Effective Tumor Annotation for Automated Diagnosis of Liver Cancer

YI-HSUAN CHUANG[1,2,3], JA-HWUNG SU[4], TZU-CHIEH LIN[5], HUE-XIN CHENG[4],
PIN-HAO SHEN[4], JIN-PING OU[4], DING-HONG HAN[5], YI-WEN LIAO[6], YEONG-CHYI LEE[7],
YU-FAN CHENG[1,2,3], TZUNG-PEI HONG[4,5], (Senior Member, IEEE),
KATHERINE SHU-MIN LI[5], YI LU[1,2,3], AND CHIH-CHI WANG[8,9]

[1]Liver Transplantation Program, Kaohsiung Chang Gung Memorial Hospital, Niaosung District, Kaohsiung 833401, Taiwan
[2]Department of Diagnostic Radiology, Kaohsiung Chang Gung Memorial Hospital, Niaosung District, Kaohsiung 833401, Taiwan
[3]College of Medicine, Chang Gung University, Taoyuan City 33302, Taiwan
[4]Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811726, Taiwan
[5]Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804201, Taiwan
[6]Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, Kaohsiung 824004, Taiwan
[7]Department of Information Management, Cheng Shiu University, Kaohsiung 833301, Taiwan
[8]Liver Transplantation Center, Kaohsiung Chang Gung Memorial Hospital, Niaosung District, Kaohsiung 833401, Taiwan
[9]Department of Surgery, Kaohsiung Chang Gung Memorial Hospital, Niaosung District, Kaohsiung 833401, Taiwan
CORRESPONDING AUTHORS: J.-H. SU (bb0820@ms22.hinet.net) AND Y.-F. CHENG (prof.chengyufan@gmail.com)

ABSTRACT In recent years, visual cancer information retrieval using Artificial Intelligence has been shown to be effective in diagnosis and treatment. Especially for a modern liver-cancer diagnosis system, the automated tumor annotation plays a crucial role. So-called tumor annotation refers to tagging the tumor in Biomedical images by computer vision technologies such as Deep Learning. After annotation, the tumor information such as tumor location, tumor size and tumor characteristics can be output into a clinical report. To this end, this paper proposes an effective approach that includes tumor segmentation, tumor location, tumor measuring, and tumor recognition to achieve high-quality tumor annotation, thereby assisting radiologists in efficiently making accurate diagnosis reports. For tumor segmentation, a Multi-Residual Attention Unet is proposed to alleviate problems of vanishing gradient and information diversity. For tumor location, an effective Multi-SeResUnet is proposed to partition the liver into 8 couinaud segments. Based on the partitioned segments, the tumor is located accurately. For tumor recognition, an effective multi-labeling classifier is used to recognize the tumor characteristics by the visual tumor features. For tumor measuring, a regression model is proposed to measure the tumor size. To reveal the effectiveness of individual methods, each method was evaluated on real datasets. The experimental results reveal that the proposed methods are more promising than the state-of-the-art methods in tumor segmentation, tumor measuring, tumor localization and tumor recognition. Specifically, the average tumor size error and the annotation accuracy are 0.432 cm and 91.6%, respectively, which suggest potential for reducing radiologists' workload. In summary, this paper proposes an effective tumor annotation for an automated diagnosis support system.

INDEX TERMS Liver cancer diagnosis, liver tumor segmentation, liver tumor measuring, liver tumor location, liver tumor recognition.

Clinical and Translational Impact Statement—The proposed methods have been evaluated and shown to significantly improve the efficiency and accuracy of liver tumor annotation, reducing the time required for radiologists to complete reports on tumor segmentation, liver partition, tumor measuring and tumor recognition. By integrating into existing clinical decision support systems, it has the potential to reduce diagnostic errors and treatment delays, thereby improving patient outcomes and clinical workflow.
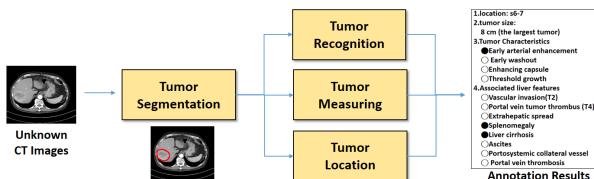
FIGURE 1. Scenario of proposed tumor annotation system.

I. INTRODUCTION

WITH advances of Biomedical Information Retrieval (BIR), significant successes have been achieved over the years in diagnosis, treatment and prognosis. This enables a large increase of medical support systems, particularly for liver cancer [1]. For these needs, a number of previous works were devoted to CT (Computed Tomography) image recognition. Based on image recognition, the practical system for diagnosing CT liver-tumors can segment the tumors automatically. Then, the radiologist manually annotates the tumor information into a diagnosis report, such as tumor location, tumor size and tumor characteristics. This inefficient procedure motivates us to propose an advanced solution to assist the radiologists in efficiently making an effective diagnosis report [2], [3] for CT liver tumors.

Figure 1 shows the scenario of proposed tumor annotation system. In this diagram, the tumor segmentation serves as the foundational component for the next three elements, namely tumor recognition, tumor measuring and tumor location. Then, the annotation results will be generated according to these three elements. Note that, in this paper, the difference between tumor segmentation and tumor recognition lies in that the tumor segmentation refers to marking the tumor region, while the tumor recognition involves analyzing the tumor by visual features of the marked region.

In fact, implementing such annotation systems will encounter challenges in simultaneously performing effective tumor segmentation, tumor recognition, tumor measuring and tumor location to achieve more reliable diagnoses. In summary, the motivations and benefits of this paper can be folded as follows.

