Noname manuscript No.
(will be inserted by the editor)

Classification of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma Based on Multi-phase CT Scans

Donlapark Pornnopparath · Papangkorn Inkeaw · Jeerayut Chaijaruwanich · Patrinee Traisathit · Patumrat Sripan · Nakarin Inmutto · Wittanee Na Chiangmai · Donsuk Pongnikorn · Imjai Chitapanarux

Received: date / Accepted: date

The total number of words of the manuscript (including entire text from title page to figure legends): approx. 7600 words
The number of words of the abstract: 182 words
The number of figures: 8 figures
The number of tables: 6 tables

Corresponding author
Papangkorn Inkeaw
Tel. +66 5394 3412 ext.111 Fax. +66 5389 2281
E-mail: papangkorn.i@cmu.ac.th

Donlapark Pornnopparath · Patrinee Traisathit
Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
E-mail: donlapark.p@cmu.ac.th, patrinee@gmail.com

Papangkorn Inkeaw (corresponding author) · Jeerayut Chaijaruwanich
Data Science Research Center, Department of Computer Science, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
E-mail: {papangkorn.i, jeerayut.c}@cmu.ac.th

Patumrat Sripan
Research Institute for Health Sciences, Chiang Mai University, Chiang Mai 50200, Thailand
E-mail: pspatumrat3@gmail.com

Nakarin Inmutto · Wittanee Na Chiangmai · Imjai Chitapanarux
Department of Radiology, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand
E-mail: {nakarin.i, wittanee.n, imjai.chitapanarux}@cmu.ac.th

Donsuk Pongnikorn
Cancer Registry Unit, Lampang Cancer Hospital, Lampang 52000, Thailand
E-mail: donsukp@hotmail.com

Abstract Liver and bile duct cancers are leading causes of worldwide cancer death. The most common ones are hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC). Influencing factors and prognosis of HCC and ICC are different. Precise classification of these two liver cancers is essential for treatment and prevention plans. The aim of this study is to develop a machine-based method that differentiates between the two types of liver cancers from multi-phase abdominal computerized tomography (CT) scans. The proposed method consists of two major steps. In the first step, the liver is segmented from the original images using a convolutional neural network model, together with task-specific pre-processing and post-processing techniques. In the second step, by looking at the intensity histograms of the segmented images, we extract features from regions that are discriminating between HCC and ICC, and use them as an input for classification using support vector machine model. By testing on a dataset of labeled multi-phase CT scans provided by Maharaj Nakorn Chiang Mai Hospital, Thailand, we have obtained 88% in classification accuracy. Our proposed method has a great potential in helping radiologists diagnosing liver cancer.

Keywords Classification · Machine Learning · Image Processing · Liver Neoplasms · Tomography

Declarations

Funding

This study was funded by Faculty of Medicine, Chiang Mai University, Thailand.

Conflicts of interest

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval

This study was approved and monitored by the Ethics Committee of Faculty of Medicine, Chiang Mai University, Thailand.

Authors' contributions

Conceptualization: Donlapark Pornnopparath, Papangkorn Inkeaw; Methodology: Donlapark Pornnopparath, Papangkorn Inkeaw; Software: Donlapark Pornnopparath, Papangkorn Inkeaw; Validation: Donlapark Pornnopparath, Papangkorn Inkeaw; Formal Analysis: Donlapark Pornnopparath, Papangkorn Inkeaw; Investigation: Donlapark Pornnopparath, Papangkorn Inkeaw; Data Curation: Donlapark Pornnopparath, Patumrat Sripan, Nakarin Inmutto; Writing Original Draft: Donlapark Pornnopparath, Papangkorn Inkeaw, Patumrat Sripan; Writing Review & Editing: Jeerayut Chaijaruwanich, Patrinee Traisathit, Nakarin Inmutto, Wittanee

Na Chiangmai, Donsuk Pongnikorn, Imjai Chitapanarux; Visualization: Donlapark Pornnopparath, Papangkorn Inkeaw; Supervision: Jeerayut Chaijaruwanich, Patrinee Traisathit, Imjai Chitapanarux; Project Administration: Nakarin Inmutto; Funding Acquisition: Nakarin Inmutto;

1 Introduction

As reported by the Global Cancer Statistics 2018 [7], liver and bile duct cancers are leading causes of worldwide cancer death as they were the second most responsible for cancer mortality in male and the sixth most responsible for cancer mortality in female, combined into an estimate of over 750,000 deaths in 2018. The highest incidence rate is found in the Southeast Asia, East Asia, and Africa [19]. It is also the most common cancer in Thai male and the second most common cancer in Thai female with an age-standardized rate (ASR) per 100,000 persons of 33.9 and 12.9, respectively [18]. The most common types of liver cancer are hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC).

Since risk factor, prognosis, and also the management of HCC and ICC are different [20, 21, 29, 35, 36, 11], accurate screening is essential for treatment and prevention plans. Radiology imaging techniques, such as multi-phase computed tomography (CT) and magnetic resonance imaging (MRI), have been commonly used for diagnosis and classification of HCC and ICC. In practice, both can be identified by relative intensity of lesions in four-phase CT images of a patient after being injected with a contrast medium. HCC lesions exhibit low attenuation on non-contrast images, early peak of enhancement on arterial phase and followed by a continuous decrease in attenuation on portal venous and delayed phases, while ICC lesions typically show homogeneously low attenuation on non-contrast scans, faint peripheral enhancement on arterial phase and gradual centripetal enhancement on portovenous and delayed phases. Accuracy of these techniques depends on the size of tumors, complications of cirrhosis and the radiologists' experience [9, 25, 32]. However, Despite of existing diagnosis guidelines for these two types of liver cancer [8, 12, 13, 30], HCC can still be mistaken with ICC due to similar enhancement patterns in CT images.

Despite the importance of the task, classification of HCC and ICC using machine learning (ML) has rarely been explored. Based on our knowledge and according to recent reports on application of artificial intelligence on hepatology [4, 5], there is no article of machine learning techniques for classifying HCC and ICC listed in PubMed, EMBASE, and the Web of Science. We found only a study from [26] which employed Convolutional Neural Network (CNN) for the classification task on CT images whose tumor regions were semi-automatically segmented by a radiologist expert. Although the CNN has excellent performance in distinguishing HCC and ICC, it requires a large training dataset in order to avoid over-fitting. Another issue of CNN is the lack of interpretability.

For ML applications of other CT-based liver diagnoses, liver segmentation is often required in order to exclude unnecessary factors from image analysis. Over the past few years, deep learning models have shown to be very successful in medical segmentation. The reports on ISBI 2017 and MICCAI 2017 liver segmentation challenges [6] indicate the success of deep learning models, especially the U-Net [31], in terms of segmentation accuracy. However, due to variation in liver shapes, scanners, contrast enhancement and contrast mediums, automating liver segmentation across different domains poses a challenging task. In the case of multi-phase images, it is expensive and time-consuming to obtain expert-verified ground truth segmentation of images in all phases. Thus, in many situations, it is desirable to build a segmentation model using images from a single phase that also works well on images from all other phases. Since many deep learning models are usually

biased toward the training set, they do not fit this description. When we apply these models to images in high-contrast medium-enhanced phases, the lesions are often excluded from the segmentation.

In this work, we propose a classifier for liver cancer diagnosis of HCC or ICC from multi-phase CT images. The proposed method takes multi-phase CT images of a patient, who has been diagnosed with a liver cancer. However, the diagnosis cannot identify whether the type of cancer is HCC or ICC. The main purpose of this method is to classify the type of cancer that appears in CT images and marks the regions that influence its decision.

Our approach to the classification is based on the changes in pixel intensity; a blob that is darker or lighter than its background is marked as a lesion, and the changes in color of the lesion across multi-phase images should be able to indicate the type of the cancer. However, without restricting the "background" region, we would run into several issues: (1) it would be quite difficult to tell the lesions apart from other small organs (2) the background is typically incongruous from having multiple organs, making it impossible to identify a liver lesion. Thus, for our approach to be effective, it is essential that the searching region is limited to only the liver. From this point of view, we chose to perform liver segmentation before the classification.

As a whole, our HCC-ICC classification method consists of two major steps:

1. Liver segmentation for capturing the region of interest
2. Classification using information from discriminative regions on liver.

During the segmentation step, several techniques are employed to address the segmentation issues. First, we use a histogram-based technique to match the color distribution between the non-contrast and high-contrast phases. Secondly, we have modified U-Net so that its output map at each pixel behaves more like a probability value instead of being strictly 0 or 1. Lastly, these probability maps are re-calibrated by a graphical model that puts less emphasis on contrast and more on proximity of high-probability regions. As a result, the lesions that were not detected before will be included in the post-processed image, which is then passed to the classification model.

Within the classification model, we propose a feature extraction method based on the differences in pixel intensity between the phases. These features indicate the discriminative regions in the multi-phase CT images, which can be used to identify the type of liver cancers. We then use these features to build a HCC-ICC classifier. Experimental results show that, among several machine learning models, support vector machine performs the best on this task. As an additional benefit, the discriminative regions can help radiologists spot the lesions and see the reasoning behind the classification results.

