

## RESEARCH ARTICLE

# Clinical application of mask region-based convolutional neural network for the automatic detection and segmentation of abnormal liver density based on hepatocellular carcinoma computed tomography datasets

Ching-Juei Yang<sup>1,2,3</sup>, Chien-Kuo Wang<sup>4</sup>, Yu-Hua Dean Fang<sup>5</sup>, Jing-Yao Wang<sup>1,2</sup>, Fong-Chin Su<sup>1,6</sup>, Hong-Ming Tsai<sup>4,7</sup>, Yih-Jyh Lin<sup>7,8</sup>, Hung-Wen Tsai<sup>7,9</sup>, Lee-Ren Yeh<sup>10\*</sup>

**1** Department of Biomedical Engineering, National Cheng-Kung University, Tainan, Taiwan, **2** Division of Medical Radiology, E-Da Cancer Hospital, I-Shou University, Kaohsiung, Taiwan, **3** E-Da Cancer Hospital, Hepatobiliary and Pancreatic Cancer Collaborative Oncology Group, Kaohsiung, Taiwan, **4** Department of Medical Imaging, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan, **5** Department of Radiology, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States of America, **6** Department of Biomedical Engineering & Medical Device Innovation Center, National Cheng Kung University, Tainan, Taiwan, **7** National Cheng Kung University Hospital, Liver Cancer Collaborative Oncology Group, Tainan, Taiwan, **8** Division of General and Transplantation Surgery, Department of Surgery, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan, **9** Department of Pathology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan, **10** Department of Radiology, E-Da Hospital, I-Shou University, Kaohsiung, Taiwan

\* [yehlr3323@gmail.com](mailto:yehlr3323@gmail.com)



## OPEN ACCESS

**Citation:** Yang C-J, Wang C-K, Fang Y-HD, Wang J-Y, Su F-C, Tsai H-M, et al. (2021) Clinical application of mask region-based convolutional neural network for the automatic detection and segmentation of abnormal liver density based on hepatocellular carcinoma computed tomography datasets. PLoS ONE 16(8): e0255605. <https://doi.org/10.1371/journal.pone.0255605>

**Editor:** Khanh N.Q. Le, Taipei Medical University, TAIWAN

**Received:** January 18, 2021

**Accepted:** July 19, 2021

**Published:** August 10, 2021

**Copyright:** © 2021 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The laboratory protocols were publicized under the DOI ([dx.doi.org/10.17504/protocols.io.buusunwwe](https://doi.org/10.17504/protocols.io.buusunwwe)), with minimal anonymized data from public medical images extracted from The Cancer Imaging Archive.

**Funding:** The financial support of this study is under NCKU E-Da Research Grant No. NCKUEDA10614. The funders had no role in study

## Abstract

The aim of the study was to use a previously proposed mask region-based convolutional neural network (Mask R-CNN) for automatic abnormal liver density detection and segmentation based on hepatocellular carcinoma (HCC) computed tomography (CT) datasets from a radiological perspective. Training and testing datasets were acquired retrospectively from two hospitals of Taiwan. The training dataset contained 10,130 images of liver tumor densities of 11,258 regions of interest (ROIs). The positive testing dataset contained 1,833 images of liver tumor densities with 1,874 ROIs, and negative testing data comprised 20,283 images without abnormal densities in liver parenchyma. The Mask R-CNN was used to generate a medical model, and areas under the curve, true positive rates, false positive rates, and Dice coefficients were evaluated. For abnormal liver CT density detection, in each image, we identified the mean area under the curve, true positive rate, and false positive rate, which were 0.9490, 91.99%, and 13.68%, respectively. For segmentation ability, the highest mean Dice coefficient obtained was 0.8041. This study trained a Mask R-CNN on various HCC images to construct a medical model that serves as an auxiliary tool for alerting radiologists to abnormal CT density in liver scans; this model can simultaneously detect liver lesions and perform automatic instance segmentation.

design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Liver cancer was the seventh most commonly diagnosed cancer (4.7%) and the leading cause of cancer death (8.2%) globally in 2018, with approximately 841,080 new cases and 781,631 deaths that year [1]. Among liver cancers, HCC accounts for >80% of primary liver tumors and is estimated to be the fourth most common cause of death from cancer worldwide [2]. HCC rates are particularly high in eastern and southeastern Asia and Africa [2, 3]. In recent decades, HCC has attracted attention in the United States and Western Europe due to its major risk factors including nonalcoholic steatohepatitis, alcoholism, and hepatitis C [4]. It frequently occurs in middle-aged men [5], and it is a major burden in many countries [6].

The surveillance of liver lesions depends on imaging features. Aggressive periodic imaging surveillance of patients with high-risk factors and liver cirrhosis (including nodule sizes of <1 cm) is practiced widely [7, 8]. The early stages of HCC as defined by the Barcelona Clinic Liver Cancer system involve a single tumor or three nodules of <3 cm [8]. In addition, the essential criteria of HCC defined by the American Joint Committee on Cancer staging system are one or multiple tumors and the largest having a diameter exceeding 2 cm or 5 cm.

Patients with early-stage HCC can be treated with curative resection or liver transplantation [8, 9]. Transarterial chemoembolization and radiofrequency ablation are treatment options for intermediate-stage HCC [8–10]. For advanced-stage HCC, systemic therapy with sorafenib remains the best option [9].

## Pathophysiology and imaging features of HCC

The liver has two major vascular inflow systems: arterial and portal systems. When HCC develops, the vascular supply system shifts to the arterial system. As a result, the typical features of dynamic imaging of HCC are early arterial enhancement and portal venous- or delayed-phase washout. The typical and atypical enhancement patterns of HCC on computed tomography (CT) were 56.4% and 43.6%, respectively [11]. The tumor growth pathology is a spectrum rather than a clear cutoff characteristic. Moreover, pathological classifications are made on the basis of variable types [12, 13]. HCCs of different sizes may manifest in the different phases of CT images as blurred margins, homogeneous enhancement, heterogeneous enhancement, poor enhancement, central scars or necrosis, or infiltrative type [14, 15]. Thus, the detection and segmentation of liver lesions are challenging.

## Human versus computer-aided detection of liver lesions

Abnormal density identification in each CT image is the crucial first step for radiologists followed by differential diagnosis in evaluating liver disease. The traditional standard imaging diagnosis for HCC consists of a semantic description of lesions on the basis of pathophysiology by radiologists trained in multiphasic dynamic imaging studies. The mean values of sensitivity for the combination of arterial and portal venous-phase imaging, arterial and delayed-phase imaging, and a combination of all three phases were 86.8%, 90.3%, and 93.8%, respectively [16]. The pooled per-lesion sensitivity and positive predictive value of contrast-enhanced CT were 73.6% and 85.8%, respectively [17]. The sensitivity for diagnosing HCC <2 cm by using contrast-enhanced CT was 52.9% [18]. The detection rates of HCCs <1 cm were between 17.4% and 20.6% [19], and small regions with abnormal density were frequently missed or not evident on CT at the initial diagnosis by humans.

Imaging feature identification and HCC analysis in dynamic imaging studies are complex [20, 21], with various sizes and morphologies observed in practical clinical situations. In the past 5 years, imaging feature analysis has rapidly developed [22, 23], and digital imaging

features can be computed from regions of interest (ROIs). It is a new field but a promising technology for liver lesion diagnosis [24].