- In terms of tumor segmentation, traditional methods often suffer from information loss and uniform feature weighting. To address these issues, this paper proposes a Multi-Residual Attention U-Net (MRAU-Net), which integrates continuous residual connections along with channel-wise and pixel-wise attention mechanisms to enhance feature representation and improve segmentation performance.
- In terms of tumor location, the liver must first be partitioned into 8 regions (called liver partition in this paper), after which the tumor is localized within these specific regions. Actually, most related works have focused on liver partition using a single, highly complex model. In contrast, we propose multi-classifier framework, called Multi-SeResUnet, to decrease the

computational complexity and enable accurate tumor localization within specific liver regions.
- For tumor measuring and tumor labeling, they have received limited attention in the literatures, as they are often considered less critical than tumor segmentation and liver partition. However, these two tasks are essential for the development of a fully functional medical support system in practice. To tackle this concern, the tumor size in this paper is calculated by the proposed regression model, which is based on the tumor segmentation. Additionally, tumor characteristics are effectively recognized through the use of a multi-labeling classifier.
- Regarding the clinical translational contribution, currently, this tumor annotation system is being tested by the radiologists at Chang Gung Memorial Hospital, Kaohsiung Branch, Taiwan. Preliminary feedbacks have been positive, as the system demonstrates the potential to significantly reduce the clinical workload associated with liver tumor assessment reporting. By streamlining the reporting process, it will help improve diagnostic efficiency and consistency in clinical practice. The average reporting time per case is expected to be reduced by at least 30%, thereby enabling radiologists to concentrate more on complex and high-priority diagnostic tasks. Additionally, it has the potential to reduce misjudgments and treatment delays. In the near future, the system is planned to be deployed across other branches of Chang Gung Memorial Hospital in Taiwan, further demonstrating its clinical applicability.

II. LITERATURE REVIEW

The related works can be categorized into 3 classes, namely liver tumor segmentation, liver tumor recognition and liver partition. In the followings, these related works will be reviewed by classes.

A. LIVER TUMOR SEGMENTATION

Liver tumor segmentation has been a hot topic in the field of Bioinformatics due to rapid advances on Deep Learning. A number of forerunners proposed effective methods on this topic. In visual-based Deep Learning, Convolutional Neural Network (CNN) [5] is the backbone optimizing the features by a set of convolutions and gradient descents. Attributed to CNN's excellent feature extraction capabilities, it has achieved successes in segmentation tasks. A fully convolutional neural network (FCN) [6], [7], [8] one of CNNs is efficient for object segmentation due to the end-to-end training. Next, Unet [9] is the most popular network for Biomedical image segmentation. Basically, the naïve Unet considers the horizontal information loss by concatenating the encoder and decoder. However, there remains another problem: vertical vanishing gradient. Here, so-called vertical vanishing gradient indicates the information loss while encoding and decoding. To attack this problem, ResUnet [10] added residual structures to the convolutions, reducing the

vertical vanishing gradients. Also, Unet++ [11], [12] was extended to merge Unets of different depths as a global Unet. These sub-Unets share an encoder and each has its own decoder. In order to train them simultaneously, their losses are calculated together, and then the gradients are backpropagated to each other. In addition to residual-based Unet, attention gates [13], [14] were embedded into Unet, making the network focus on interested objects. The other recent type of Unets is DefED-Net [15] incorporating deformable convolutions and Ladder-ASPP (Atrous-Spatial-Pyramid-Pooling) into the Unet for scale granularity concerns.

B. LIVER TUMOR RECOGNITION

In this paper, tumor recognition is based on the segmented tumor. According to this concept, how to effectively recognize the tumor by feature analysis is the major aim. For example, Jiang et al. [16] detected microvascular invasion in hepatocellular carcinoma by comparing XGBoost and deep learning based on Radiomics features. Zhu et al. [17], and Shan et al. [18] adopted Radiomics features to predict early recurrence of hepatocellular carcinoma (HCC) after treatments. For microvascular invasion and outcome, Sun et al. [19] and Xu et al. [20] used Deep Learning and multivariate logistic regression to achieve the predictions for MRI and CT images, respectively. On the contrary, without microvascular invasion, Wei et al. [21] attempted to predict early recurrence of solitary HCC by MRI images. For early recurrence in HCC after hepatectomy, Zhao et al. [22] offered comparisons of clinical and imaging data for early recurrence and non-early recurrence. Later, Yan et al. [23] extracted the significant features from MRI images by VGG19 and compared 5 classifiers then. Tian et al. [24] provided a diagnosis support system tagging tumors with shape, contour, and intensity by a semi-supervised attention mechanism. Srinivasan et al. [25] aimed at the X-ray reporting by TCNet (Tag Classification Net) with Multi-head attention mechanism. Also, Endo et al. [26] projected visual and textual features on a space and retrieved the most similar report by similarity comparisons. Alksas et al. [27] utilized Random Forest to classify the liver tumor grades of LR1, LR2, LR3, LR4, and LR5.

C. LIVER PARTITION

Typically, the liver can be divided into 8 functionally independent regions called Couinaud Segments by the portal vein, as shown in Figure 2. In general, at least 2 Couinaud segments will be reserved after a liver hepatectomy due to concerns of segment remnant functions. Hence, preoperative assessment for hepatic functional reserve is an important issue. Hernandez-Alejandro et al. [28] proposed a proposal for staged liver resections by bridging couinaud segments to portal vein ligation. Both Lebre et al. [29] and Oliveira et al. [30] performed liver partition by stages. First, the liver was segmented. Then, the portal veins were identified, and finally,
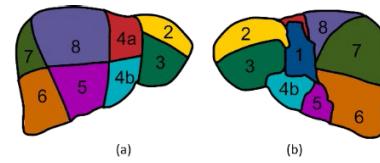

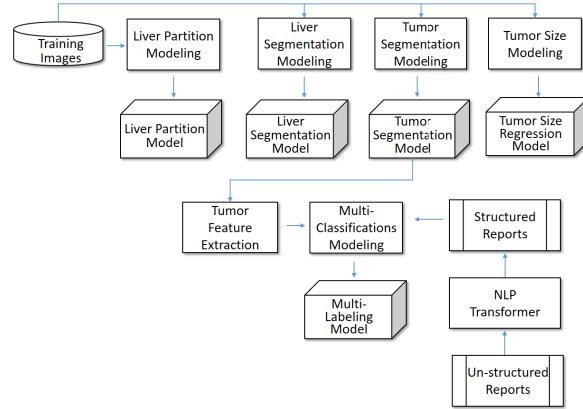**FIGURE 2.** Couinaud segments in (a): anterior view and (b): posterior view.