2 Materials

In this study, two datasets of 2D CT images were used: (1) a dataset for classification models, consisting of multi-phase images of liver cancer patients (four images per patient) who has already been diagnosed with either HCC or ICC. Note that this dataset does not come with ground truth liver segmentation. (2) a dataset for segmentation models, consisting of single-phase images, either in non-contrast or

delayed phase, labeled with ground-truth liver segmentation. These datasets are explored in more details below.

2.1 CMU HCC-ICC dataset

Under an institutional review board-approved protocol, we obtained CT imaging data and clinical data of 187 HCC cases and 70 ICC cases diagnosed during 2013-2014 from Maharaj Nakorn Chiang Mai hospital. The diagnosis of HCC and ICC were based on clinical information, laboratory test, CT imaging and treatment response.

The abdominal CT scans, starting from the dome of the diaphragm to the iliac crest, were performed with a multi-detector CT scan (SOMATOM definition, Siemens, Germany). The CT parameters were 120 kVp, 200-400 mAs, 0.6 mm $\times$ 64 section collimation with a single breath-hold helical acquisition.

With these settings, the non-contrast images were obtained with 5-mm-thick axial sections. After that, the patient was intravenously administered with approximately 100-150 ml of nonionic contrast material containing 300-350 mg/ml of iodine. The triple-phase high-contrast images were then acquired successively after the administration; the hepatic arterial phase images were acquired after 25-30 seconds, the portal venous phase images after 60-70 seconds and the delayed phase images after five minutes.

The CT images, stored in DICOM format, were displayed in a viewing software (Synapse PACS;FUJIFILM Medical Systems U.S.A) with 50 HU window level and 300-500 window width. Using the image capture function, liver masses were converted to 8-bit Joint Photographic Experts Group (JPEG) format (512 $\times$ 512 pixels).

Medical professionals then use their knowledge to pick a single image from each phase, combined into a sequence of four CT scans that best identifies the type of liver cancer. An example of multi-phase images of a patient is shown in Figure 1. As the names suggest, the lesions become clearly visible from the contrast media in high-contrast phases.

Because of the small number of samples, we settled with the 10-fold cross validation as our out-of-sample testing method. This means that, in each trial, we will be training our classification model on 168-169 HCC samples and 63 ICC samples, and testing on 18-19 HCC samples and 7 ICC samples.

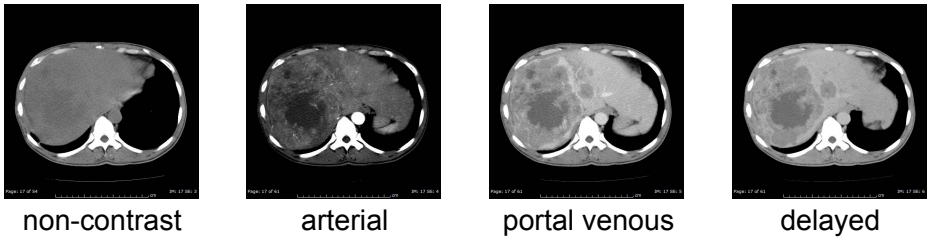

Fig. 1 A sample of multi-phase CT images in CMU HCC-ICC dataset.

2.2 Datasets for liver segmentation

Due to high variability of liver's shapes in 2D cross-sectional images, we need to obtain as much labeled data as possible in order to train an efficient deep learning model for liver segmentation. Our training data for liver segmentation was extracted from various sources as indicated in Table 1, combined into a total of 171 DICOM series of CT scans in the abdominal area, together with the manual segmentation performed by several expert radiologists. Since each image in our CMU HCC-ICC dataset (see details in subsection 2.1) always contains the liver, we can reduce the number of samples (slices) by picking only slices that contain the liver. From the original series, we extracted a total number of 18,739 slices with ground truth segmentations, which were then split into a training set of 16,697 slices from 157 patients and a validation set of 2,042 slices from 17 patients for training the model.

Table 1 List of publicly available datasets of liver segmentation that we used to train our segmentation model

Dataset	Institution	#Volumes	References
3Dircadb-01	IRCAD	20	[14]
Sliver'07	IRCAD	20	[17]
LiTS	Various	131	[6]

3 Proposed method

Assume that each set of multi-phase CT images obtained in subsection 2.1 has either HCC lesions or ICC lesions, but not both. The goal is to build an automatic system that correctly classifies the liver cancer type from these four images. As mentioned in the introduction, our approach to this problem consists of (1) liver segmentation and (2) classification. The details of each step will be explained in the next sections.

3.1 Liver segmentation

The segmentation follows the usual preprocess-segment-postprocess workflow Figure 2, specifically designed to be robust to complication of lesions in high-contrast phases. In this section, we describe all the methods that we used to obtain the robustness to complication of liver lesions.

3.1.1 Histogram matching

Although the segmentation technique can be directly apply to CT images in the CMU HCC-ICC dataset in non-contrast, portal venous and delayed phases, CT images in arterial phase need to be adjusted in contrast level because they are relatively darker than the other phases. Therefore, we match its histogram with the

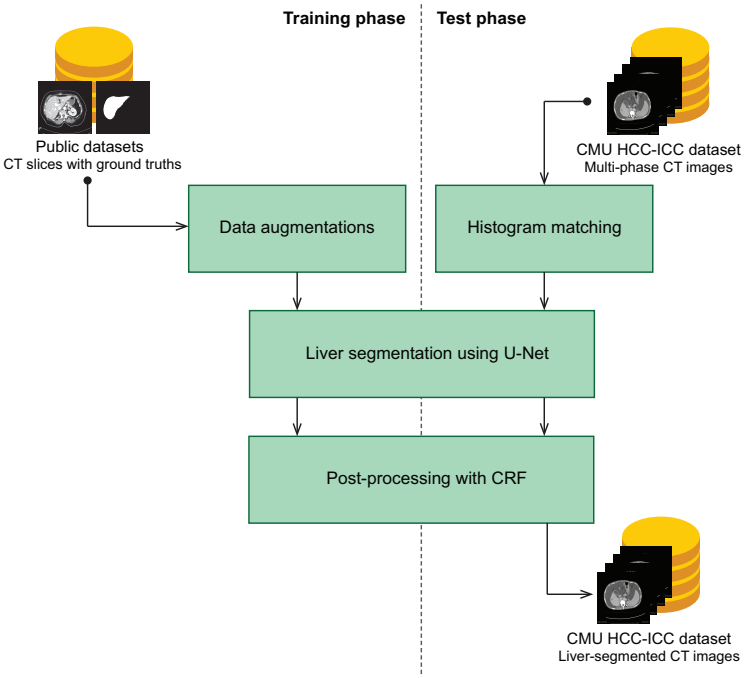

Fig. 2 An overview of proposed segmentation workflow.

delayed phase image of the same patient before segmenting with U-Net. Suppose that $P_A(x)$ and $P_D(x), 0 \leq x \leq 255$ is the cumulative histogram (with 256 bins) of the arterial phase image A and delayed-phase image D, respectively. Then, by linearly interpolating P_D into a continuous cumulative distribution function P'_D, we match P_A to $P'(D)$ by transforming A as follows:

$$\text{For each } x_i \in A, \quad x_i \mapsto x'_i \quad \text{if} \quad P_A(x_i) = P'_D(x'_i).$$

See Online Resource 1 for an example of matching images from two different phases. This does not only brighten the liver in the arterial phase but also reduces the contrast in that area.

3.1.2 Liver segmentation using U-Net

U-Net architecture For the segmentation task, we adopt U-Net [31]. The network architecture of U-Net consists of a downward encoding path where the model aggregates local information and the upward decoding path where the encoded images was upsampled and reinforced by the information from the downward path in order to recover spatial information. In contrast to the network architecture in [31], we reduced the output dimensions in some of the layers in order to reduce the training time (Figure 3).