Computer-aided detection (CAD) involves the identification of specific digital imaging features of target lesions; this assists clinicians in reducing observational oversights [25, 26]. Recently, convolutional neural networks (CNNs) have become mainstream for digital imaging analysis [27, 28]. A CNN is an advanced filter that transforms low-level imaging features into abstract high-level features. Some studies have used early CAD for liver tumor detection [29] and segmentation [30] and differentiation of liver masses on dynamic contrast medium-enhanced CT [31].

These CAD studies have revealed a segmentation quality for liver tumors similar to human performance (mean Dice coefficients of 0.69 vs. 0.72) but inferior to detection performance (recall: 63% vs. 92%) in the liver tumor segmentation test [32]. According to our limited knowledge, the best volume Dice score for liver tumor segmentation is 0.70 based on the Liver Tumor Segmentation Benchmark [33] in 2019. Until July 27, 2020, the optimal Dice value per patient was 0.7990 for three-dimensional (3D) volume segmentation of liver lesions on contrast-enhanced CT scans obtained from the Liver Tumor Segmentation Challenge (LiTs) with automatic segmentation algorithms (<https://competitions.codalab.org/competitions/17094#results>).

### Semantic segmentation using convolutional neural network

Many deep learning (DL) architectures for liver segmentation perform well on different datasets (e.g., LiTs, 3D Image Reconstruction for Comparison of Algorithm Database) or their own datasets with different window ranges.

The semantic segmentation architecture U-Net was first proposed in 2015 [34], with the 3D U-Net variant developed in 2016 [35]. Christ et al. [36] used a two-step modified form of U-Net [34] to perform liver lesion segmentation and reported a Dice value greater than 0.94. Xiao developed a deep convolutional neural network [37] combining the features of U-Net [34] and ResNet [38] and achieved an average Dice value of 0.67. The H-DenseUNet [39] demonstrated a global Dice value of 0.965 for liver segmentation and 0.824 for liver lesion segmentation using a hybrid feature layer. AHCNet [40] demonstrated global Dice values of 0.959 for liver segmentation and 0.734 for tumor segmentation using a U-Net variant, attention mechanism, and skip connections. U<sup>n</sup>-Net, with an n-fold network architecture [41], resulted in a Dice value of 0.9638 for liver segmentation and 0.7369 for tumor segmentation. E<sup>2</sup>NET [42], a two-dimensional (2D) edge-enhanced network with a multiscale feature extraction backbone and a decoder similar to that of U-Net [34], yielded global Dice values of 0.968 and 0.829 for liver and tumor segmentation, respectively, on the LiTs dataset.

U-Net [34] is a semantic segmentation architecture that performs segmentation of classes without instance segmentation, which distinguishes objects of the same class. The original U-Net was tested on cell tracking. For more complex medical images, such as those of liver tumors, the careful design of liver-oriented CNN layers and the reconciliation of image features from different scales to U-Net are crucial.

### Mask region-based convolutional neural network

Girshick et al. proposed a region-based CNN (R-CNN) in 2013–2014 that consists of four basic components: image input, region extraction proposals, computation of CNN features, and classification of regions [43]. The R-CNN can converge the diverse morphology, size, color, density, and texture characteristics as well as the visualization of domain knowledge into a digital informatics system for further transferring applications and engagements to other

experts. On the basis of this structure, fast [44], faster [45], and Mask R-CNN by He et al. [46] were developed and exhibited high efficiency for simultaneous object detection and instant segmentation regarding the relationship between object imaging features and background. For Mask R-CNN, ROI Align was proposed to solve the problem of misalignment. Thus, the object size is no longer a problem if the target item provides sufficient pixel information. As a result, the inference ability of Mask R-CNN for images with complicated pixel information of HCC is unknown.

In this study, we used datasets from two hospitals in Taiwan rather than the open-source dataset. The assessment was conducted from radiologists' viewpoint in addition to engineering considerations.

To implement Mask R-CNN, this study used Google TensorFlow, the pretrained Microsoft Common Objects in Context model, and CT datasets to develop an inference model for abnormal liver density detection and segmentation simultaneously.

In this study, we trained the model based on the CT images of HCC obtained from one institute and further tested this model on the dataset obtained from the second institute.

## Materials and methods

### Collection of clinical training and testing datasets

This retrospective study was approved by the institutional review boards of E-Da Hospital and National Cheng Kung University Hospital (EMRP-106-041/NCKUEDA10614; A-ER-107-023), which also waived the requirement of informed consent. HCC CT data from 2012 to 2017 were retrospectively obtained from the cancer databanks of two hospitals in Taiwan (Table 1 and Fig 1); all data were fully anonymized.

The CT scanning parameters were 120 kVp with automatic exposure control depending on the CT machine. The time settings of the high-pressure syringe dynamic CT scans were 15–20 sec for the early arterial phase, 45 sec for the late arterial phase, 80 sec for the portal venous phase, and 4 min for the equilibrium delay phase, with an injection speed of 2.0–3.5 mL/sec. For nondynamic CT, the contrast medium was manually injected into the venous system of

**Table 1. Properties of training and testing data.**

Characteristic	Training set (training: validation = 9:1 randomly distributed)	Positive test set	Negative test set
Data source	Hospital 1	Hospital 2	Hospital 2
Number of patients	394	214	123
Male/ Female	303/ 91	154/ 60	90/ 33
Age (mean, SD)	(64.9, 11.1)	(62.6, 11.2)	(59.9, 9.7)
Dynamic/ Non-dynamic CT	376/ 18	208/ 6	117/ 6
Number of liver CT images	10130	1833	20283
Number of tumor ROIs	11258	1874	-
Number of hyperdense ROIs	3078	334	-
Number of hypodense ROIs	6544	1313	-
Number of heterogeneous density ROIs	1636	227	-
LD of GT tumor ROI (mean, SD) (cm)	(6.37, 3.97)	(4.75, 2.16)	-
Area of GT tumor ROI (mean, SD) (cm <sup>2</sup> )	(32.08, 37.29)	(17.03, 15.83)	-

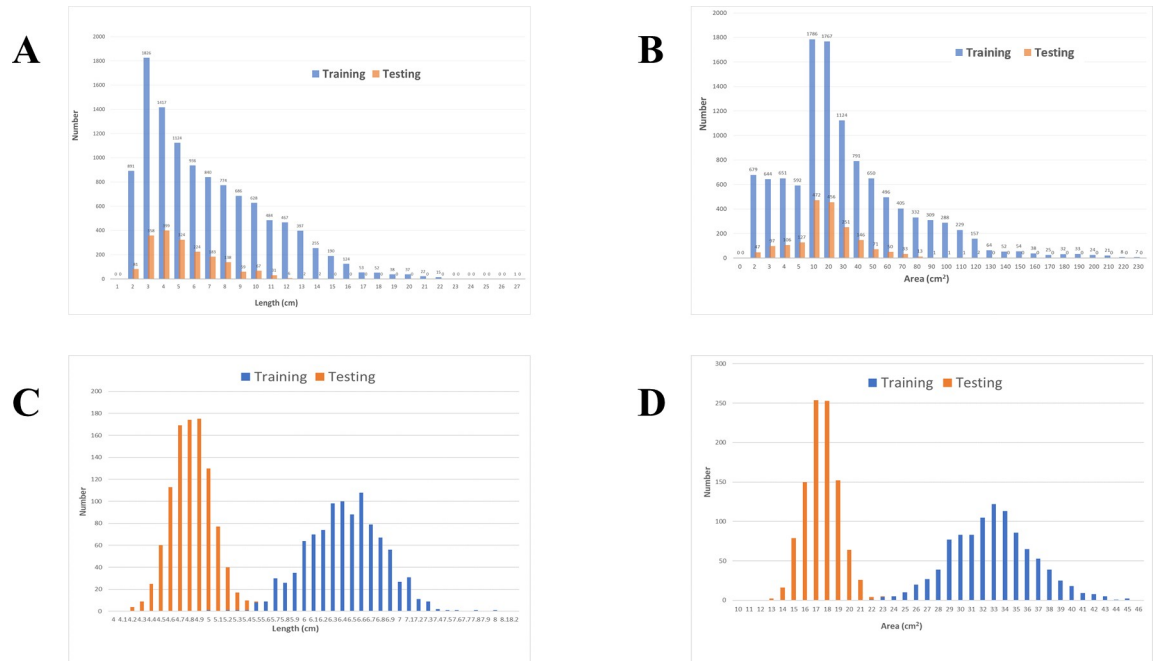
LD: longest diameter.

