IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 12, 2020, accepted July 7, 2020, date of publication July 22, 2020, date of current version August 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011145

Deep Fusion Models of Multi-Phase CT and Selected Clinical Data for Preoperative Prediction of Early Recurrence in Hepatocellular Carcinoma

WEIBIN WANG[1], QINGQING CHEN[2], YUTARO IWAMOTO[1], (Member, IEEE), PANYANAT AONPONG[1], LANFEN LIN[3], (Member, IEEE), HONGJIE HU[2], QIAOWEI ZHANG[2], AND YEN-WEI CHEN[1,4], (Member, IEEE)

[1]Graduate School of Information Science and Engineering, Ritsumeikan University, Kusatsu 525-8577, Japan
[2]Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou 310000, China
[3]College of Computer Science and Technology, Zhejiang University, Hangzhou 310000, China
[4]Research Center for Healthcare Data Science, Zhijiang Laboratory, Hangzhou 310000, China

Corresponding authors: Lanfen Lin (llf@zju.edu.cn), Hongjie Hu (hongjiehu@zju.edu.cn), and Yen-Wei Chen (chen@is.ritsumei.ac.jp)

This work was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry of Education, Science, Culture, and Sports (MEXT), under Grant 18H03267 and 20K21821; in part by the Zhejiang Laboratory Program under Grant 2018DG0ZX01; in part by the Key Research and Development Plan of Zhejiang Province under Grant 2019C03064; and in part by the Program Co-sponsored by Province and Ministry under the Grant No.WKJ-ZJ-1926.

ABSTRACT Post-operative early recurrence (ER) of hepatocellular carcinoma (HCC) is one of the leading causes of death. The prediction of the ER of HCC before treatment contributes to guiding treatment and follow-up protocols. In recent years, CT radiomics signatures have been proven effective in several studies in predicting early recurrence of HCC, there are still two major challenges. First, the radiomics features extracted were low or mid-level features, which may not fully characterize HCC heterogeneity. Second, the fusion approach of clinical textual data and image information is in little consensus. In this paper, we proposed a deep-learning based prediction model to extract high-level features from the triple-phase CT images and compare its performance with traditional radiomics model and clinical model. The accuracy and area under the curve (AUC) of receiver operating characteristics of three models was 69.52%/0.723, 67.04%/0.64, 76.03%/0.75, respectively. In addition, we proposed four fusion models to combine clinical data and high-level features. Among them, Fusion model D performed best, achieving a higher prediction accuracy of 78.66% and AUC of 0.8248. Moreover, fusion models with a joint loss function can further improve the prediction performance to 80.49% and 0.8331.

INDEX TERMS Hepatocellular carcinoma, early recurrence, deep learning, multi-phase CT images, clinical data, fusion model.

I. INTRODUCTION

Hepatocellular carcinoma (HCC) is the sixth most commonly diagnosed malignancy and the third leading cause of cancer-related mortality worldwide, especially in the Asia-Pacific region [1], [2]. The mainstay treatments of HCC include surgical resection, radiofrequency ablation (RFA), transarterial chemoembolization (TACE), and liver transplantation. For patients with well-preserved liver function, hepatic resection is the first-line treatment strategy [3], [4]. Moreover, surgical resection is actively considered even when the patient is diagnosed with vascular tumor thrombosis or extrahepatic metastases in many experienced hepatobiliary surgical centers [5].

However, postoperative recurrence either intrahepatic or extrahepatic is still the major cause of patient death [6]. The peak time of HCC recurrence is one year after resection, which is defined as ''early recurrence'' (ER) [7]. Previous studies have reported that some pathological indicators such as worse histological differentiation, the presence of microvascular invasion (MVI) and microsatellite nodules were significant factors for predicting ER [8]. But all of them could be only determined after surgery. As such, researcher effort to identify high-risk ER individuals before treatment.

The associate editor coordinating the review of this manuscript and approving it for publication was Ting Li.

For whom, a more aggressive treatment like liver transplantation or expanding resection margin as well as adjuvant therapy or closer follow-up after surgery.