Loss function As U-Net is trained via minimizing a *loss function*, we have to define a loss function that is appropriate for our task. Let M be the total number

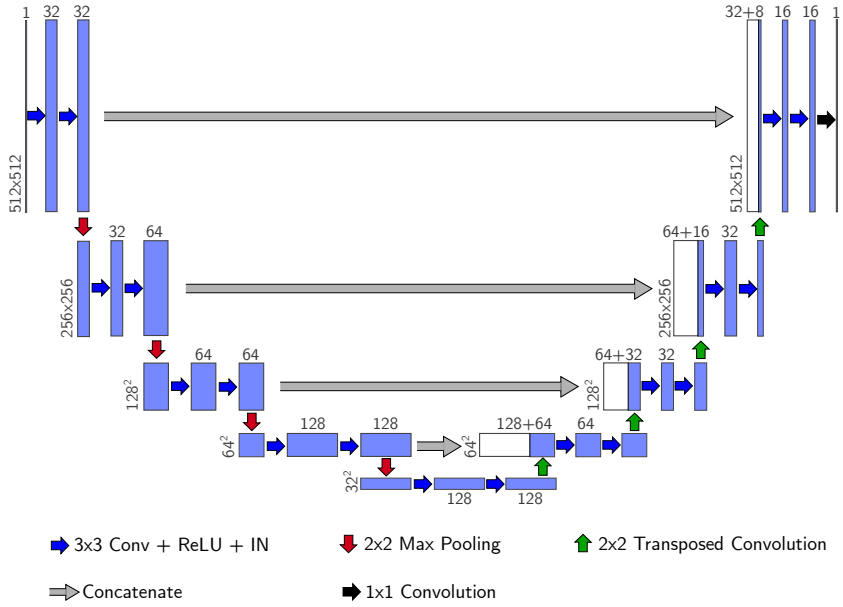

Fig. 3 A modified U-Net architecture in which an instance normalization layer is inserted after every convolutional layer.

of images in one training batch, N the total number of pixels in one image, $\mathbf{u}^m = [u_1^m, u_2^m, \ldots, u_N^m] \in [0,1]^N$ be the probabilistic output of applying U-Net on m-th image and $\mathbf{v}^m = [v_1^m, v_2^m, \ldots, v_N^m] \in \{0,1\}^N$ be the ground truth segmentation of m-th image. Our loss function is:

$$L_{dcce} = L_{dc} + L_{ce}. \tag{1}$$

which is a combination of the DICE loss

$$L_{dc} = -\frac{1}{M}\sum_{m=1}^{M} \mathrm{DICE}(\mathbf{u}^m, \mathbf{v}^m) \tag{2}$$

$$\mathrm{DICE}(\mathbf{u}^m, \mathbf{v}^m) = \frac{2\mathbf{u}^m \cdot \mathbf{v}^m}{\sum_i u_i^m + \sum_i v_i^m}. \tag{3}$$

and the weighted cross-entropy loss

$$L_{ce} = -\frac{1}{M}\sum_{m=1}^{M} [\eta^{(1)} v_i^m \log u_i^m + \eta^{(2)}(1 - v_i^m)\log(1 - u_i^m)],$$

Intuitively, $\mathrm{DICE}(\mathbf{u}^m, \mathbf{v}^m)$ measures the similarity between the prediction and the ground truth. The value is one only if both images match, and zero if they are disjoint. $\eta^{(1)}$ and $\eta^{(2)}$ are per-class weights, chosen in order to counteract class imbalance. In this study, we set $(\eta^{(1)}, \eta^{(2)}) = (0.7, 0.3)$ in order to emphasize more on positive labels incorrectly classified as negative.

3.1.3 Post-processing with Conditional Random Field

The segmentations from U-Net might contain some false negatives due to contrast enhancement. Fortunately, training with weighted-cross entropy loss gives us probability-valued predictions (in contrast to strictly 0's and 1's). We then post-process U-Net's outputs with fully connected conditional random fields (CRF), which changes the probability values based on proximity and color similarity of pixels in the original images [22]. The mathematical formulation of the fully connected CRF for image segmentation in Online Resource 2.

We give an example in Online Resource 3 which shows that using the mixed loss L_{dcce} allows CRF to effectively re-calibrate the output from U-Net which leads to more accurate segmentation.

Figure 4 shows segmentation of the liver in high-contrast phases. The results shown in the last row indicate that the segmentation can be improved with artificial lesions and instance normalization.

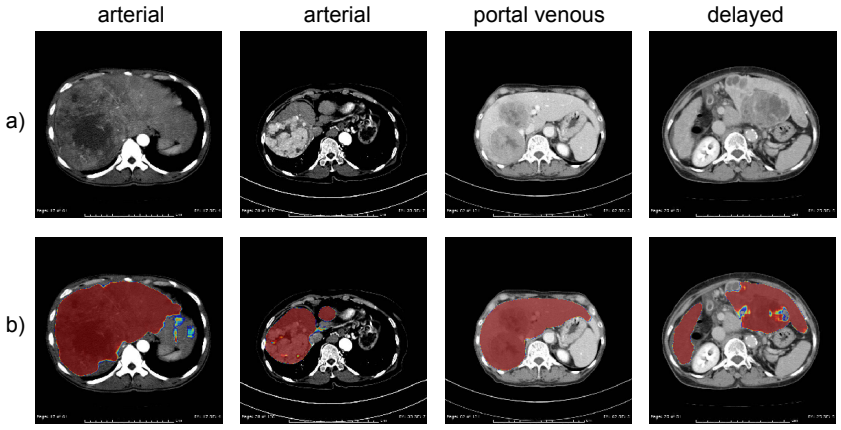

Fig. 4 Examples of post-processed liver segmentations from HCC-ICC dataset. From top to bottom row: **a) original image, b) segmentation with U-Net and CRF.**

3.2 Classification

In this section, we describe the process of classifying HCC and ICC based on multi-phase CT images. The diagram of the proposed classification technique is shown in Figure 5, consisting of four steps: (1) analysis of pixel intensity (2) determination of potentially discriminative regions (3) discriminative feature extraction and (4) classification of HCC and ICC. Each of these steps is described as follows.

3.2.1 Analysis of pixel intensity

As HCC and ICC can be diagnosed by observing the change of color of contrast medium in multi-phase CT images, we first analyzed the color intensity distributions of the CT images. For each CT image, the color intensity distribution is

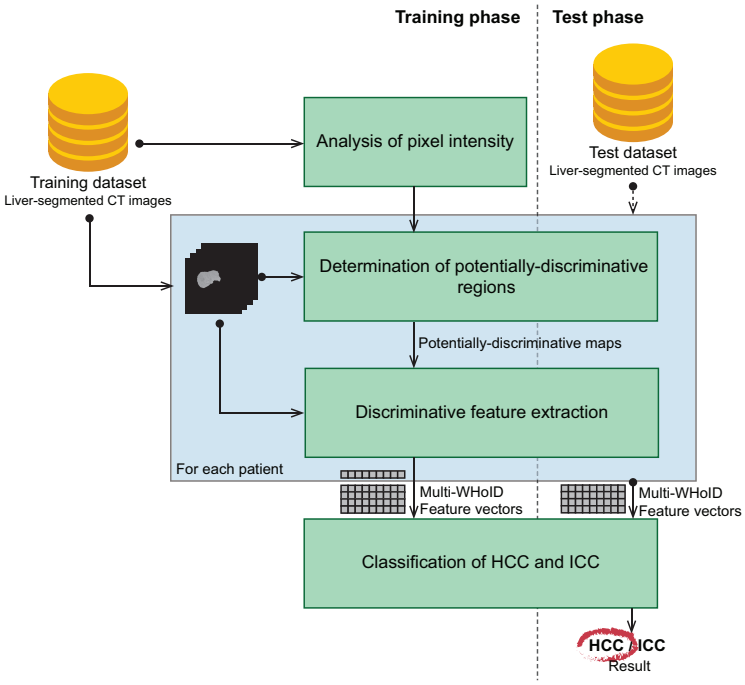

Fig. 5 Flowchart of the proposed classification technique

constructed. We found that the color intensity values tend to be around a central value and their distributions are almost normal (see examples in Online Resource 4). The color intensity values of pixels on normal regions are tightly clustered around the mode, while that of the lesions are either lower (called hypodensity region) or higher (called hyperdensity region) than the mode.

Next, for each enhancement phase, we constructed the color intensity distributions from all CT images of HCC and ICC as given in Online Resource 5. As the CT images of the same enhancement phase were created under the same settings, the mode of color intensity distributions of HCC and ICC are not significantly different. Moreover, the color intensity of normal regions on CT images of HCC and ICC varies in the same distribution. As a result, each enhancement phase can be represented as the color intensity distribution constructed from all HCC and ICC images (Online Resource 6). We denote the most frequent value of color intensities for non-contrast, arterial, portal and delayed phases by t_N, t_A, t_P and t_D, respectively. These values will be used as parameters to determine lesion regions in the next step.