ROI: region of interest.

GT: ground truth.

SD: standard deviation.

<https://doi.org/10.1371/journal.pone.0255605.t001>



**Fig 1.** (A) Distribution of LDs of Abnormal Liver Density. (B) Distribution of areas of abnormal liver density. (C) Distribution of LDs after bootstrap resampling with a mean value of 100 samples repeated 1,000 times ( $r = 0.9980, p < 0.0001$ ). (D) Distribution of areas after bootstrap resampling with a mean value of 100 samples repeated 1,000 times ( $r = 0.9988, p < 0.0001$ ).

<https://doi.org/10.1371/journal.pone.0255605.g001>

the upper extremity by a qualified nurse, and an immediate scan was conducted after the contrast medium was pushed out for the post-contrast-medium phase.

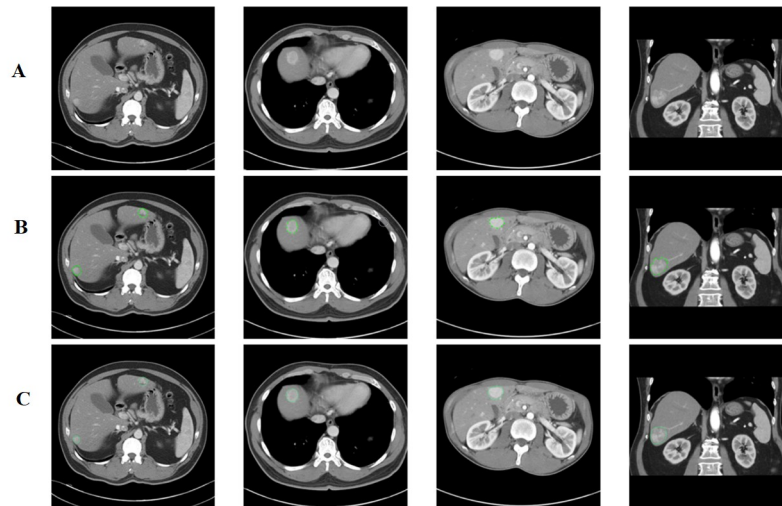
All images were acquired using 5-mm-thick transverse planes spaced at 5 mm in addition to coronal or sagittal reconstructed planes. All training images comprised only the field of view of the liver, and images of patients with severe fatty liver [47], hemochromatosis, and severe distortion of the liver parenchyma after transarterial chemoembolization, radiofrequency ablation, or partial resection were removed.

The images were converted from DICOM to 8-bit PNG format with a ratio of  $512 \times 512$  pixels corresponding to  $35 \times 35 \text{ cm}^2$  in 2D images under a window width of 450 HU and window level of 45 HU.

### Determination of ROI area and longest diameter

Because of varying ROI morphologies, the study used the fitEllipse and contourArea functions of the Open Source Computer Vision Library to automatically measure longest diameters (LDs) and areas through contour delineation of the lesions.

Finally, we acquired 10,130 images containing liver tumor densities of 11,258 ROIs (LD:  $6.37 \pm 3.97 \text{ cm}$ ; area:  $32.08 \pm 37.29 \text{ cm}^2$ ) obtained from 394 CT images (dynamic CT to nondynamic CT ratio was 376:18) of 394 patients (age:  $64.92 \pm 11.14$  years; male to female ratio was 303:91) as the training dataset from the first hospital (Table 1). In clinical viewpoints, the arbitrary 2D imaging features of one tumor may be similar with those of another tumor. For the maximal acquisition of abnormal imaging features, each 2D image was treated independently and randomly distributed into the training and validation datasets at a ratio of 9:1. Moreover, in the second hospital, 1,833 images containing liver tumors with 1,874 ROIs (LD:  $4.75 \pm 2.16 \text{ cm}$ ; area:  $17.03 \pm 15.83 \text{ cm}^2$ ) obtained from 214 CT images (dynamic CT to nondynamic CT



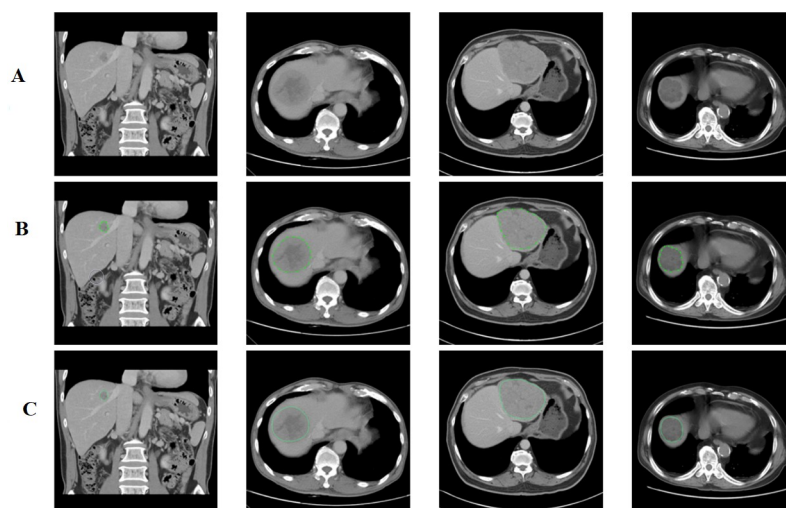
**Fig 2.** (A) Abnormal Hyperdensities Without Labeling ROI. (B) Hyperdensities with hand-marked ROIs. (C) Hyperdensities with AI-predicted ROIs.

<https://doi.org/10.1371/journal.pone.0255605.g002>

ratio was 208:6) of 214 patients (age:  $62.59 \pm 11.24$  years; male to female ratio was 154:60) were regarded as positive testing data, and 20,283 CT images obtained from 123 CT images (dynamic CT to nondynamic CT ratio was 117:6) of 123 patients (age:  $59.96 \pm 9.76$  years; male to female ratio was 90:33) with liver fields of view without abnormal densities in the liver parenchyma were regarded as negative testing data (Table 1). Fig 1A and 1B show the LD and area distribution obtained from the two hospitals. Fig 1C and 1D show the correlation values of LD and area distribution.