FIGURE 3. Offline training of proposed system framework.

the liver was partitioned into 8 segments by the portal veins. Han et al. [31] used a single 3D Unet to directly segment liver segments from MRI images. In contrast to the reference [31], this paper proposes multi-classifiers from viewpoints of different segments, and further conducts the better results for both 2D and 3D images.

III. PROPOSED FRAMEWORK for LIVER-TUMOR ANNOTATION

A. FRAMEWORK

To achieve high quality of tumor annotation, Figures 3 and 4 show the framework of proposed system comprising offline training and online annotation.

- Offline training (illustrated in Figure 3): This stage trains models for four main components. In terms of tumor segmentation, it needs a liver segmentation model and a tumor segmentation model. In terms of tumor recognition, it needs a multi-labeling model combining information of tumor features and structured reports labels. In terms of tumor measuring, it needs a tumor size regression model. In terms of tumor location, it needs a liver partition model.
- Online annotation (illustrated in Figure 4): On the basis of the models trained, first, the liver and tumor are sequentially segmented from the unknown image. That is, the whole liver is split from the image and then the tumor is segmented from the liver. Next, the liver is partitioned into 8 regions and the tumor will be located accordingly. Simultaneously, the tumor features are extracted for multi-labeling the tumor. Also, the tumor size is predicted by the regression model.
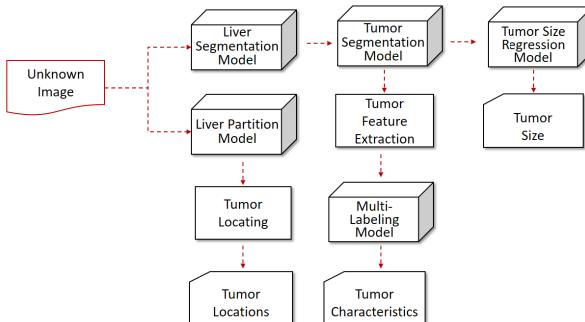
FIGURE 4. Online annotation of proposed system framework.

FIGURE 5. Architecture of proposed multi-residual attention Unet.

Finally, the tumor size, tumor location and tumor characteristics can be incorporated as annotations into the diagnosis report.

B. OFFLINE TRAINING

1) GENERATING GROUND TRUTHS FROM UNSTRUCTURED REPORTS

This is an essential operation which can be viewed as knowledge discovery. Since existing diagnostic reports are expressed in natural language, it is necessary to transform these unstructured sentences into structured labels. For this need, the unstructured reports are transformed into the labels as ground truths. Because this operation is not the primary focus of this paper, we just utilized the previous work [2] to perform it, achieving an accuracy of approximately 90.99%.

2) TUMOR SEGMENTATION

In our proposed method, the tumor segmentation consists of two steps, namely liver segmentation and tumor segmentation. Thus, it needs two segmentation models for livers and tumors, respectively. For modeling the liver segmentation, it is implemented by Unet [9] with data augmentations of flipping. Finally, the dice of the liver segmentation reached 0.951. For modeling the tumor segmentation, Figure 5 shows the architecture of the proposed Multi-Residual Attention Unet consisting of an encoder and a decoder. The encoder is composed of 4 multi-residual blocks, while the decoder is composed of 4 de-convolution blocks. In the followings, the main ideas of multi-residual blocks, pixel-wise attention and loss functions will be presented in detail.

C. MULTI-RESIDUAL BLOCK

In this component, the major intent is to cope with the issues of *information loss* and *uniform feature weighting* by leveraging two key concepts: multiple residuals and attention mechanisms. In terms of *information loss*, information in traditional deep networks could be lost through multiple convolutional layers. To address this issue, we employ multiple continuous residual connections to preserve and enhance information during the convolution process. In terms of
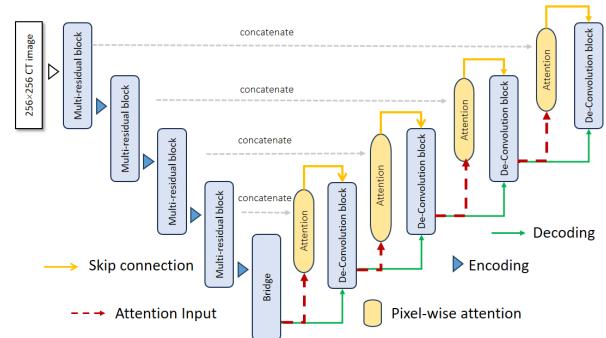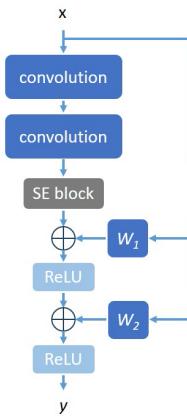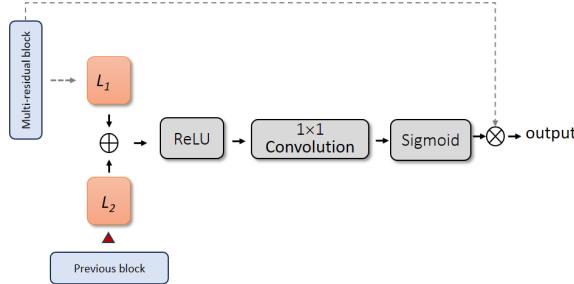

FIGURE 6. Architecture of proposed multi-residual block.

uniform feature weighting, traditional models process input features uniformly, making it difficult to emphasize the more important parts. To aim at this issue, we employ the attention mechanism [33] to dynamically learn the importance of each input position, thereby improving model performance. In overall, Multi-residual Block takes a dual mechanism of dynamic weighting and information enhancement, enabling it to effectively handle the challenges of tumor segmentation.

Figure 6 shows the architecture of the proposed multi-residual block, consisting an input x, an output y, two convolutions, a Squeeze-and-Excitation block (called SE block) [34], two activation functions ReLU and two weight functions W_1 and W_2. Here, W_1 and W_2 are the linear mapping functions using the 1×1 convolutions to achieve the linear transformations. For the squeeze-and-excitation block, it can actually be regarded as a channel-wise attention mechanism consisting of two main modules. The first module Squeeze is set for aggregating the features. The formula for aggregations is defined as:

$$z_c = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} u_c(i,j)' \tag{1}$$

where z denotes a scalar for the attention coefficient, c denotes the channel number, h and w denote height and width, respectively, and u denotes the aggregation function

FIGURE 7. Workflow of pixel-wise attention.