3.2.2 Determination of potentially discriminative regions

After obtaining the most frequent value of color intensities on a set of CT images for each enhancement phase, the value is used as a parameter for determining the regions on CT images that have high potential for discriminating between HCC and ICC. In this work, lesions (both hypodensity and hyperdensity regions) are emphasized as potentially discriminative regions because physicians consider the

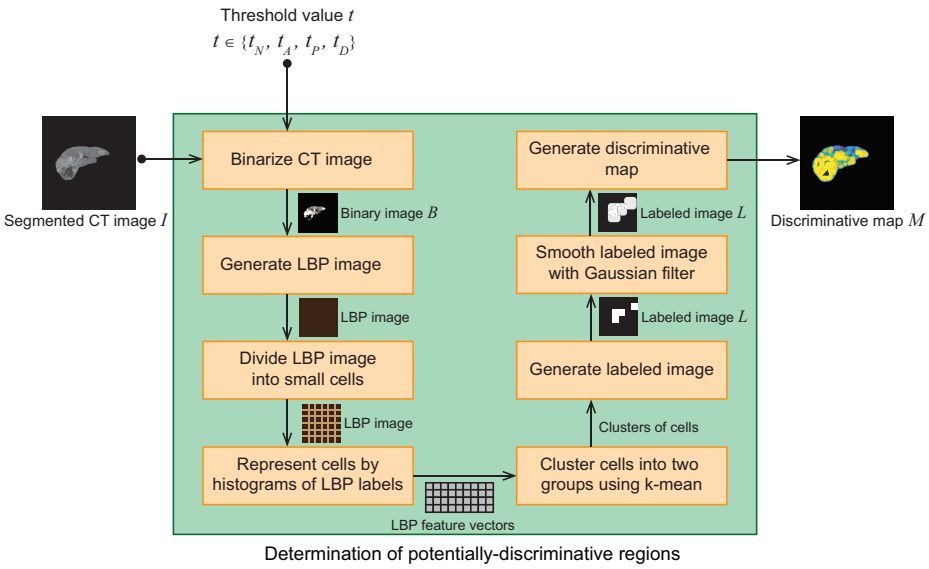

Fig. 6 Processing flow of determination of potentially discriminative regions

intensity enhancement of those regions on multi-phase CT images to distinguish between HCC and ICC. Figure 6 summarizes the processing flow of this step. To determine the potentially discriminative regions, a CT image $I(i,j)$ is first transformed into a binary image B defined by:

$$B(i,j) = \begin{cases} 1 & \text{if } I(i,j) \geq t \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where the threshold value t is selected from $\{t_N, t_A, t_P, t_D\}$ depending on the enhancement phase of the input CT image. An example binary image generated from a CT image on arterial phase of HCC is given in Figure 7. As the intensity value of all pixels on the same lesion region is lower or higher than the threshold value, the texture of lesion regions on binary image is uniform. In contrast, the texture of the other areas is rough because the pixel intensity values vary around the threshold value. Therefore, we can distinguish the lesion regions from the others by analyzing the patterns of the texture.

Second, we adopted Local Binary Patterns (LBP) , which is one of the most popular features used for texture classification, as texture descriptor. For each pixel, a binary number is extracted by comparing the pixel with its 8 neighbors. The neighbors having smaller value than that of the central pixel will have the bit 0, and the other neighbors having value equal to or greater than that of the central pixel will have the bit 1. For each given central pixel, the binary number is obtained by concatenating all these binary bits in a clockwise manner. It can be represented as a decimal value called LBP label. A new image, called LBP image, is generated by replacing each pixel on the binary image with its LBP label.

Then, the LBP image is divided into equally sized smaller cells. For each cell, a histogram of LBP labels is calculated. This histogram can be seen as a 256-

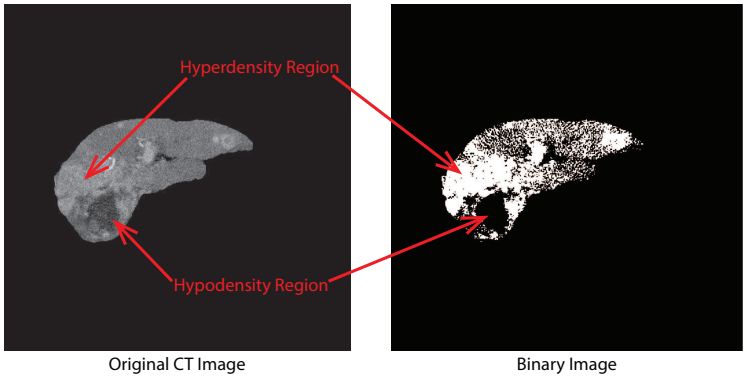

Fig. 7 An example binary image of a CT image. The binary image (right) was generated from a CT image (left) on arterial phase of HCC. Dark-uniform and light-uniform areas on the binary image correspond to hypodensity and hyperdensity regions on the CT image, respectively.

dimensional feature vector. The length of the feature vector can be reduced to 36 by considering rotational invariant patterns.

Next, all cells represented by LBP feature vectors are clustered into two groups i.e. uniform and non-uniform patterns. In this work, k-mean clustering (i.e. $k = 2$) was used. Then, a labeled image L is constructed. The value of $L(i, j)$ will be 1 if its corresponding cell has the uniform pattern; otherwise it will be 0. As the labeled image L has the size of $\frac{\text{height}}{\text{cell's size}} \times \frac{\text{width}}{\text{cell's size}}$, it is enlarged to have the same size with the input CT image. Then, the image L is smoothed by convolution with a Gaussian function defined as:

$$G(i, j) = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}}, \tag{5}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

Finally, a discriminative map M is generated by

$$M(i, j) = \begin{cases} L(i, j) & \text{if } (i, j) \in \mathbf{R} \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where $\mathbf{R}$ is the set of pixels on the liver area. An example of a discriminative map is shown in Figure 8. The value of $M(i, j)$ is 1 if the position (i, j) lies on lesion regions. On the other hand, the value is close to zero if the position (i, j) is far from lesion regions. The function M can then be viewed as a discriminative score of pixel (i, j).

3.2.3 Discriminative feature extraction

As the color intensities are used as a criterion to distinguish HCC and ICC based on CT images, the intensity histogram is adopted to pack intensity information of multi-phase CT images in a feature vector. Normally, an intensity histogram is constructed by considering all pixels on image with the same weight. In this work,

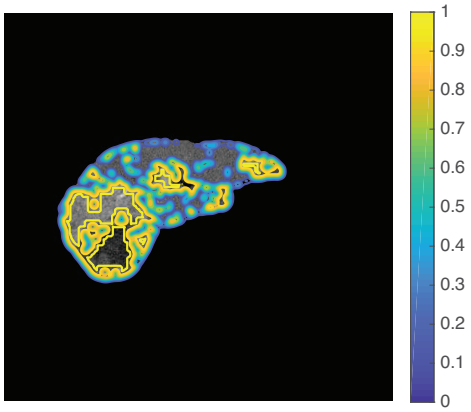

Fig. 8 Contour plot illustrating a discriminative map of a CT image

Algorithm 1: Weighted Histogram of Intensity (WHoI)

Input: An input image I, a weight value matrix W and the number of bins b.
Output: A weighted histogram of intensity H
begin
  $[nRow, nCol] \leftarrow \text{size}(I)$
  $binSize \leftarrow \lfloor \frac{256}{b} \rfloor$
  $H \leftarrow \text{zeros}(1, b)$
  for $i = 1$ **to** $nRow$ **do**
    for $j = 1$ **to** $nCol$ **do**
      $H(\lfloor \frac{I(i,j)}{binSize+1} \rfloor + 1) \leftarrow H(\lfloor \frac{I(i,j)}{binSize+1} \rfloor + 1) + W(i,j)$

we propose Weighted Histogram of Intensity (WHoI) that all pixels are weighted by their discriminative scores. Algorithm 1 describes the construction of a WHoI. A feature vector of a CT image I is constructed by:

$$F_I = \text{WHoI}(I, M, b). \tag{7}$$

For a partial, we have four CT images generated from different enhancement phases i.e. non-contrast, arterial, portal and delayed phases. A feature vector can be formed by:

$$F = F_N \oplus F_A \oplus F_P \oplus F_D, \tag{8}$$

where N, A, P and D are the CT image of non-contrast, arterial, portal and delayed phased, respectively. We named the feature vector as Multi-phase WHoI Descriptor (Multi-WHoID).

3.2.4 Classification of HCC and ICC

Using the extracted features, we adopted the support vector machine (SVM) as the baseline classifier. The goal of SVM is to find a separating hyperplane which

maximizes the margin between the two classes [10]. The SVM predicts the class label $\hat{y}$ for any given data point $\mathbf{x}$ by:

$$\hat{y} = f(\mathbf{x}) = \text{sign}(w^T \phi(\mathbf{x}) + b), \tag{9}$$

where $\phi(\mathbf{x})$ maps $\mathbf{x}$ into a higher-dimensional space, w is the weight vector and b is the bias.

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ be a training dataset where $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{ip}]^T$ is a data point represented by a feature vector and $y_i \in \{+1, -1\}$ is a class label of $\mathbf{x}_i$. As we know that $w = \sum_{i=1}^n \alpha_i y_i \phi(\mathbf{x}_i)$, the SVM algorithm maximizes the dual representation of the maximum margin problem defined as

$$L(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j), \tag{10}$$

with respect to $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ subject to the constraints

$$0 \leq \alpha_i \leq c, i = 1, 2, ..., n \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0, \tag{11}$$

where $c > 0$ is the regularization parameter that controls a trade-off between training error and generalization, and $k(\cdot)$ is the kernel function. Thus, the decision function becomes

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} y_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b \right). \tag{12}$$

Any data point for which $\alpha_i = 0$ will not appear in Equation 12 and plays no role in making the prediction for new data point. The remaining data points that satisfy $\alpha_i > 0$, commonly named as *support vectors*, correspond to points that lie on the maximum margin hyperplanes in feature space. As we found a value of $\boldsymbol{\alpha}$, we can determine the value of bias b by

$$b = \frac{1}{|\mathbf{S}|} \sum_{i \in \mathbf{S}} \left(y_i - \sum_{j \in \mathbf{S}} \alpha_j y_j k(\mathbf{x}_i, \mathbf{x}_j) \right), \tag{13}$$

where $\mathbf{S}$ is the set of indices of the support vectors.