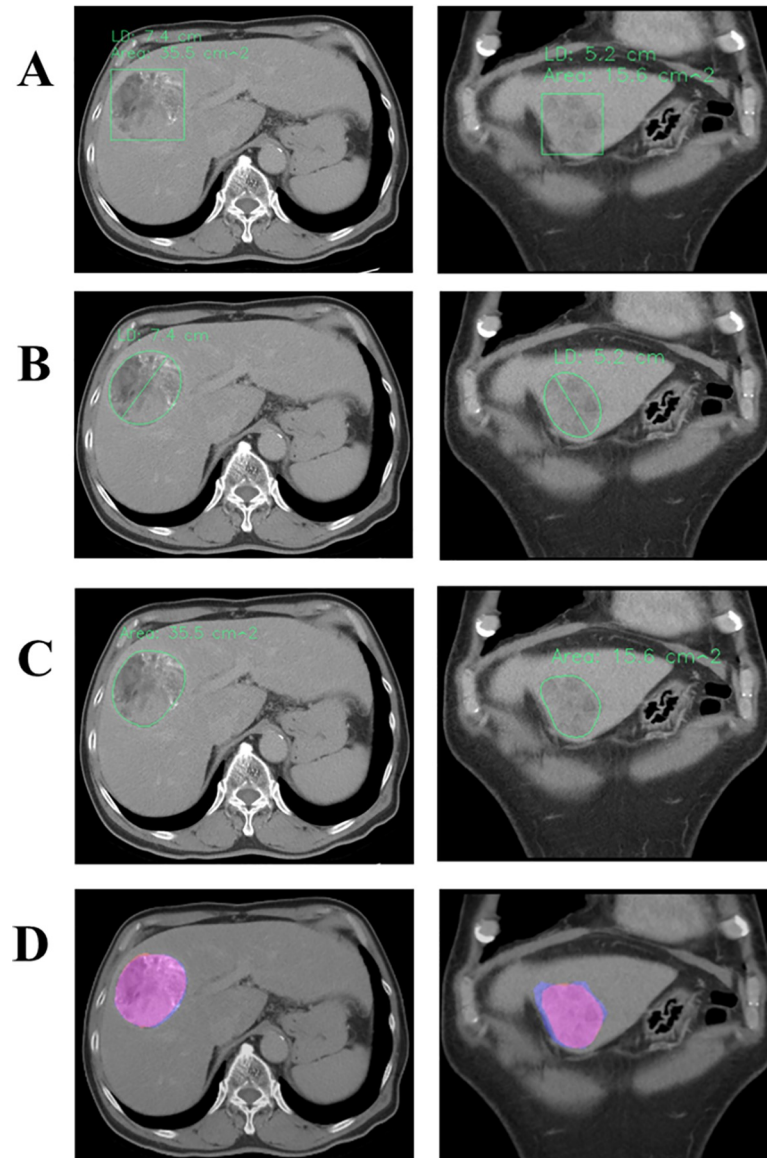
### Positive ground truth ROI

The positive ground truth (GT) tumors were verified against the cancer data bank document and original reporting document and further confirmed by a second radiologist. Abnormal liver densities were identified based on hyperdensity (Fig 2A) in the arterial phase, hypodensity (Fig 3A) in



**Fig 3.** (A) Abnormal Hypodensities Without Labeling ROI. (B) Hypodensities with hand-marked ROIs. (C) Hypodensities with AI-predicted ROIs.

<https://doi.org/10.1371/journal.pone.0255605.g003>



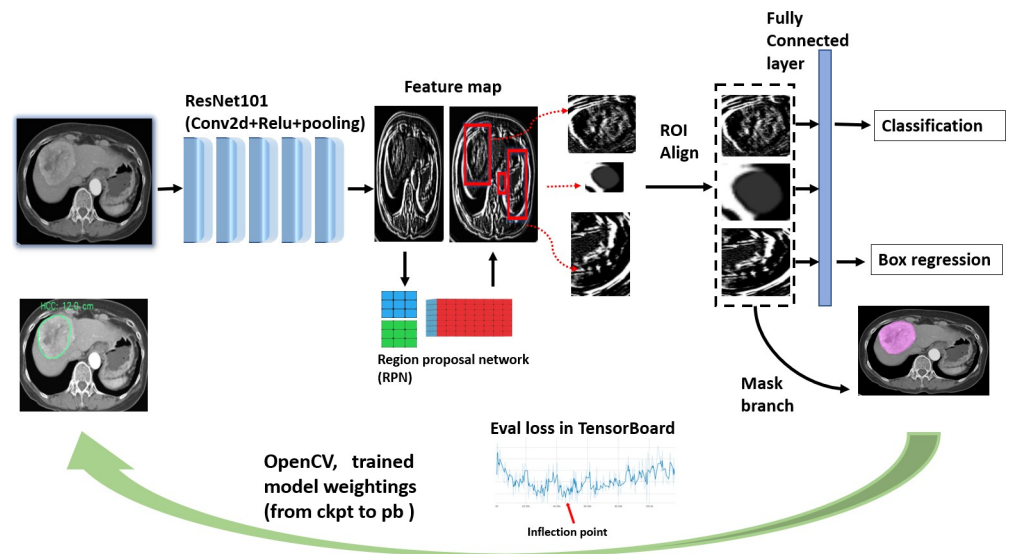
**Fig 4.** (A) AI-predicted Rectangle ROIs of Heterogeneous Densities with Concurrent Outputs of LD and Area. (B) AI-labeled LDs (7.4 and 5.2 cm) of abnormal densities through ellipse fitting. (C) AI-labeled areas of abnormal densities through contour delineation (area = 35.5 and 15.6 cm<sup>2</sup>). (D) Overlapping hand-drawn GT ROIs (light blue) and Mask R-CNN-predicted ROIs (pink) of abnormal densities (Dice coefficients = 0.9667 and 0.8922).

<https://doi.org/10.1371/journal.pone.0255605.g004>

the noncontrast phase or washout patterns during either the portal venous or delayed phase, or heterogeneous density (Fig 4A–4C) [11].

### Negative GT dataset

The prevalence rate of incidental hepatobiliary findings based on chest CT screening was 6.1% [48]. The prevalence of small hepatic lesions discovered through CT in patients with cancer was 12.7% [49]. In addition to determining the positive GT, this study included the negative GT dataset without abnormal density in the liver parenchyma to mimic the larger number of normal images encountered in the daily practice of radiologists.



**Fig 5. Diagram of mask R-CNN modeling and inference building.**

<https://doi.org/10.1371/journal.pone.0255605.g005>

## Implementation of artificial intelligence inference

The experimental protocols of Protocol\_mask\_RCNN\_CT\_tensorflow were deposited in protocols.io with the following digital object identifier: [dx.doi.org/10.17504/protocols.io.buusnww](https://doi.org/10.17504/protocols.io.buusnww). We used public medical images from The Cancer Genome Atlas Liver Hepatocellular Carcinoma data collection (<https://wiki.cancerimagingarchive.net/display/Public/TCGA-LIHC>) [50] of The Cancer Imaging Archive [51] for preliminary program testing.

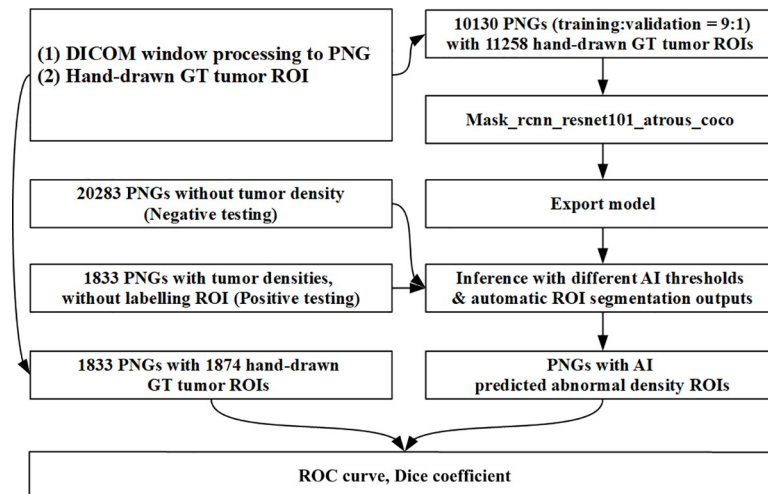
Fig 5 depicts the end-to-end process of this inference model. Its backbone is ResNet101, which filters input images using 64 randomly generated filters in the first stage and constructs 101 layers. The residual learning of ResNet employs shortcut connections from one layer to subsequent layers that facilitate training of the deeper network [38]. We assumed that this backbone could capture more diverse features from HCC images. The design of the region proposal network incorporates the ROIs into a feature map following the RoI Align to avoid misalignment of image size and pixel information. In the final stage, the Mask R-CNN, employs bounding box regression and object classification using the neural networks. Thus, with sufficient pixel information, the two-stage architecture of the Mask R-CNN can perform accurate object detection and instance segmentation regardless of the size of object.