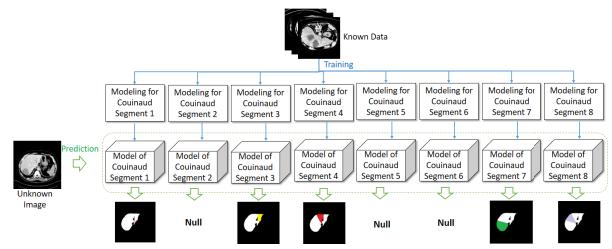**FIGURE 8.** Architecture of proposed Multi-SeResUnet.

for pixels of i and j in a channel. The major idea of this formula is to apply Global Average Pooling to each channel (i.e., averaging over the entire spatial dimensions $h \times w$) to obtain a single value per channel. This effectively compresses each channel's feature map into one scalar, forming a vector of size $c \times 1$. In this step, all pixels in the spatial dimension are averaged, indicating that the pixels are not independent but they collectively contribute to the overall representation of the channel. After aggregations, the second module Excitation is set to extract attention features from the aggregated information, generating the attention coefficient. The formula for excitation is defined as:

$$s = \sigma(F_2 \delta(F_1 z)), \qquad (2)$$

where z stands for the feature vector derived by Equation (1), F_1 and F_2 are the linear transformation functions used in the fully connected layer, δ and σ stand for the ReLU activation function and the sigmoid activation function, respectively. Therefore, the attention coefficient derived by Equation (2) can be employed to enhance the model segmentation quality. In summary, through our proposed multi-sequential weighted identity mapping, the feature information can be enhanced. That is, the potentially interested features can be focused by the squeeze-and-excitation block. This is the main uniqueness of the proposed multi-residual block.

D. PIXEL-WISE ATTENTION

In order to make the segmentation focus on the tumor more, the pixel-wise attention mechanism [13] (also called Spatial-wise attention) is embedded into the decoder. Figure 7 illustrates the pixel-wise attention mechanism. In this framework, L1 and L2 are two linear transformation functions implemented using 3×3 convolutions, applied respectively to the two inputs of the pixel-wise attention module. Specifically, L1 applies a 3×3 convolution to the input feature map from the multi-residual block to obtain the first transformed feature map, while L2 applies a 3×3 convolution to the input feature map from the previous block to obtain the second transformed feature map. Next, the addition of two feature maps is passed through a RelU function, followed by a 1×1 convolution and a sigmoid function. Finally, these weights are element-wise multiplied with the original

multi-residual block. As a result, the whole segmentation network is guided to focus more effectively on the target object.

E. LOSS COEFFICIENT

The final idea to present in Multi-Residual Attention Unet is the used loss coefficients, including Dice Coefficient, Binary Cross-entropy Coefficient and Matthews Correlation Coefficient. The dice coefficient [35] is a popular metric widely used in evaluating image segmentation quality, indicating the intersection ratio between the prediction and the ground truth. The Binary Cross-Entropy BCE is adopted to measure the object classification quality in this paper, which is defined as:

$$BCE = -((1 - y) * log(1 - p) + y \times log(p)), \qquad (3)$$

where y and p denote the true binary and the probability of the predicted positive class, respectively. The Matthews correlation coefficient MCC [36] is a binary measurement for classifications, measuring the relationships among False Positives (FP), True Positives (TP), False Negatives (FN), and True Negatives (TN) in a confusion matrix. It is defined as:

$$MCC = MCC_n / \sqrt{MCC_d}, \qquad (4)$$

where $MCC_d = (TP + FN) * (TP + FP) * (TN + FN) * (TN + FP)$ and $MCC_n = (TP * TN) - (FP * FN)$.

1) TUMOR LOCATION

In the online stage, as shown in Figure 8, it needs to partition the liver into 8 couinaud segments before locating the tumor. Therefore, in the training, the liver partition model has to be constructed. In this paper, innovative segmentation methods for 2D and 3D images are proposed where the core creativity is to reduce the segmentation complexity. Traditional segmentation methods attempt to partition the objects by a single segmentor but the effectiveness is limited. In our approach, 8 segmentors are conducted individually. That is, the computer vision will be limited in the smaller space and therefore the effectiveness will be increased obviously.

For 2D images, the proposed method, called Multi-SeResUnet, utilizes SeResUnet [4] to construct eight recognition models. The detailed framework can be found in the related work [37]. For a 3D image [31], [38], which contains 3D kernels used for convolutions and max-pooling operations. Except for the final convolution, the kernel with size

of 3×3 is used for the other convolutions. Also, the batch normalization (BN) along with the LeakyReLU activation function is used for these convolutions. The transposed convolutional layers and max pooling operations use a kernel size 2×2. The final convolutional layer generates the output, while the related kernel size is 1×1 and the related activation function is sigmoid. Whether the image is 2D or 3D, the major uniqueness for liver partition in this paper is to split a single partition model into 8 individual partition models to reduce the partition complexity. That is, for each of couinaud segments, a referred model is trained.

2) TUMOR MEASURING

Tumor size is an important evidence for stage diagnosis, surgical treatment and prognosis estimation. In general, the tumor size ranges from 0.5cm to more than 20cm, with the majority being less than 5cm, followed by 5-10cm. Few tumors will grow to more than 10cm. Currently, treatments for liver tumors below 3cm are relatively selective, and the prognosis is relatively good. This is why the aimed annotation contains tumor size. Because the tumor is segmented in a visual format of pixels, it needs a mechanism to transform the pixels into centimeters. Hence, the size here is approximated by a linear regression function, which is defined as:

$$size = b + a \times length, \tag{5}$$

where b stands for the intercept, a stands for the regression coefficient and $length$ stands for the longest Euclidean distance between the tumor border points. In this paper, b and a are set as -0.2011 and 0.0748, respectively.