The kernel functions investigated in this study are the following:

$$\text{Linear: } k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

$$\text{Radial basis function: } k(\mathbf{x}_i, \mathbf{x}_j) = \exp(\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

$$\text{Chi-square: } k(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{p} \frac{2 x_{ik} x_{jk}}{x_{ik} + x_{jk}}$$

$$\text{Generalized histogram intersection: } k(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{p} \min(|x_{ik}|^\alpha, |x_{jk}|^\alpha).$$

with $\gamma > 0$ and $\alpha > 0$ being the kernel parameter of Radial Basis Functions (RBF) and Generalized Histogram Intersection kernel (GHI), respectively. The linear and

RBF kernels are mostly used in pattern recognition works while the Chi-square and GHI kernels are developed for histogram based feature vectors.

In this work, the SVM hyper-parameter c and the kernel's parameter γ were optimized by grid-search method with cross validation. The values of c and γ in the set $\{2^{-10}, 2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 2^{0}, 2^{2}, 2^{4}, 2^{6}, 2^{8}, 2^{10}\}$ were evaluated while the kernel's parameter α was set to be 1.

4 Results

4.1 Experiment protocol

We evaluate the segmentation method on the public datasets and the classification method on the HCC-ICC dataset. Note that the public datasets mentioned in Table 1 come with the ground truth segmentation but do *not* have the HCC-ICC labels, while the CMU HCC-ICC dataset does *not* come with the segmentation while having the HCC-ICC labels.

Evaluation of liver segmentation Let $u_i \in [0,1]^{512\times512}$ be the U-Net's segmentation and $v_i \in \{0,1\}^{512\times512}$ be the ground truth segmentation of the i–th image. The dice score is defined by:

$$\text{Dice score} = \frac{2}{n}\sum_{k=1}^{n}\frac{\sum_i u_i^k v_i^k}{\sum_i u_i^k + \sum_i v_i^k},$$

where n is the number of samples in the test set. To measure the performance of U-Net under different normalization techniques (none, BN or IN) and loss functions (L_{dc} or L_{dcce}), DICOM series of 17 patients were taken from the segmentation dataset (Table 1) as a hold-out test set. Then, we trained each of these models on the training set 10 times and compute the mean of the Dice scores on the test set. Each of these models was implemented in Keras (available on `https://keras.io`). The training consisted of 8 instances per batch, 1,000 batches per epoch and 20 epochs per session. After each epoch, the model was evaluated on the validation set and the training would be stopped early if the validation dice score had not been improved for four consecutive epochs. We set the initial learning rate to be 10^{-3} and reduce the learning rate by a factor of 0.2 if the validation Dice score has not been improved for three epochs.

Evaluation of HCC-ICC classification To find the most suitable features and models for the classification, we performed several experiments, each of which consisted of 30 trials of 10-fold cross validation (10-CV) on the HCC-ICC dataset. For each trial, the dataset was randomly shuffle. The available samples were partitioned into 10 groups. Then nine of the groups are used to train the classifier that are then evaluated on the remaining group. This process is then repeated for all 10 possible choices for the hold-out group. The performance of the proposed method was evaluated in term of recognition accuracy, True Positive Rate (TPR), True Negative Rate (TNR), Positive Predictive Value (PPV) and Negative Predictive Value (NPV). Note that both PPV and NPV are dependent on the prevalence of the diseases on the population of interest [23].

To compare the performance of two methods with the average performance score of multiple-experiments, we adopt the Wilcoxon signed-rank test at the significance level of 0.05, and we reject the null hypothesis if the recognition accuracy provided by the classifier A is significantly higher than that provided by the classifier B.

4.2 Performance of proposed segmentation method

We firstly evaluated the performance of U-Net under different normalization techniques and loss functions. Figure 9 reports the performances of segmentation models in term of the mean Dice scores with standard deviations.

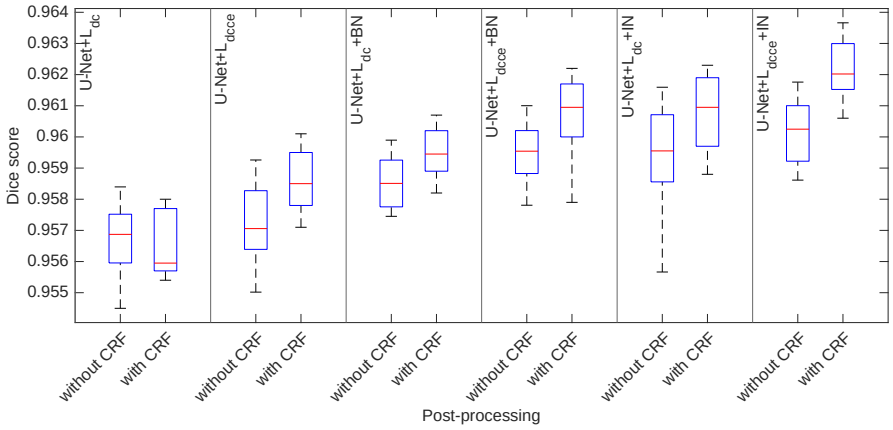

Fig. 9 Dice scores of U-Net with various design choices.

We can see improvements in segmentation by incorporating normalization techniques in the model, and U-Net with the mixed loss L_{dcce} and IN gives the best predictions compared to other designs. We also observe that CRF almost always helps improve the segmentations, with an exception of simple U-Net with the Dice loss.

4.3 Performance of Multi-WHoID with different SVM kernel functions

In this experiment, we determine the most effective SVM kernel function and the number of bins for distinguishing HCC from ICC using multi-WHoID. The recognition accuracies are shown in Figure 10. The highest average accuracy of 88.19% was achieved when the number of bins in each WHoI was set to be 64. Thus, in this work, we decided to use SVM with the GHI kernel function as the classifier. The multi-WhoID formed by four WHoI with 64 bins was used as feature descriptor.

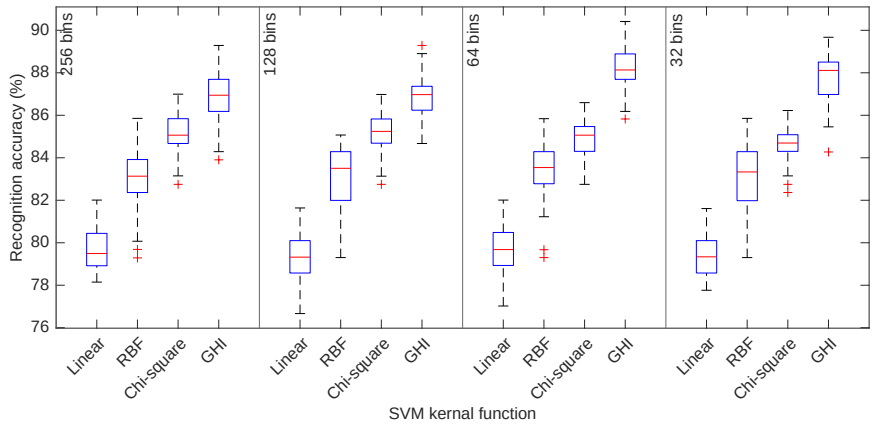

Fig. 10 The recognition performance of Multi-WHoID (with different numbers of bins for each WHoI) with four different SVM kernel functions.

4.4 Performance of proposed method on distinguishing HCC and ICC

The performance of the proposed method in terms of TPR, TNR, PPV and NPV is given in Figure 11. The results indicate that the proposed method correctly detected 89.30% of the HCC patients and 84.42% of the ICC patients.

Among all patients classified as having HCC, 95.18% of them were HCC patients. In contrast, only 69.44% of all patients identified with ICC were actually ICC patients. We can see that the NPV is significantly lower than the PPV because of a small sample size of ICC patients in the CMU HCC-ICC dataset.

4.5 Performance comparison between Multi-WHoID and standard intensity histograms

We compared the performance between Multi-WHoID and standard intensity histograms (Figure 11 and 12). Among different values of the number of bins, the combining of standard intensity histograms reaches the highest accuracy of 86.43% when the number of bins is 64. However, as can be seen in Figure 12, the multi-WHoID, which is the concatenation of single-phase WHoIDs (defined as Eq. (8)), provides a significantly higher accuracy than the aggregation of standard intensity histograms for all considered values of the number of bins. Also, as can be seen in Figure 11, TPR, TNR and NPV provided by the multi-WHoID are significantly higher than that provided by intensity histograms. Lastly, the multi-WHoID provided a significantly higher PPV than the intensity histograms.

From the results, the multi-WHoID outperforms the combination of standard intensity histograms in distinguishing HCC and ICC, showing that emphasizing the pixels on lesion regions can improve the performance of discriminating the two liver cancers. Thus, the WHoI is appropriate to be used as feature descriptor.