After building the training dataset, we used mask\_R-CNN\_resnet101\_atrous\_coco as the pretrained model. Microsoft's Common Objects in Context is a large-scale image dataset designed for object detection and segmentation [52]. The training dataset comprises 10,130 images with a mixture of hypodensity, hyperdensity, and heterogeneous density (Table 1). The next step was the optimization of digital imaging features by using the Google TensorFlow open-source platform for machine learning on a NVIDIA 2080 Ti GPU. All training images were input in batches (1,000–3,000 training images per batch) to the DL model and inspected using the TensorBoard web interface. The training endpoint was set as the inflection point of the validation data in the training dataset to avoid overfitting errors (Fig 5).

Finally, the optimal model parameters were frozen and bridged to the Open Source Computer Vision Library to build the inference for further application, that is, testing with the dataset from the second hospital (Fig 6).

For this artificial intelligence (AI) inference, we designed two automatic detection output interfaces for use depending on user preferences: rectangle inference–predicted ROI (Fig 4A)





**Fig 6. Process of mask R-CNN modeling and inference building.**

<https://doi.org/10.1371/journal.pone.0255605.g006>

and contour inference–predicted ROI (Fig 4C). Simultaneously, LD, area, and overlapping maps were revealed under contour delineation results (Fig 4B–4D).

### Statistical analysis

Because the training dataset was larger than the testing dataset in the right-skewed distribution (Fig 1A and 1B), a bootstrapping method [53] was employed (1,000 random samples of 100); an independent-samples *t* test [54] was used to compare the distribution of mean values by randomly resampling LD and area values from the training and testing datasets (Fig 1C and 1D).

The testing section was composed of two sets of mask thresholds (THs; 0.25, 0.5, and 0.75) and confidence THs (0.99, 0.95, 0.9, 0.85, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, and 0.2) of 51 combinations for determining variations in detection sensitivity.

In medical imaging detection, the most essential aspect of lesion detection is alerts of any suspicious abnormal density in the image, regardless of the lesion size. Thus, true positive was defined as any pixel intersection of the positive GT ROI and inference-predicted contour ROI larger than zero pixels. Furthermore, the Dice coefficient was determined through comparing the GT ROI and inference-predicted contour ROI for segmentation ability.

$$\text{Dice coefficient} = 2|\text{GT.ROI} \cap \text{Predicted.ROI}| / (\text{GT.ROI} + \text{Predicted.ROI}) \quad (1)$$

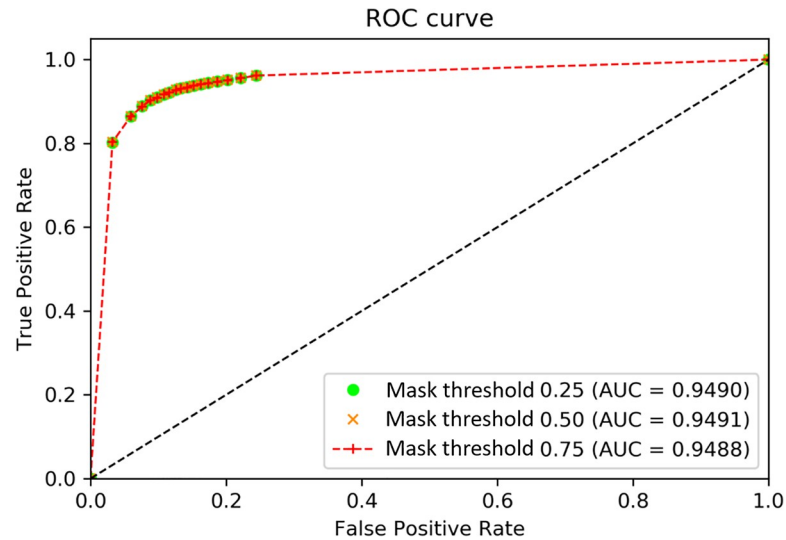
If the contour inference detected any pixel in the group of negative testing images, the result was regarded as a false positive.

The area under the curve (AUC), true positive rate (TPR), false positive rate (FPR), and Dice coefficient were used to evaluate liver lesion detection and segmentation ability.

### Results

The LD correlation and area correlation values between training and testing datasets were 0.9980 and 0.9988, respectively, and the *p* value of both was <0.0001 (Fig 1C and 1D). During training, in TensorFlow, the validation dataset at the inflection point revealed a loss value of 1.181.

Detection evaluation was assumed for 3 and 17 types of mask THs and confidence THs, respectively, to tune different sensitivities of the AI system for lesion detection. The average



**Fig 7. Receiver operating characteristic curve of lesion detection inference for all tumor densities.**

<https://doi.org/10.1371/journal.pone.0255605.g007>

processing time by a 2080 Ti GPU for a single image was 0.51 sec. Because making absolute cutoff margins for lesion and non-lesion areas in a single image is impractical, the pathological scale of tumors cannot be precisely revealed through CT imaging. We designed positive and negative testing datasets for evaluating detection ability. The testing group included 1,833 images with 1,874 lesion ROIs (positive testing dataset) and 20,283 images without abnormal lesion densities (negative testing dataset).

This model obtained AUC values of 0.9490 (95% confidence interval [CI]: 0.9471–0.9508), 0.9491 (95% CI: 0.9472–0.9509), and 0.9488 (95% CI: 0.9469–0.9506) for the three mask THs of 0.25, 0.50, and 0.75, respectively (Fig 7), for a mixture of all three density conditions.

According to Fig 7, the optimal discrimination point is at the upper left corner of the receiver operating characteristic curve (mask TH = 0.50, confidence TH = 0.85). We further analyzed the performance of this inference for hyperdensity, hypodensity, and heterogeneous density under mask TH = 0.5, with different confidence THs.

The optimal AUC value was 0.9581 (95% CI: 0.9561–0.9600) for hypodense lesions, followed by 0.9389 (95% CI: 0.9331–0.9446) for those of heterogeneous density and 0.9197 (95% CI: 0.9140–0.9253) for hyperdense lesions (Fig 8).

The mean AUC, TPR, and FPR were 0.9490, 91.99%, and 13.67%, respectively, for the objective detection of abnormal lesions with 51 TH combinations and a mixture of all three density conditions (Table 2).

For the true positive detection of lesions, automatic CAD segmentation was performed. This inference attained mean Dice coefficients of 0.7962 (standard deviation [SD] = 0.0306), 0.8041 (SD = 0.0310), and 0.7912 (SD = 0.0303) for the three mask THs (Table 2). The segmentation ability achieved a mean Dice coefficient of 0.7971 for 51 TH condition combinations, with the best coefficient being 0.8041 under a mask TH of 0.5 and 0.8355 under a mask TH of 0.5 and confidence TH of 0.2. This inference exhibited acceptable detection and segmentation abilities.