3) TUMOR RECOGNITION

In this work, tumor recognition refers to the multi-labeling algorithm learning from visual features on recognizing the tumor characteristics. For this purpose, the *Radiomics* features [39], [40] are selected as the visual features. In this stage, the *Radiomics* features are extracted first. Next, a set of nearly optimal features are filtered by the feature selector. Finally, the multi-labeling classifier is performed to recognize the potential characteristics. In order to pursue a high-quality classification, sets of feature selectors and multi-labeling classifiers were evaluated in the experiments. Specifically, feature selectors such as Information Gain and Pearson Correlation Coefficients, and classifiers including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Random Forest (RF), Neural Network (NN), AdaBoost, and XGBoost were assessed.

F. ONLINE ANNOTATION

On the basis of the above training models, the automated annotation starts with a submission of an unknown CT image. First, the liver is segmented by Unet. Second, the tumor is initially split from the liver by Multi-Residual Attention Unet. Because the initial tumor segmentation is not very solid, it needs an additional post-process. Here, the post-process comprises two further operations, namely thresholding and morphological operations. Through thresholding, the most potential pixels will be determined from the initial tumor segmentation, while morphological operations such as closing and opening make the tumor segmentation more solid. In particular, morphological operations are performed in three dimensions in a CT image cube, with respect to xy, xz and yz planes. Note that, the threshold is set as 2 in this paper. Simultaneously, the liver is partitioned by Multi-SeResUnet. Third, the *Radiomics* features are extracted from the tumor and the corresponding characteristics are subsequently classified. Fourth, the tumor is measured by Equation (5). Fifth, the tumor is located by checking every couinaud segment with Equation (6) which is defined as:

$$location(s_k)$$
$$= \begin{cases} true, & if \dfrac{predicted \cap ground\_truth}{ground\_truth} > 0.25 \\ false, & otherwise \end{cases}, \tag{6}$$

where s_k indicates the kth couinaud segment label, *predicted* indicates the partitioned couinaud segment, and *ground_truth* indicates the actual couinaud segment. That is, if $location(s_k)$ is true, the resulting locations will include s_k. Finally, the tumor characteristics, tumor size and tumor location are merged into the final diagnosis report.

IV. EXPERIMENTS
A. EXPERIMENTAL SETTINGS
1) EXPERIMENTAL DATA

The overall experiments were conducted on an aspect of evaluating the proposed components, including tumor segmentation, tumor location, tumor measuring and tumor recognition. Therefore, three datasets were set for evaluating these components. In terms of tumor segmentation, Lits data [41] were set as the experimental data, including 131 patients. In terms of tumor location, it needs to partition the liver in advance. For this need, MedSeg data [42] were set for liver partition, including 50 patients. In terms of tumor measuring and tumor recognition, the experimental data were gathered from the hospital in Taiwan, including 70 patients. The number of labels for tumor recognition is 11. In the related evaluations, 5-fold cross validations were set for tumor segmentation, liver partition and tumor recognition, while tumor measuring was experimented by a 2-fold cross validation. Note that, the proposed evaluation shows the robustness and generalization of proposed system because the evaluation was conducted using models constructed from three different datasets. The LiTS dataset was used to build the liver segmentation model. Based on its results, the Chang Gung dataset was employed for tumor segmentation model construction. Additionally, the MedSeg dataset was used to develop the liver partition model. Finally, tumors in the Chang Gung dataset were located and annotated by integrating the results of tumor segmentation and liver partition.

TABLE 1. Settings of experimental network parameters.

Network	batch size	resize	epoch	optimizer	learning rate
Multi-Residual Attention Unet	16	256×256	50	Adam	0.0001
SeResUnet	8	256×256	50	Adam	0.0001
3D Unet	2	64×128×128	80	Adam	0.0001

TABLE 2. Terminologies of involved methods compared in the experiments.

Aspect	Compared Method	Terminology
Tumor Segmentation	Unet [9]	Unet
	Attention Unet [13]	AUnet
	Unet with Backbone ResNet	RUnet
	SE-Unet [34]	SeUnet
	Unet++ [11]	Unet++
	Unet with residuals and Backbone InceptionNet v3	IUnet
	MRAUnet	proposed
Feature Selection	Pearson Correlation Coefficient	PCC
	Information Gain	IG
Tumor Recognition	Support Vector Machine	SVM
	Linear Discriminant Analysis	LDA
	Random Forest	RF
	Neural Network	NN
	AdaBoost	AdB
	XGBoost	XGB

2) EXPERIMENTAL NETWORK PARAMETERS

The involved Unets in this paper include Multi-Residual Attention Unet, SeResUnet and 3D Unet. Table 1 shows the detailed settings of experimental network parameters.
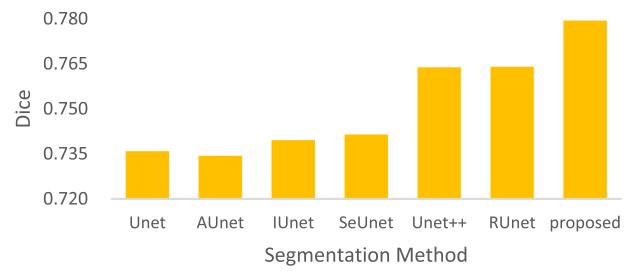
3) EXPERIMENTAL MEASURES

The measures used in the evaluations can be divided into 3 categories for segmentation/partition, measuring and recognition/location, respectively. For segmentation and partition, dice is the main metric revealing the Harmonic mean between True Positive Rate and Positive Predictive Value. For measuring, MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) are two related metrics. For recognition and location, Accuracy, Precision, Recall and F1-score widely used in machine learning evaluations are four main metrics.

B. EVALUATIONS RESULTS

1) EXPERIMENTAL RESULTS for TUMOR SEGMENTATION

The experimental results for tumor segmentation are analyzed in two main aspects: approximation of settings and overall performance comparison. For approximation of settings, the better settings for #residuals, attention mechanisms, loss functions and morphological operations are elicited before overall comparisons. Next, the proposed method was compared with a set of state-of-the-art methods.