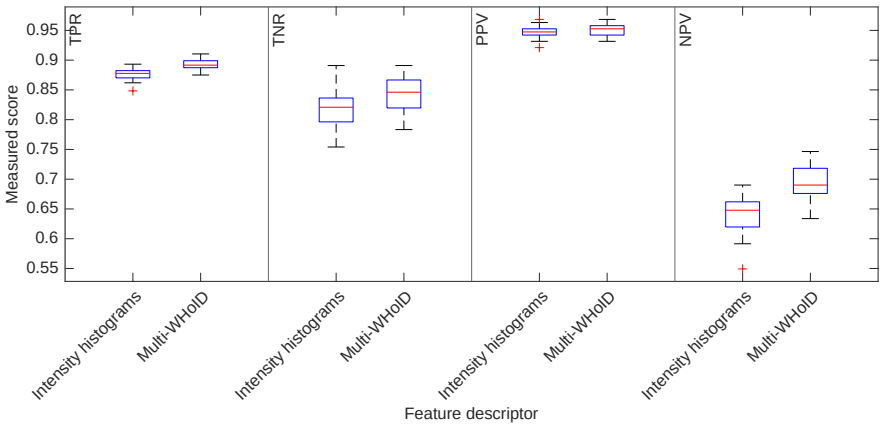

Fig. 11 The performance of Multi-WHoID descriptor compared to that of standard intensity histograms. The SVM with GHI kernel is used as the classifier. The results show True Positive Rate (TPR), True Negative Rate (TNR), Positive Predictive Value (PPV) and Negative Predictive Value (NPV) from multiple evaluations as box-plots.

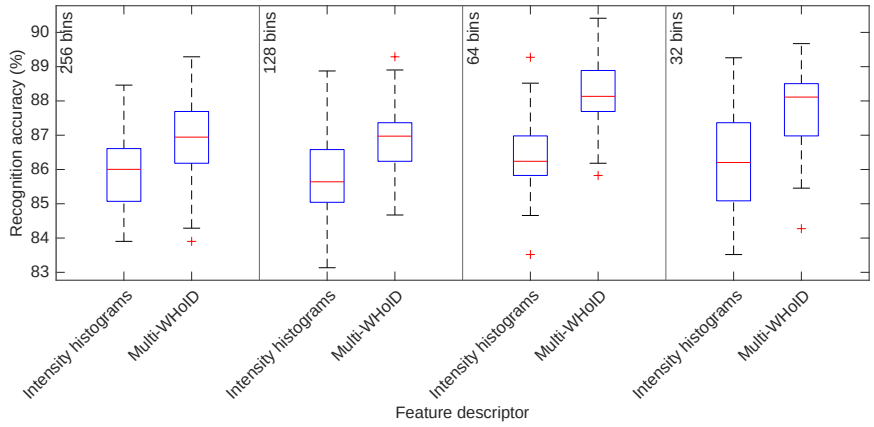

Fig. 12 Performance comparison between standard intensity histograms and Multi-WHoID with different numbers of bins.

4.6 Performance of proposed features with well-known classifiers

We adopted some well-known classifiers with the proposed features to discriminate HCC and ICC. The SVM with GHI kernel used as the classifier was replaced by k-nearest neighbors (kNN), decision tree, random forest and multi-layer perception (MLP). The parameter k of kNN in a set $\{1, 3, 5\}$ was evaluated. As well, the number of trees in the random forest was set to be 50 and 100. For the MLP, the architecture was designed following the suggestions in [16], which consists of one

hidden layer with 173 hidden nodes, and the activation function employed was the familiar hyperbolic tangent function.

The performance comparison of applying the Multi-WHoID with kNNs, decision tree, random forests and MLP and SVM is given in Figure 13. The highest recognition accuracy is achieved by the SVM with GHI kerneli, which is significantly greater than the one provided by other comparing classifiers at 0.05 significance level. The MLP and random forests gave accuracies higher than 80% while the kNNs and decision tree gave lower accuracies. Although selecting a classification model as the classifier depends on empirical results, for suggestion, SVM, random forest and MLP should be firstly considered to be used with the proposed features.

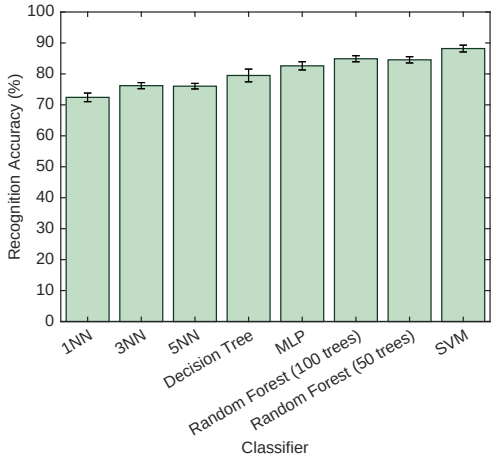

Fig. 13 The performance of Multi-WHoID with eight classifiers. The results show mean recognition accuracies (%) with standard deviation. The highest accuracy is significantly higher that the others at 0.01 significant level.

4.7 Error analysis

To find the causes of the misclassifications, we looked into the images that were falsely classified for more than 15 out of 30 cross-validation trials, 75% of which consist of ICC samples misclassified as HCC. After closely inspecting the segmentation and feature maps on these images, we found that they fall into one of the following three categories.

1. *Segmentation errors.* Since U-Net was trained on a dataset that only contains non-contrast and delayed phase, the model does not perform as well on images in the other two phases. After examining the misclassified images, we found that the lesions might be excluded from the segmentation when the lesions are very close to the liver's boundary. Two examples of such cases are shown in Figure 14, where the ICC lesions are indicated by the arrows. Since the lesions

are not only excluded from the segmentation but also the subsequent feature analysis, the images will most likely be misclassified.

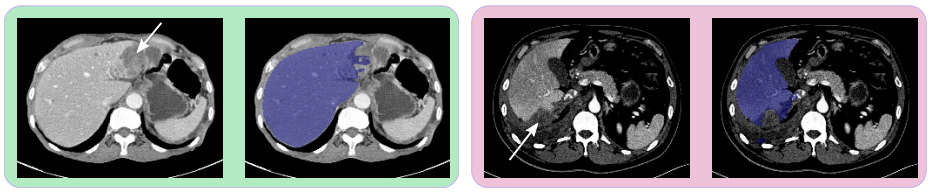

Fig. 14 Two examples of segmentation errors caused by the ICC lesion being close to the liver's boundary. The lesion areas are indicated by the arrows. **Left**: an image in portal venous phase. **Right**: an image in arterial phase.

2. *Unusual brightness in arterial phase.* This is caused by human errors during the image acquisition, either from incorrect settings in the viewer or the JPEG conversion software. Examples of such cases are shown in Figure 15. Notice that the feature maps of the arterial phase are covering the whole area. This is because most of the liver have pixel values greater than the arterial-phase threshold t_A that we set in subsection 3.2. Consequently, most of the images were retained from the binarization, causing errors in subsequent analyses.

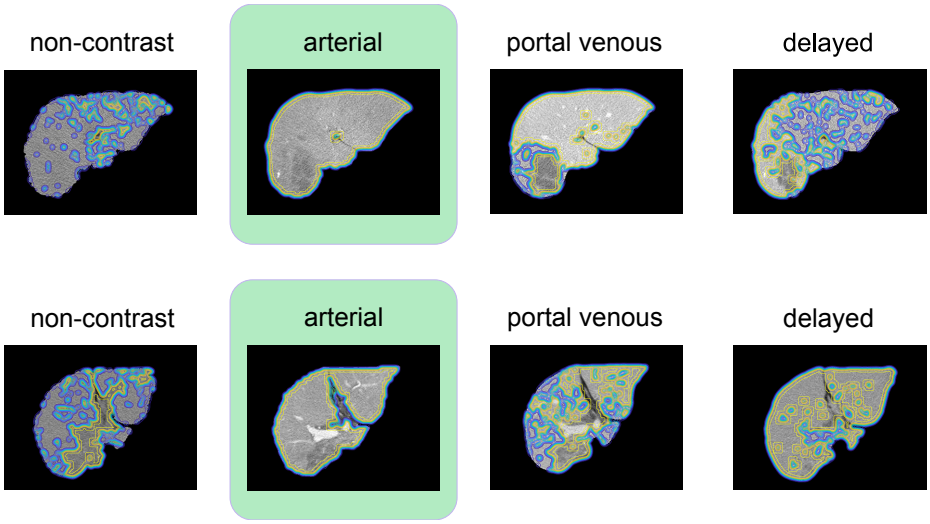

Fig. 15 Two examples of images with unusual brightness in the arterial phase. This causes the feature maps to cover the whole liver area.

3. *Barely visible lesions.* Small lesions are very hard to detect in all contrast-enhanced phases. An example is shown in Figure 16, where the ICC lesion is a small blurry dot in the center (two other dots are cysts) that is barely visible in the arterial and portal venous phase and hardly visible in non-contrast and

delayed phase. We can see that the feature maps capture the lesion in the non-contrast and arterial phase, but not in the portal venous and delayed phase.

