkbey, E., Wang, X., Zhu, W., Carrafiello, G., Patella, F., Cariati, M., Obinata, H., Mori, H., Tamura, K., An, P., Wood, B. J., & Xu, D. (2021). Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. *Medical Image Analysis*, 70, 101992. https://doi.org/10.1016/j.media.2021.101992
- You, C., Xiang, J., Su, K., Zhang, X., Dong, S., Onofrey, J., Staib, L., & Duncan, J. S. (2022). Incremental Learning Meets Transfer Learning: Application to Multi-site Prostate MRI Segmentation. In S. Albarqouni, S. Bakas, S. Bano, M. J. Cardoso, B. Khanal, B. Landman, X. Li, C. Qin, I. Rekik, N. Rieke, H. Roth, D. Sheet, & D. Xu (Eds.), Distributed, Collaborative, and Federated Learning, and Affordable AI and Healthcare for Resource Diverse Global Health (pp. 3–16). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-18523-6_1
- Zhang, D., Chen, B., Chong, J., & Li, S. (2021). Weakly-Supervised teacher-Student network for liver tumor segmentation from non-enhanced images. *Medical Image Analysis*, 70, 102005. https://doi.org/10.1016/j.media.2021.102005
- Zhang, S., Wang, W.-S., Zhong, B.-Y., & Ni, C.-F. (2022). Subsequent Treatment after Transarterial Chemoembolization Failure/Refractoriness: A Review Based on Published Evidence. *Journal of Clinical and Translational Hepatology*, 10(4), 740–747. https://doi.org/10.14218/JCTH.2021.00336
- Zhao, L., Dong, Q., Luo, C., Wu, Y., Bu, D., Qi, X., Luo, Y., & Zhao, Y. (2021). DeepOmix: A scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis.
 Computational and Structural Biotechnology Journal, 19, 2719–2725. https://doi.org/10.1016/j.csbj.2021.04.067
- Zheng, H., Lin, L., Hu, H., Zhang, Q., Chen, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Tong, R., & Wu, J. (2019). Semi-supervised Segmentation of Liver Using Adversarial Learning with Deep Atlas Prior.
 In D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, & A. Khan (Eds.), Medical Image Computing and Computer Assisted Intervention MICCAI 2019 (Vol. 11769, pp. 148–156). Springer International Publishing. https://doi.org/10.1007/978-3-030-32226-7
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). *UNet++: A Nested U-Net Architecture for Medical Image Segmentation* (arXiv:1807.10165). arXiv. https://doi.org/10.48550/arXiv.1807.10165