Table 2 lists the compared methods with terminologies. Because there have been so many recent works

FIGURE 9. Dices of compared tumor segmentation methods.

accompanying Unets with InceptionNet [43] and RseNet [32], we do not list the specific cited references here.

Figure 9 depicts the dices of compared tumor segmentation methods, delivering several observations. First, the results can briefly be clustered into 4 groups, namely {Unet, AUnet}, {IUnet, SeUnet}, {Unet++, ResUnet} and {proposed}, where the performance rank is {proposed}, {Unet++, ResUnet}, {IUnet, SeUnet} and {Unet, AUnet}. In the results, the potential reason for why Unet and AUnet perform closely could be that, the bridge attention does not deal with small tumors very well. Second, pure residual-based Unets such as Unet++ and ResUnet perform better than the others except the proposed method. Third, the proposed MRAUnet achieves the best dice. From above results, the technical contribution can be echoed with a concept that, the proposed MRAUnet with multiple sequential residuals and attention mechanisms brings out the better segmentation quality than the compared methods.

2) EXPERIMENTAL RESULTS for TUMOR LOCATION

The tumor location was implemented by two phases, including liver partition and tumor location. In terms of liver partition, we compared the proposed multi-models with the related single models [4], [31]. The liver-partition dices in 2D and 3D spaces are shown in Figure 10. From this result, we can obtain two observations that, first, 3D models outperform 2D models. Second, the proposed multi-models work better than the compared single models. Here, this result can be viewed as one of our technical contributions. Following this, we will look into the tumor location. On the basis of liver-partition results, the evaluation results for tumor location are shown in Figure 11. The precision, recall, and accuracy were found to be satisfactory by the doctors after reviewing the results.

3) EXPERIMENTAL RESULTS FOR TUMOR MEASURING

Tumor measuring is proposed to predict the tumor size by Equitation (5). Table 3 shows the tumor size distribution. Most tumor sizes range from 1 to 5 centimeters. Table 4 reveals the evaluation results in terms of MAE and RMSE, which can be explained by some concepts. First, the averaged MAE is much smaller than size of 5cm, around an error ratio of 0.0866cm. Second, in contrast to MAE, RMSE is slightly higher than MAE. This is because RMSE is more sensitive
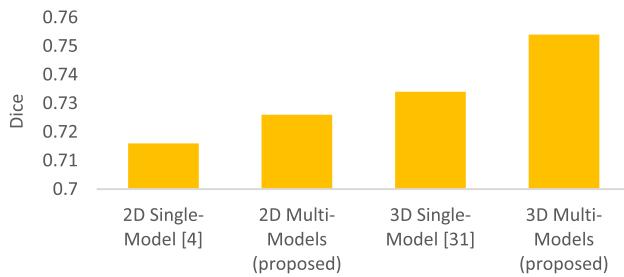
FIGURE 10. Dices of liver partition using multi-models and single models under 2D and 3D spaces.
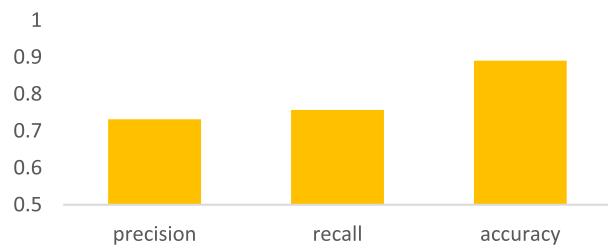

FIGURE 11. Precision, recall and accuracy of tumor location using 2D multi-models.

TABLE 3. Tumor size distribution.

	<1cm	1-5cm	6-10cm	>10cm
#tumors	1	61	7	1

TABLE 4. Experimental results for tumor measuring.

Metric	Average
MAE	0.432996
RMSE	0.638784

TABLE 5. Statistical analysis for tumor measuring.

Metric	value
PCC	0.951458
p value (<0.01)	0.000

to the larger prediction losses. Overall, the results show the proposed method is stable. For robustness concern, it is further analyzed statistically. Table 5 depicts that the proposed tumor measuring is robust while the PCC and p value achieve 95.15% and 0.000, respectively. For PCC, it indicates the prediction results are highly related to the ground truths. For p value, it is 0.000 smaller than the threshold of 0.01, revealing there exists an obvious relevance between prediction results and ground truths. These evaluation results evidence the robustness of the proposed tumor measuring.

4) EXPERIMENTAL RESULTS for TUMOR RECOGNITION

Figure 12 shows the comparative results for tumor recognition. In this evaluation, 1316 visual features are extracted
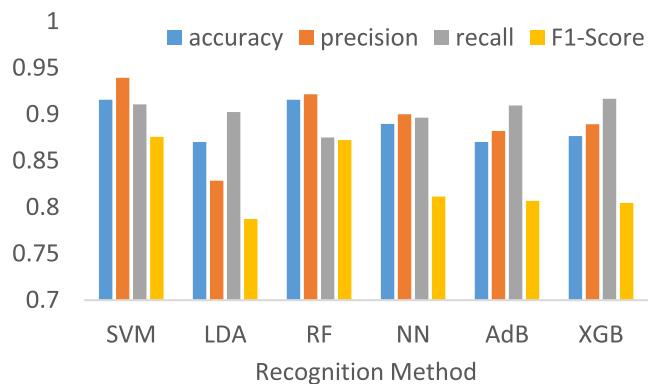

FIGURE 12. Effectiveness comparisons for tumor recognition models.