Uneven liver parenchyma (26.17%) accounted for the majority of false positives, followed by the inferior vena cava (15.95%) and other organs (Fig 9). These results provide essential information for the future improvement of this inference.

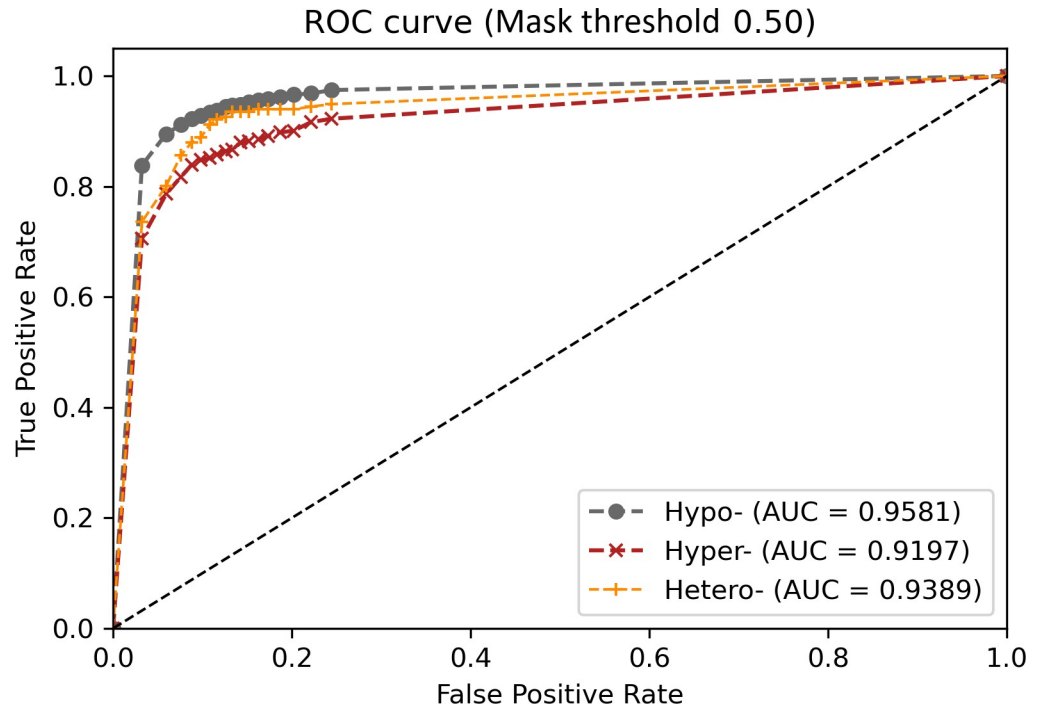


Fig 8. Receiver operating characteristic curve of lesion detection inference for different densities.

<https://doi.org/10.1371/journal.pone.0255605.g008>

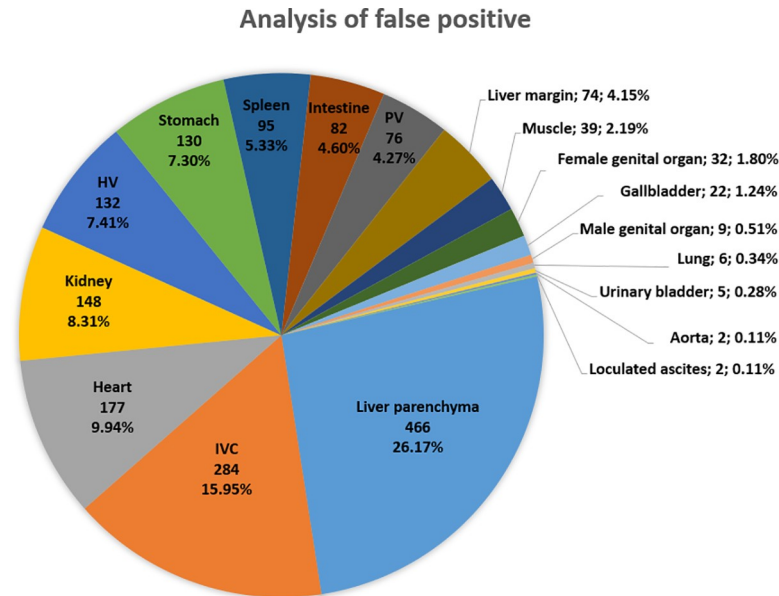
Table 2. TPR (%), FPR (%), and dice coefficients of different confidence THs and mask THs.

Confidence TH	Mask TH = 0.25			Mask TH = 0.5			Mask TH = 0.75		
	FPR (%)	TPR (%)	Dice	FPR (%)	TPR (%)	Dice	FPR (%)	TPR (%)	Dice
0.99	3.24	80.25	0.7032	3.24	80.31	0.7100	3.24	80.31	0.6994
0.95	5.95	86.42	0.7545	5.95	86.47	0.7619	5.95	86.47	0.7499
0.9	7.6	88.87	0.7723	7.6	88.87	0.7796	7.6	88.82	0.7671
0.85	8.79	90.29	0.7838	8.79	90.29	0.7914	8.79	90.23	0.7787
0.8	9.84	90.94	0.7884	9.84	90.94	0.7961	9.84	90.89	0.7834
0.75	10.81	91.71	0.7944	10.81	91.71	0.8021	10.81	91.65	0.7892
0.7	11.64	92.14	0.7975	11.64	92.14	0.8053	11.64	92.09	0.7922
0.65	12.58	92.8	0.8030	12.58	92.8	0.8108	12.58	92.74	0.7975
0.6	13.37	93.13	0.8057	13.37	93.13	0.8135	13.37	93.07	0.8001
0.55	14.27	93.4	0.8076	14.27	93.4	0.8155	14.27	93.34	0.8021
0.5	15.18	93.78	0.8103	15.18	93.78	0.8182	15.18	93.73	0.8050
0.45	16.3	94.16	0.8131	16.3	94.16	0.8211	16.3	94.11	0.8078
0.4	17.38	94.44	0.8144	17.38	94.44	0.8225	17.38	94.38	0.8092
0.35	18.73	94.76	0.8173	18.73	94.76	0.8255	18.73	94.71	0.8123
0.3	20.21	95.14	0.8211	20.21	95.14	0.8294	20.21	95.09	0.8161
0.25	22.15	95.64	0.8236	22.15	95.64	0.8319	22.15	95.58	0.8187
0.2	24.47	96.24	0.8268	24.47	96.24	0.8355	24.47	96.18	0.8225
mean	13.68	92.01	0.7962	13.68	92.01	0.8041	13.68	91.96	0.7912
SD	5.8	3.95	0.0306	5.8	3.93	0.0310	5.8	3.92	0.0303

TH: threshold.

SD: standard deviation.

<https://doi.org/10.1371/journal.pone.0255605.t002>



**Fig 9. Analysis of 1781 false positive images.** (IVC: inferior vena cava; HV: hepatic vein; PV: portal vein).

<https://doi.org/10.1371/journal.pone.0255605.g009>

## Discussion

This study used the previously proposed Mask R-CNN to engineer and design of a method of data collection and model evaluation from a medical viewpoint. This 2D validation study performed training and testing on HCC datasets with different size distributions in an end-to-end manner.

The Mask-R-CNN extracts the feature map and simultaneously performs the RoI Align for objective detection and instance segmentation with variable sizes. It provided AUC values greater than 0.9 for different liver lesion densities for a 2D Dice value of approximately 0.80.