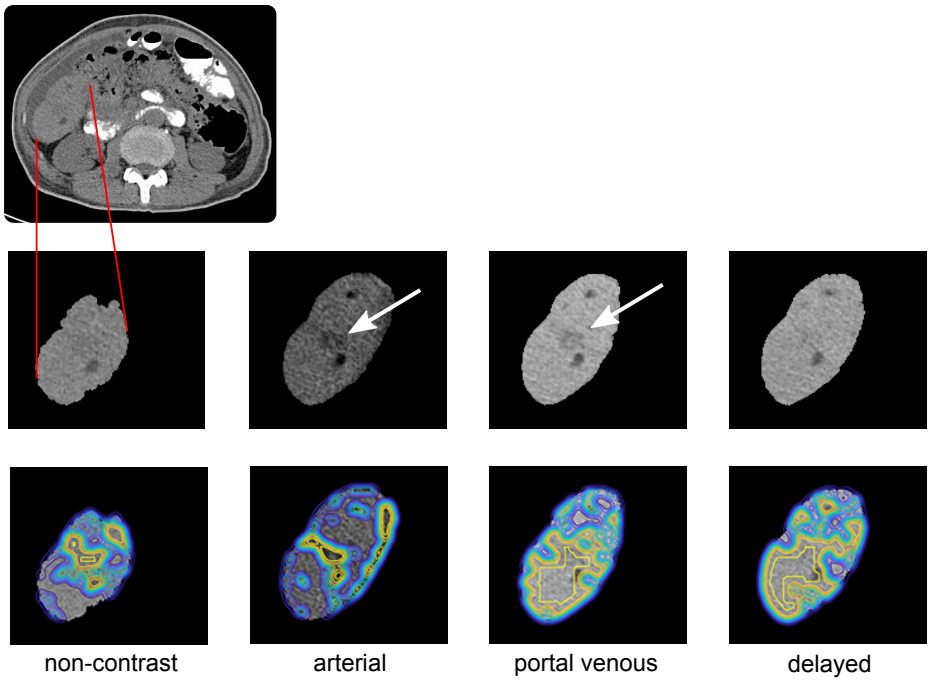

Fig. 16 An example of misclassifications caused by a small and blurry ICC lesion. The contrast-enhanced lesion is indicated by the arrows, and two other small dots are cysts.

5 Discussion

In this study, we designed several experiments in order to determine the best settings for our two-step workflow. First, we focused on liver segmentation. The results show that U-Net can be improved by (1) adding an instant segmentation layer after every convolutional layer (2) using a combination between the DICE loss and the weighted cross-entropy as our loss function and (3) post-processing the segmentation with a conditional random field. This is because the differences in contrast enhancement between the non-contrast/delayed phase images in the segmentation dataset and arterial-portal venous phase images in the CMU HCC-ICC dataset.

The segmented livers were then passed to the next experiments, in which we tried to find suitable features and machine learning models for the HCC-ICC classification. The results show that the Multi-WHoID, built from 64-bin color intensity distributions, is the best feature descriptor, and the SVM with the GHI kernel is the best classifier.

The only previous work that studies the same problem as ours is [26], where only images from portal venous phase were used. In this case, the tumor regions were semi-automatically segmented under the supervision of radiologists and given to SVM for classification between HCC and ICC. Compared to their method which gives 69.70% test accuracy, our method which utilizes multi-phase images is able to get higher test accuracy of 88%.

One of the advantages of our method is that it does not require as many labeled images as a standard CNN model, since all of the heavy lifting is done by training U-Net on the public data, and the classification is done by an SVM which is not as data-hungry as deep learning models. This is beneficial in the field of medical imaging where acquiring and labeling data are time-consuming. Compared to [37], in which a standard CNN model for liver mass was trained with a much larger sample of 1068 sets of images and gave 84% test accuracy, we only used 257 sets and obtained 88% accuracy.

Another advantage of our method over pure deep learning methods is the interpretability of the discriminative map, as shown in Figure 8. By automatically highlighting highly discriminative regions, our method can be used to assist radiologists for fast lesion tracking and help them make final diagnosis.

Our method can be applied to other lesion classification tasks, such as HCC and cirrhosis, HCC and cysts or HCC and normal liver [28, 33, 37]. It is also applicable to related problems that require some texture analysis. For example, in the problem of liver graft hepatic steatosis assessment for liver transplant, our multi-grid approach to feature extraction can be applied to liver images from RGB cameras [27]. Alternatively, we can apply our entire workflow to multi-phase CT scans of livers for automatic assessments [24]. Another task that utilizes multi-phase liver images is the radioembolization therapy (RE), a locoregional therapy for advanced-stage liver cancer, where the contrast-enhanced images are taken pre- and post-treatment [1, 2, 3, 15, 34], on which our method can be applied to evaluate the patient response to RE.

However, there can be some misclassifications due to various issues. First, segmentation errors from the differences in contrast enhancement between the images in the segmentation dataset and the CMU HCC-ICC dataset. This is because our method of segmenting high-contrast images relies heavily on histogram matching to "blur out" lesions, which does not always work on large lesions.Alternatively, we could make U-Net adaptive to lesions by adding manually segmented data in high-contrast phases to the training set. However, this approach can be quite costly, both in increased training time and manual labors; to find the optimal number of manually segmented images to be put in the training set, some experiments are required.

The second issue is the unusual brightness in the arterial phase. This issue can be fixed by obtaining more training data so that our threshold can be adjusted accordingly. Alternatively, we can try to develop a new feature descriptor that is invariant to brightness of the liver and retains the color intensity of the lesions.

Lastly, the images might be misclassified if the lesions are too small to be detected from our feature descriptor. We believe that the nearby cysts might be responsible since they are detected in Figure 16 in some of the phases. Hence, the first step toward solving this issue is to "teach" our classification model to differentiate between the cysts, HCC and ICC. If we follow this path, the feature descriptor or the SVM model has to be modified in some way to achieve such goal.

Another weakness of our method is that the model cannot handle a rare case where an image contains both HCC lesions and ICC lesions. And it would be fruitful to develop a model that can deal with such case. All in all, we think that these possible improvements pose interesting but challenging problems that we will look into in future studies.

6 Conclusion

In this paper, we propose an automatic two-step method of classifying two types of liver cancer, namely HCC and ICC, from multi-phase abdominal CT scans. In the first step, we segment the liver using a deep learning model U-Net with instance segmentation and CRF post-processing. In the second step, we extract the features based on pixel intensities of lesions, which differs across multi-phase images between these two types of cancer. The features are then fed into a SVM classifiers for the final prediction. As a result, we achieved well recognition accuracy of over 88% in the HCC-ICC classification from CT images of the liver on CMU HCC-ICC dataset. We believe that the proposed method has a great potential for classification of liver cancers, which could become radiologists' essential assist tool for better planning of prevention and treatment strategies. In addition, we believe that the integrated program will help the radiologists in varying levels of expertise in diagnosis and classification of liver cancer.

Acknowledgements This study was funded by Faculty of Medicine, Chiang Mai University, Thailand. We would like to thank Chiang Mai University, Thailand, for financial support and computing resources.

References

1. Adcock CS, Florez E, Zand KA, Patel A, Howard CM, Fatemi A (2018) Assessment of treatment response following yttrium-90 transarterial radioembolization of liver malignancies. Cureus DOI 10.7759/cureus.2895, URL `https://doi.org/10.7759/cureus.2895`
2. Alsultan AA, Smits MLJ, Barentsz MW, Braat AJAT, Lam MGEH (2019) The value of yttrium-90 PET/CT after hepatic radioembolization: a pictorial essay. Clinical and Translational Imaging 7(4):303–312, DOI 10.1007/s40336-019-00335-2, URL `https://doi.org/10.1007/s40336-019-00335-2`
3. Altenbernd J, Wetter A, Forsting M, Umutlu L (2016) Treatment response after radioembolisation in patients with hepatocellular carcinoma—an evaluation with dual energy computed-tomography. European Journal of Radiology Open 3:230–235, DOI 10.1016/j.ejro.2016.08.002, URL `https://doi.org/10.1016/j.ejro.2016.08.002`
4. Azer SA (2019) Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: A systematic review. World journal of gastrointestinal oncology 11(12):1218–1230, DOI 10.4251/wjgo.v11.i12.1218, URL `https://pubmed.ncbi.nlm.nih.gov/31908726`
5. Berre CL, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smal-Tabbone M, Danese S, Peyrin-Biroulet L (2020) Application of artificial in-

telligence to gastroenterology and hepatology. Gastroenterology 158(1):76–94.e2, DOI https://doi.org/10.1053/j.gastro.2019.08.058, URL http://www.sciencedirect.com/science/article/pii/S0016508519414121