from a tumor. Here, several observations are summarized. First, in overall, SVM is the best, and RF is the second top whether using IG or PCC. However, the performances of SVM and RF are pretty close. This could be explained by the fact that, SVM uses a nonlinear kernel function, while RF inherently exhibits nonlinear characteristics. Both models can effectively handle the nonlinear data in this experiment. On average, SVM is more stable than RF due to its higher precision and recall. Second, LDA performs the worst among the compared methods. The potential reason could be that, LDA is based on idealized linearity assumptions of linearity and Gaussian distribution. Because the experimental data in this study do not meet these assumptions, LDA naturally performs worse than more flexible, nonlinear, or ensemble-based models. In contrast, RF and XGB are ensemble learning models capable of constructing complex and flexible decision boundaries. NN can learn deep features and handle high-dimensional and nonlinear data. AdB adapts by focusing more on the misclassified samples. Third, although the precisions and recalls of NN, AdB and XGB are not low, their F1-scores are lower than those of SVM and RF. This might be caused by the fact that, F1-score is calculated by first taking the harmonic mean of precision and recall for each individual record, and then averaging the results. It penalizes cases where there is a large discrepancy between precision and recall. Therefore, if the model's performance across records is diverse, it can significantly lower the overall F1-score. In other words, the precisions and recalls of NN, AdB and XGB show significant instability, with substantial variability in individual records. In summary, SVM and RF are two notable classifiers recommended in this paper due to their consistently high and closely matched performances. Note that, for feature selections, IG is slightly better than PCC on the whole. To reveal the model interpretability, Table 6 presents the top 10 Radiomics features used in the recognition models, highlighting that shape and first-order statistical features are the most influential. Shape features describe the geometric properties of tumors, such as volume, surface area, aspect ratio, roundness, and boundary irregularity. In contrast, first-order statistics are derived from the distribution of

TABLE 6. Top 10 radiomics features.

Rank	Feature
1	original_shape_Elongation
2	original_shape_MajorAxisLength
3	original_shape_MeshVolume
4	original_shape_MinorAxisLength
5	original_shape_Sphericity
6	original_shape_SurfaceArea
7	original_shape_SurfaceVolumeRatio
8	original_firstorder_Energy
9	original_firstorder_Entropy
10	original_firstorder_Kurtosis

gray-level intensities within the image, without accounting for the spatial relationships between pixels.

V. CONCLUSION

Recent Bioinformatics techniques have been effectively applied to liver tumor segmentation over the past few years. Yet, there remain some issues in these methods. First, their effectiveness could be improved. Second, no study has proposed a comprehensive tumor annotation system. To address these issues, in this paper, we propose a set of effective methods to achieve an automated tumor annotation system for diagnosis. First, the annotation format is defined by the doctors based on practical requirements. Second, unstructured reports are transformed into structured labels for training. Third, tumor segmentation is performed using the proposed MRAUnet, which incorporates Unet, SE-Bolck, pixel-wise attention, channel-wise attention and post-processing. Fourth, the tumor is located within the specific Couinaud segments, which are partitioned by the proposed Multi-SeResUnet. Fifth, the tumor is measured by the proposed regression model. Sixth, the tumor characteristics are classified by the approximated multi-labeling method. Finally, the location, size and characteristics of the tumors are viewed as annotations merged into the diagnosis report. To assess the effectiveness of proposed methods, a number of evaluations were conducted on real datasets. The evaluation results reveal that the proposed methods indeed yield satisfactory outcomes for radiologists. Accordingly, this approved annotation system will be integrated into the existing medical infrastructure to assist radiologists in generating effective reports more efficiently and thereby help prevent treatment delays.

In the future, several issues need to be addressed. First, a general system using different datasets is demanded by the doctors because the support hospital has different branches with different CT capturing devices. Federated learning and collaborative learning are the next intents in our future work. Second, the effectiveness is related to the GPU performances, making the model un-afford big data. Hence, the data augmentation was not used in the tumor segmentation and liver partition. In the future, we will make further training models with data augmentations by the greater GPU. Third, for liver partition, 3D Unet is a potential model but its performance was limited in GPU performances. Therefore, the better version will be researched in the future by using the more powerful GPU. Fourth, the location, size and characteristics are the aimed attributes in this paper. In addition to these attributes, future work will incorporate more tumor information, such as stage, invasion, recurrence, and others, into the annotation system to enhance its robustness.

AUTHOR CONTRIBUTIONS

Conceptualization: Yi-Hsuan Chuang, Ja-Hwung Su, Yi-Wen Liao, Yeong-Chyi Lee, and Yu-Fan Cheng; Methodology: Yi-Hsuan Chuang, Ja-Hwung Su, Tzung-Pei Hong, and Katherine Shu-Min Li; Validation: Tzu-Chieh Lin, Hue-Xin Cheng, Pin-Hao Shen, Jin-Ping Ou, and Ding-Hong Han; Formal Analysis: Ja-Hwung Su, Yi-Wen Liao, and Yeong-Chyi Lee; Investigation: Yu-Fan Cheng, Tzung-Pei Hong, Katherine Shu-Min Li, and Chih-Chi Wang; Data Curation: Yi-Hsuan Chuang, Yu-Fan Cheng, Yi Lu, and Chih-Chi Wang; Resources: Yi-Hsuan Chuang, Yu-Fan Cheng, Yi Lu, and Chih-Chi Wang; Writing—Original Draft Preparation: Ja-Hwung Su; and Writing—Review and Editing: Ja-Hwung Su.

REFERENCES

[1] A. Bakrania, N. Joshi, X. Zhao, G. Zheng, and M. Bhat, "Artificial intelligence in liver cancers: Decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases," *Pharmacological Res.*, vol. 189, Mar. 2023, Art. no. 106706.

[2] Y.-H. Chuang et al., "Effective natural language processing and interpretable machine learning for structuring CT liver-tumor reports," *IEEE Access*, vol. 10, pp. 116273–116286, 2022.

[3] P. D. Poullos et al., "Structured reporting of multiphasic CT for hepatocellular carcinoma: Effect on staging and suitability for transplant," *Amer. J. Roentgenology*, vol. 210, no. 4, pp. 766–774, Apr. 2018.

[4] *ResUNetPlusPlus*. Accessed: Dec. 2022. [Online]. Available: https://github.com/DebeshJha/ResUNetPlusPlus/blob/master/m_resunet.py

[5] Y.-C. Tung et al., "High-performance scaphoid fracture recognition via effectiveness assessment of artificial neural networks," *Appl. Sci.*, vol. 11, no. 18, p. 8485, Sep. 2021.

[6] P. Ferdinand Christ et al., "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," 2017, *arXiv:1702.05970*.

[7] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, and H. Meine, "Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing," *Sci. Rep.*, vol. 8, no. 1, p. 15497, Oct. 2018.

[8] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Munich, Germany, Jan. 2015, pp. 234–241.