Setting an appropriate window is the first step in the process of image inspection for radiologists and in feature extraction based on the DL model.

The initial window setting of this study was within the abdominal window range for increased ease of transfer learning—the machine learning method whereby a pretrained model is reused as the starting training point for another similar task—for use with the abdominal window ranges of other radiologists.

Although the optimal theoretical TH setting was mask TH = 0.50 and confidence TH = 0.85, the end user can select different TH combinations of varying sensitivities for particular medical situations.

In the analysis of the false positives, we identified uneven liver parenchyma, such as that caused by fatty liver or severe liver cirrhosis, as the primary weakness of this model because we excluded such extreme liver parenchyma images from our training dataset. The majority of the other false positives were related to major vessels or hypervascular organs such as the inferior vena cava or heart. We surmised that these may present similar features to the imaging features of hypervascular tumors under a single DL model. In addition, this 2D model lacks 3D information (i.e., z-axis); thus, its recognition of 3D organ structures is suboptimal. In addition, the model was trained on single class, with limited objects in each image; 3D voxel multiclass training has greater requirements for hardware, graphical processing, and training time than its 2D counterpart. The balance between the available CAD system and the hardware requirements for DL training is a crucial concern in public health care worldwide.

In this model, we did not include other malignant tumors such as cholangiocarcinoma, metastasis, and other benign liver lesions such as liver cysts or hemangiomas for multiple object detection.

For long-term usage and development, this model can be used for purposes other than single object liver lesion AI training through transfer learning, thus avoiding data scarcity dilemmas to achieve the desired training endpoint.

Detailed region segmentation and sophisticated classifications of dynamic CT phases, other liver lesions, diverse liver parenchymal backgrounds, and normal variations of organ structures should all be considered for this DL model in the future.

## Conclusion

This study demonstrated the application of the Mask R-CNN for usage in clinical imaging to improve the accuracy of liver lesion detection based on CT imaging from a medical perspective. Our study revealed that this single DL model cannot replace the complex and subtle medical evaluations of radiologists, but it can reduce tedious labor.

In the future, an AI CAD should be developed to help radiologists with the challenging task of lesion detection and analysis of different domains.

## Acknowledgments

The authors would like to thank the E-Da Hospital, E-Da Cancer Hospital and National Cheng Kung University (NCKU) Hospital Cancer Centers in Taiwan for the access provided to the Cancer Databanks. We also thank Miss Tzu-Shan Chen from the Department of Medical Research, E-Da Hospital in Taiwan, for assistance in statistical analysis.

## Author Contributions

**Conceptualization:** Ching-Juei Yang.

**Data curation:** Ching-Juei Yang, Hong-Ming Tsai, Yih-Jyh Lin, Hung-Wen Tsai.

**Formal analysis:** Ching-Juei Yang, Yu-Hua Dean Fang, Jing-Yao Wang.

**Funding acquisition:** Chien-Kuo Wang, Lee-Ren Yeh.

**Methodology:** Ching-Juei Yang, Hong-Ming Tsai.

**Resources:** Fong-Chin Su.

**Writing – review & editing:** Fong-Chin Su, Lee-Ren Yeh.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68(6):394–424. <https://doi.org/10.3322/caac.21492> PMID: 30207593
2. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol.* 2019; 16(10):589–604. <https://doi.org/10.1038/s41575-019-0186-y> PMID: 31439937
3. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet.* 2018; 391(10127):1301–1314. [https://doi.org/10.1016/S0140-6736\(18\)30010-2](https://doi.org/10.1016/S0140-6736(18)30010-2) PMID: 29307467
4. Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: epidemiology, etiology, and carcinogenesis. *J Carcinog.* 2017; 16:1. [https://doi.org/10.4103/jcar.JCar\\_9\\_16](https://doi.org/10.4103/jcar.JCar_9_16) PMID: 28694740
5. Zheng B, Zhu YJ, Wang HY, Chen L. Gender disparity in hepatocellular carcinoma (HCC): multiple underlying mechanisms. *Sci China Life Sci.* 2017; 60(6):575–584. <https://doi.org/10.1007/s11427-016-9043-9> PMID: 28547581

6. Global Burden of Disease Liver Cancer C, Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, et al. The Burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015. *JAMA Oncol.* 2017; 3(12):1683–1691. <https://doi.org/10.1001/jamaoncol.2017.3055> PMID: 28983565
7. McEvoy SH, McCarthy CJ, Lavelle LP, Moran DE, Cantwell CP, Skehan SJ, et al. Hepatocellular carcinoma: illustrated guide to systematic radiologic diagnosis and staging according to guidelines of the American Association for the Study of Liver Diseases. *Radiographics.* 2013; 33(6):1653–1668. <https://doi.org/10.1148/rg.336125104> PMID: 24108556
8. European Association for the Study of the Liver. Electronic address eee, European Association for the study of the L. EASL Clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol.* 2018; 69(1):182–236. <https://doi.org/10.1016/j.jhep.2018.03.019> PMID: 29628281
9. Omata M, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, et al. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatol Int.* 2017; 11(4):317–370. <https://doi.org/10.1007/s12072-017-9799-9> PMID: 28620797
10. Raoul JL, Forner A, Bolondi L, Cheung TT, Kloekner R, de Baere T. Updated use of TACE for hepatocellular carcinoma treatment: how and when to use it based on clinical evidence. *Cancer Treat Rev.* 2019; 72:28–36. <https://doi.org/10.1016/j.ctrv.2018.11.002> PMID: 30447470
11. Lee JH, Lee JM, Kim SJ, Baek JH, Yun SH, Kim KW, et al. Enhancement patterns of hepatocellular carcinomas on multiphasic multidetector row CT: comparison with pathological differentiation. *Br J Radiol.* 2012; 85(1017):e573–e583. <https://doi.org/10.1259/bjr/86767895> PMID: 22919011
12. Chung YE, Park MS, Park YN, Lee HJ, Seok JY, Yu JS, et al. Hepatocellular carcinoma variants: radiologic-pathologic correlation. *Am J Roentgenol.* 2009; 193(1):W7–W13. <https://doi.org/10.2214/AJR.07.3947> PMID: 19542386
13. Torbenson MS. Morphologic subtypes of hepatocellular carcinoma. *Gastroenterol Clin North Am.* 2017; 46(2):365–391. <https://doi.org/10.1016/j.gtc.2017.01.009> PMID: 28506370
14. Shah S, Shukla A, Paunipagar B. Radiological features of hepatocellular carcinoma. *J Clin Exp Hepatol.* 2014; 4(Suppl 3):S63–S66. <https://doi.org/10.1016/j.jceh.2014.06.009> PMID: 25755613
15. Reynolds AR, Furlan A, Fetzer DT, Sasatomi E, Borhani AA, Heller MT, et al. Infiltrative hepatocellular carcinoma: what radiologists need to know. *Radiographics.* 2015; 35(2):371–386. <https://doi.org/10.1148/rg.352140114> PMID: 25763723
16. Monzawa S, Ichikawa T, Nakajima H, Kitanaka Y, Omata K, Araki T. Dynamic CT for detecting small hepatocellular carcinoma: usefulness of delayed phase imaging. *Am J Roentgenol.* 2007; 188(1):147–153. <https://doi.org/10.2214/AJR.05.0512> PMID: 17179357
17. Hanna RF, Miloushev VZ, Tang A, Finklestone LA, Brejt SZ, Sandhu RS, et al. Comparative 13-year meta-analysis of the sensitivity and positive predictive value of ultrasound, CT, and MRI for detecting hepatocellular carcinoma. *Abdom Radiol (NY).* 2016; 41(1):71–90. <https://doi.org/10.1007/s00261-015-0592-8> PMID: 26830614
18. Mita K, Kim SR, Kudo M, Imoto S, Nakajima T, Ando K, et al. Diagnostic sensitivity of imaging modalities for hepatocellular carcinoma smaller than 2 cm. *World J Gastroenterol.* 2010; 16(33):4187–4192. <https://doi.org/10.3748/wjg.v16.i33.4187> PMID: 20806437
19. Choi YR, Chung JW, Yu MH, Lee M, Kim JH. Diagnostic accuracy of contrast-enhanced dynamic CT for small hypervascular hepatocellular carcinoma and assessment of dynamic enhancement patterns: results of two-year follow-up using cone-beam CT hepatic arteriography. *PLoS One.* 2018; 13(9): e0203940. <https://doi.org/10.1371/journal.pone.0203940> PMID: 30231076
20. Choi JY, Lee JM, Sirlin CB. CT and MR imaging diagnosis and staging of hepatocellular carcinoma: part I. Development, growth, and spread: key pathologic and imaging aspects. *Radiology.* 2014; 272(3):635–654. <https://doi.org/10.1148/radiol.14132361> PMID: 25153274
21. Choi JY, Lee JM, Sirlin CB. CT and MR imaging diagnosis and staging of hepatocellular carcinoma: part II. Extracellular agents, hepatobiliary agents, and ancillary imaging features. *Radiology.* 2014; 273(1):30–50. <https://doi.org/10.1148/radiol.14132362> PMID: 25247563
22. Ji GW, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, et al. Radiomic features at contrast-enhanced CT predict recurrence in early stage hepatocellular carcinoma: a multi-institutional study. *Radiology.* 2020; 294(3):568–579. <https://doi.org/10.1148/radiol.2020191470> PMID: 31934830
23. Park HJ, Park B, Lee SS. Radiomics and deep learning: hepatic applications. *Korean J Radiol.* 2020; 21(4):387–401. <https://doi.org/10.3348/kjr.2019.0752> PMID: 32193887
24. Wakabayashi T, Ouhmich F, Gonzalez-Cabrera C, Felli E, Saviano A, Agnus V, et al. Radiomics in hepatocellular carcinoma: a quantitative review. *Hepatol Int.* 2019; 13(5):546–559. <https://doi.org/10.1007/s12072-019-09973-0> PMID: 31473947