6. Bilic P, Christ PF, Vorontsov E, Chlebus G, Chen H, Dou Q, Fu CW, Han X, Heng PA, Hesser J (2019) The Liver Tumor Segmentation Benchmark (LiTS). arXiv e-prints arXiv:1901.04056, 1901.04056
7. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 68(6):394–424, DOI 10.3322/caac.21492
8. Bruix J, Sherman M (2011) Management of hepatocellular carcinoma: An update. Hepatology 53(3):1020–1022, DOI 10.1002/hep.24199
9. Colli A, Fraquelli M, Casazza G, Massironi S, Colucci A, Conte D, Duca P (2006) Accuracy of ultrasonography, spiral ct, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: A systematic review: Cme. American Journal of Gastroenterology 101(3):513–523, DOI 10.1111/j.1572-0241.2006.00467.x
10. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20(3):273–297, DOI 10.1007/BF00994018
11. DeOliveira ML, Cunningham SC, Cameron JL, Kamangar F, Winter JM, Lillemoe KD, Choti MA, Yeo CJ, Schulick RD (2007) Cholangiocarcinoma: Thirty-one-year experience with 564 patients at a single institution. Annals of Surgery 245(5)
12. European Association For The Study Of The Liver EOFRTOC (2012) Easl-eortc clinical practice guidelines: Management of hepatocellular carcinoma. Journal of Hepatology 56(4):908–943, DOI 10.1016/j.jhep.2011.12.001
13. Forner A, Vilana R, Ayuso C, Bianchi L, Sol M, Ayuso JR, Boix L, Sala M, Varela M, Llovet JM, Br C, Bruix J (2008) Diagnosis of hepatic nodules 20 mm or smaller in cirrhosis: Prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. Hepatology 47(1):97–104, DOI 10.1002/hep.21966
14. ircad France (2009) 3dircadb. URL https://www.ircad.fr/research/3dircadb, Accessed on 06.21.2019
15. Gensure RH, Foran DJ, Lee VM, Gendel VM, Jabbour SK, Carpizo DR, Nosher JL, Yang L (2012) Evaluation of hepatic tumor response to yttrium-90 radioembolization therapy using texture signatures generated from contrast-enhanced CT images. Academic Radiology 19(10):1201–1207, DOI 10.1016/j.acra.2012.04.015, URL https://doi.org/10.1016/j.acra.2012.04.015
16. Heaton J (2008) Introduction to Neural Networks for Java, 2nd Edition, 2nd edn. Heaton Research, Inc.
17. Heimann T, van Ginneken B, Styner MA, Arzhaeva Y, Aurich V, Bauer C, Beck A, Becker C, Beichel R, Bekes G, Bello F, Binnig G, Bischof H, Bornik A, Cashman PMM, Chi Y, Cordova A, Dawant BM, Fidrich M, Furst JD, Furukawa D, Grenacher L, Hornegger J, Kainmller D, Kitney RI, Kobatake H, Lamecker H, Lange T, Lee J, Lennon B, Li R, Li S, Meinzer H, Nemeth G, Raicu DS, Rau A, van Rikxoort EM, Rousson M, Rusko L, Saddi KA, Schmidt G, Seghers D, Shimizu A, Slagmolen P, Sorantin E, Soza G, Susomboon R, Waite JM, Wimmer A, Wolf I (2009) Comparison and evaluation of methods for liver segmentation from ct datasets. IEEE Transactions on Medical Imaging

28(8):1251–1265
18. Imsamran W, Chaiwerawattana A, Wiangnon S, Suwanrungruang K, Sangrajrang S, Buasom R (2018) Cancer in Thailand, vol IX. National Cancer Institute, Thailand
19. Jemal A, Center MM, DeSantis C, Ward EM (2010) Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiology and Prevention Biomarkers 19(8):1893–1907, DOI 10.1158/1055-9965.EPI-10-0437, URL `https://cebp.aacrjournals.org/content/19/8/1893`, `https://cebp.aacrjournals.org/content/19/8/1893.full.pdf`
20. Kaewpitoon N, Kaewpitoon SJ, Pengsaa P (2008) Opisthorchiasis in thailand: review and current status. World journal of gastroenterology 14(15):2297–2302, DOI 10.3748/wjg.14.2297
21. Kobayashi M, Ikeda K, Saitoh S, Suzuki F, Tsubota A, Suzuki Y, Arase Y, Murashima N, Chayama K, Kumada H (2000) Incidence of primary cholangiocellular carcinoma of the liver in japanese patients with hepatitis c virusrelated cirrhosis. Cancer 88(11):2471–2477, DOI 10.1002/1097-0142(20000601)88:11⟨2471::AID-CNCR7⟩3.0.CO;2-T
22. Krähenbühl P, Koltun V (2011) Efficient inference in fully connected crfs with gaussian edge potentials. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ (eds) Advances in Neural Information Processing Systems 24, Curran Associates, Inc., pp 109–117
23. Lalkhen AG, McCluskey A (2008) Clinical tests: sensitivity and specificity. BJA Education 8(6):221–223, DOI 10.1093/bjaceaccp/mkn041
24. Lawrence DA, Oliva IB, Israel GM (2012) Detection of hepatic steatosis on contrast-enhanced CT images: Diagnostic accuracy of identification of areas of presumed focal fatty sparing. American Journal of Roentgenology 199(1):44–47, DOI 10.2214/ajr.11.7838
25. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, Roberts LR, Heimbach JK (2018) Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the american association for the study of liver diseases. Hepatology 68(2):723–750, DOI 10.1002/hep.29913
26. Midya A, Chakraborty J, MD LMP, MD JZ, MD WRJ, MD RKGD, Simpson AL (2018) Deep convolutional neural network for the classification of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. In: Petrick N, Mori K (eds) Medical Imaging 2018: Computer-Aided Diagnosis, International Society for Optics and Photonics, SPIE, vol 10575, pp 501 – 506, DOI 10.1117/12.2293683
27. Moccia S, Mattos LS, Patrini I, Ruperti M, Poté N, Dondero F, Cauchy F, Sepulveda A, Soubrane O, Momi ED, Diaspro A, Cesaretti M (2018) Computer-assisted liver graft steatosis assessment via learning-based texture analysis. International Journal of Computer Assisted Radiology and Surgery 13(9):1357–1367, DOI 10.1007/s11548-018-1787-6
28. Nayak A, Baidya Kayal E, Arya M, Culli J, Krishan S, Agarwal S, Mehndiratta A (2019) Computer-aided diagnosis of cirrhosis and hepatocellular carcinoma using multi-phase abdomen ct. International Journal of Computer Assisted Radiology and Surgery 14(8):1341–1352, DOI 10.1007/s11548-019-01991-5
29. Ohishi W, Fujiwara S, Cologne JB, Suzuki G, Akahoshi M, Nishi N, Tsuge M, Chayama K (2011) Impact of radiation and hepatitis virus infection on risk of hepatocellular carcinoma. Hepatology 53(4):1237–1245, DOI 10.1002/hep.

24207
30. Rimola J, Forner A, Reig M, Vilana R, de Lope CR, Ayuso C, Bruix J (2009) Cholangiocarcinoma in cirrhosis: Absence of contrast washout in delayed phases by magnetic resonance imaging avoids misdiagnosis of hepatocellular carcinoma. Hepatology 50(3):791–798, DOI 10.1002/hep.23071
31. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, pp 234–241
32. Sangiovanni A, Manini MA, Iavarone M, Romeo R, Forzenigo LV, Fraquelli M, Massironi S, Della Corte C, Ronchi G, Rumi MG, Biondetti P, Colombo M (2010) The diagnostic and economic impact of contrast imaging techniques in the diagnosis of small hepatocellular carcinoma in cirrhosis. Gut 59(5):638–644, DOI 10.1136/gut.2009.187286
33. Sato M, Morimoto K, Kajihara S, Tateishi R, Shiina S, Koike K, Yatomi Y (2019) Machine-learning approach for the development of a novel predictive model for the diagnosis of hepatocellular carcinoma. Scientific Reports 9(1):7704, DOI 10.1038/s41598-019-44022-8
34. Spina JC, Hume I, Pelaez A, Peralta O, Quadrelli M, Monaco RG (2019) Expected and unexpected imaging findings after 90y transarterial radioembolization for liver tumors. RadioGraphics 39(2):578–595, DOI 10.1148/rg.2019180095, URL https://doi.org/10.1148/rg.2019180095
35. Tomimatsu M, Ishiguro N, Taniai M, Okuda H, Saito A, Obata H, Yamamoto M, Takasaki K, Nakano M (1993) Hepatitis c virus antibody in patients with primary liver cancer (hepatocellular carcinoma, cholangiocarcinoma, and combined hepatocellular-cholangiocarcinoma) in japan. Cancer 72(3):683–688, DOI 10.1002/1097-0142(19930801)72:3⟨683::AID-CNCR2820720310⟩3.0.CO;2-C
36. Welzel TM, Graubard BI, ElSerag HB, Shaib YH, Hsing AW, Davila JA, McGlynn KA (2007) Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the united states: A population-based case-control study. Clinical Gastroenterology and Hepatology 5(10):1221 – 1228, DOI https://doi.org/10.1016/j.cgh.2007.05.020
37. Yasaka K, Akai H, Abe O, Kiryu S (2018) Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced ct: A preliminary study. Radiology 286(3):887–896, DOI 10.1148/radiol.2017170706