[10] W. Qi, M. Wei, W. Yang, C. Xu, and C. Ma, "Automatic mapping of landslides by the ResU-net," *Remote Sens.*, vol. 12, no. 15, p. 2487, Aug. 2020.

[11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, vol. 11045, 2018, pp. 3–11.

[12] Z. Zhou et al., "UNet++: Redesigning skip connections to exploit multi-scale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Dec. 2019.

[13] O. Oktay et al., "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[14] E. Thomas et al., "Multi-res-attention UNet: A CNN model for the segmentation of focal cortical dysplasia lesions from magnetic resonance images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1724–1734, May 2021.

[15] T. Lei, R. Wang, Y. Zhang, Y. Wan, C. Liu, and A. K. Nandi, "DefED-net: Deformable encoder–decoder network for liver and liver tumor segmentation," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 1, pp. 68–78, Jan. 2022.

[16] Y.-Q. Jiang et al., "Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning," *J. Cancer Res. Clin. Oncol.*, vol. 147, no. 3, pp. 821–833, Mar. 2021.

[17] Y. Zhu et al., "Radiomics-based model for predicting early recurrence of intrahepatic mass-forming cholangiocarcinoma after curative tumor resection," *Sci. Rep.*, vol. 11, no. 1, p. 18347, Sep. 2021.

[18] Q.-Y. Shan et al., "CT-based peritumoral radiomics signatures to predict early recurrence in hepatocellular carcinoma after curative tumor resection or ablation," *Cancer Imag.*, vol. 19, no. 1, pp. 1–11, Feb. 2019.

[19] B.-Y. Sun et al., "Deep-learning-based analysis of preoperative MRI predicts microvascular invasion and outcome in hepatocellular carcinoma," *World J. Surgical Oncol.*, vol. 20, no. 1, pp. 1–13, Jun. 2022.

[20] X. Xu et al., "Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma," *J. Hepatology*, vol. 70, no. 6, pp. 1133–1144, Jun. 2019.

[21] Y. Wei, W. Pei, Y. Qin, D. Su, and H. Liao, "Preoperative MR imaging for predicting early recurrence of solitary hepatocellular carcinoma without microvascular invasion," *Eur. J. Radiol.*, vol. 138, May 2021, Art. no. 109663.

[22] Q.-Y. Zhao, S.-S. Liu, and M.-X. Fan, "Prediction of early recurrence of hepatocellular carcinoma after resection based on Gd-EOB-DTPA enhanced magnetic resonance imaging: A preliminary study," *J. Gastrointestinal Oncol.*, vol. 13, no. 2, pp. 792–801, Apr. 2022.

[23] M. Yan et al., "Deep learning nomogram based on Gd-EOB-DTPA MRI for predicting early recurrence in hepatocellular carcinoma after hepatectomy," *Eur. Radiol.*, vol. 33, no. 7, pp. 4949–4961, Feb. 2023.

[24] J. Tian, C. Li, Z. Shi, and F. Xu, "A diagnostic report generator from CT volumes on liver tumor with semi-supervised attention mechanism," in *Proc. MICCAI*, vol. 11071, Sep. 2018, pp. 702–710.

[25] P. Srinivasan, D. Thapar, A. Bhavsar, and A. Nigam, "Hierarchical X-ray report generation via pathology tags and multi head attention," in *Proc. Asian Conf. Comput. Vis.*, Jan. 2021, pp. 600–616.

[26] M. Endo, R. Krishnan, V. Krishna, A. Y. Ng, and P. Rajpurkar, "Retrieval-based chest X-ray report generation using a pre-trained contrastive language-image model," in *Proc. PLMR*, vol. 158, 2021, pp. 209–219.

[27] A. Alksas et al., "A novel computer-aided diagnostic system for accurate detection and grading of liver tumors," *Sci. Rep.*, vol. 11, no. 1, p. 13148, Jun. 2021.

[28] R. Hernandez-Alejandro, L. I. Ruffolo, R. Alikhanov, B. Björnsson, O. J. M. Torres, and A. Serrablo, "Associating liver partition and portal vein ligation for staged hepatectomy (ALPPS) procedure for colorectal liver metastasis," *Int. J. Surgery*, vol. 82, pp. 103–108, Oct. 2020.

[29] M.-A. Lebre et al., "Automatic segmentation methods for liver and hepatic vessels from CT and MRI volumes, applied to the Couinaud scheme," *Comput. Biol. Med.*, vol. 110, pp. 42–51, Jul. 2019.

[30] D. A. Oliveira, R. Q. Feitosa, and M. M. Correia, "Segmentation of liver, its vessels and lesions from CT images for surgical planning," *Biomed. Eng. OnLine*, vol. 10, no. 1, p. 30, Apr. 2011.

[31] X. Han et al., "Automated segmentation of liver segment on portal venous phase MR images using a 3D convolutional neural network," *Insights into Imag.*, vol. 13, no. 1, p. 26, Feb. 2022.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 770–778.

[33] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10076–10085.

[34] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2018, pp. 7132–7141.

[35] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.

[36] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica Biophysica Acta (BBA)-Protein Struct.*, vol. 405, no. 2, pp. 442–451, Oct. 1975.

[37] Y.-C. Tung et al., "Effective invasiveness recognition of imbalanced data by semi-automated segmentations of lung nodules," *Biomedicines*, vol. 11, no. 11, p. 2938, Oct. 2023.

[38] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense, volumetric segmentation from sparse annotation," in *Proc. MICCAI*, vol. 9901, Athens, Greece, Oct. 2016, pp. 424–432.

[39] J. J. M. van Griethuysen et al., "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017.

[40] J. D. Shur et al., "Radiomics in oncology: A practical guide," *RadioGraphics*, vol. 41, no. 6, pp. 1717–1732, Oct. 2021.

[41] P. Bilic et al., "The liver tumor segmentation benchmark (LiTS)," *Med. Image Anal.*, vol. 84, Nov. 2022, Art. no. 102680.

[42] H. B. Jenssen and T. Sakinis, "MedSeg liver segments dataset, figshare," Tech. Rep., 2022, doi: 10.6084/m9.figshare.13643252.v2.

[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.