25. Castellino RA. Computer aided detection (CAD): an overview. *Cancer Imaging*. 2005; 5:17–19. <https://doi.org/10.1102/1470-7330.2005.0018> PMID: 16154813
26. Keil S, Behrendt FF, Stanzel S, Suhling M, Koch A, Bubenzer J, et al. Semi-automated measurement of hyperdense, hypodense and heterogeneous hepatic metastasis on standard MDCT slices. Comparison of semi-automated and manual measurement of RECIST and WHO criteria. *Eur Radiol*. 2008; 18(11):2456–2465. <https://doi.org/10.1007/s00330-008-1050-6> PMID: 18523775
27. Azer SA. Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: a systematic review. *World J Gastrointest Oncol*. 2019; 11(12):1218–1230. <https://doi.org/10.4251/wjgo.v11.i12.1218> PMID: 31908726
28. Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. *Radiographics*. 2017; 37(7):2113–2131. <https://doi.org/10.1148/rg.2017170077> PMID: 29131760
29. Ben-Cohen A, Klang E, Kerpel A, Konen E, Amitai MM, Greenspan H. Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations. *Neurocomputing*. 2018; 275:1585–1594. <https://doi.org/10.1016/j.neucom.2017.10.001>
30. Deng Z, Guo Q, Zhu Z. Dynamic regulation of level set parameters using 3D convolutional neural network for liver tumor segmentation. *J Healthc Eng*. 2019; 2019:4321645. <https://doi.org/10.1155/2019/4321645> PMID: 30918620
31. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology*. 2018; 286(3):887–896. <https://doi.org/10.1148/radiol.2017170706> PMID: 29059036
32. Chlebus G, Schenk A, Moltz JH, van Ginneken B, Hahn HK, Meine H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci Rep*. 2018; 8(1):15497. <https://doi.org/10.1038/s41598-018-33860-7> PMID: 30341319
33. Bilic P, Christ PF, Vorontsov E, Chlebus G, Chen H, Dou Q, et al. The liver tumor segmentation benchmark (LiTS)2019 January 01, 2019:[arXiv:1901.04056 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2019arXiv190104056B>.
34. Ronneberger O, Fischer P, Brox T, editors. U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer; 2015.
35. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, editors. 3D U-Net: learning dense volumetric segmentation from sparse annotation. *International conference on medical image computing and computer-assisted intervention*. Springer; 2016.
36. Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. *arXiv e-prints*2016. p. arXiv:1610.02177.
37. Han X. Automatic liver lesion segmentation using a deep convolutional neural network method. *arXiv preprint arXiv:170407239*. 2017.
38. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
39. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: hybrid densely connected unet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging*. 2018; 37(12):2663–2674. <https://doi.org/10.1109/TMI.2018.2845918> PMID: 29994201
40. Jiang H, Shi T, Bai Z, Huang L. Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes. *IEEE Access*. 2019; 7:24898–24909.
41. Tran S-T, Cheng C-H, Liu D-G. A Multiple Layer U-Net, Un-Net, for Liver and Liver Tumor segmentation in CT. *IEEE Access*. 2020.
42. Tang Y, Tang Y, Zhu Y, Xiao J, Summers RM, editors. E<sup>2</sup>Net: an edge enhanced network for accurate liver and tumor segmentation on CT scans. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2020.
43. Girshick RB, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014:580–587.
44. Girshick RB. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015:1440–1448.
45. Ren S, He K, Girshick RB, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015; 39:1137–1149.

46. He K, Gkioxari G, Dollár P, Girshick RB. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV). 2017:2980–2988.
47. Graffy PM, Sandfort V, Summers RM, Pickhardt PJ. Automated liver fat quantification at nonenhanced abdominal CT for population-based steatosis assessment. *Radiology*. 2019; 293(2):334–342. <https://doi.org/10.1148/radiol.2019190512> PMID: 31526254
48. Nguyen XV, Davies L, Eastwood JD, Hoang JK. Extrapulmonary findings and malignancies in participants screened with chest CT in the national lung screening trial. *J Am Coll Radiol*. 2017; 14(3):324–330. <https://doi.org/10.1016/j.jacr.2016.09.044> PMID: 28259326
49. Schwartz LH, Gandras EJ, Colangelo SM, Ercolani MC, Panicek DM. Prevalence and importance of small hepatic lesions found at CT in patients with cancer. *Radiology*. 1999; 210(1):71–74. <https://doi.org/10.1148/radiology.210.1.r99ja0371> PMID: 9885589
50. Erickson B, Kirk S, Lee Y, Bathe O, Kearns M, Gerdes C, et al. Radiology data from the cancer genome atlas liver hepatocellular carcinoma [TCGA-LIHC] collection. *Cancer Imaging Arch*. 2016.
51. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit imaging*. 2013; 26(6):1045–1057. <https://doi.org/10.1007/s10278-013-9622-7> PMID: 23884657
52. Lin T-Y, Maire M, Belongie SJ, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: common objects in context. *ArXiv*. 2014;abs/1405.0312.
53. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*. 1986:54–75.
54. Rousseelet G, Pernet C, Wilcox RR. A practical introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations. 2019.