Imaging examination, represented by computed tomography (CT) and Magnetic Resonance (MR) Imaging, plays an essential role in HCC detection, diagnosis and treatment. Additionally, many scholars suggested that qualitative imaging features such as multiple tumors, corona enhancement, irregular tumor margin could predict ER of HCC to some extent [9]–[11]. But those factors remain quite subjective and difficult to explain complicated intratumor heterogeneity by naked eyes.

In 2012, Lambin first proposed the concept of radiomics, believing that images are not only pictures but also data [12]. In 2016, Radiology published an authoritative review of radiomics to re-emphasize the core idea of radiomics as converting digital images into mineable data and pointed out that such conversion would become the norm for future radiological image analysis [13]. Elsewhere, a few studies have demonstrated that radiomics features can be used as prognostic imaging biomarkers for HCC recurrence after surgery [14]–[18].

Zhou *et al.* extracted 300 hand-crafted low-level radiomics features from CT arterial- and portal venous phase, and 21 optimal radiomics signatures was selected using LASSO regression method to predict ER of HCC [19]. Ning *et al.* also developed a CT-based model for predicting the ER of HCC [20]. Both of the studies demonstrated that the radiomics signature is a potential biomarker for the ER prediction of HCC. The authors also found that integrating radiomics features and relevant clinical data can effectively improve the performance of prediction model. However, the image features extracted by these authors were low- or mid-level features, which was limited to fully characterize the potential information associated with early recurrence.

In recent years, among many methods, the deep convolutional neural networks (DCNN) structure have achieved the highest classification accuracy on the ImageNet dataset. The high-level feature extracted by DCNN has proven to be superior to explicitly designed low-level and mid-level features [21]–[23]. Several studies have applied DCNN to medical images of liver. Here, Bi *et al.* applied ResNet to the segmentation of liver lesions and won fourth place in the ISBI 2017 Liver Tumor Segmentation Challenge [24], while Liang *et al.* proposed a DCNN with local and global pathways for focal liver lesion classification [25], [26]. These studies demonstrated the effectiveness of DCNN in the field of liver image analysis. Moreover, Deep learning-based radiomics (DLR) model has been illustrated to perform better traditional radiomics model in noninvasive prediction of IDH1 mutation for low grade glioma and predicting survival of glioblastoma multiforme [27], [28]. However, to date, there is no study focus on early recurrence of HCC using deep learning method.

Thus, the primary aim of our study was to propose a deep learning-based model for predicting ER of HCC and compare the performance with conventional radiomics model and clinical model. The second aim was to develop four fusion models combining clinical data and deep features to explore the best fusion approach for two heterogeneous data. The contributions of our study are as follows: (1) in contrast to previous works, our proposed method identified that high-level features can better characterize ER of HCC; (2) we designed a DCNN model fusing CT images and clinical data together to make full use of two different types of data; and (3), we used a joint loss function to further improve the prediction performance.

II. MATERIALS

This study was approved by Zhejiang University, Ritsumeikan University, and Sir Run Run Shaw Hospital. The medical images and clinical data used in this study were collected from Sir Run Run Shaw Hospital. Initially, 331 consecutive HCC patients underwent curative partial hepatectomy from 2012 to 2016 were enrolled in this retrospective study. Inclusion criteria were as follows: (1) surgically proven HCC; (2) multi-phase dynamic CT performed within 1 month before surgery; (3) followed up at least one year after surgery; (4) no preoperative HCC treatment history; and (5) negative surgical margin. Ultimately, a cumulative total of 167 HCC patients (140 males and 27 females) were included in the study. Early recurrence was identified in 65(38.9%) patients, while the remaining 102(61.1%) patients did not recur (non-ER). The patients were thus divided into an ER group and a NER group.

A. CLINICAL DATA

Numerous literatures have discussed the prognostic indicators in the recurrence of HCC. For example, Portolani *et al.* showed that chronic active hepatitis such as hepatitis C virus (HCV) infection, and tumor vascular invasion were associated with ER [29], while Chang *et al.* recommended that the patient age of 60 as a cut-off value between ER and NER [30]. Okamura *et al.* [31] found preoperative neutral lymphatic ratio (N/L ratio), an indicator of inflammation, was related to disease-free survival (DFS) and the overall survival (OS) of HCC patients. Compared with the patients NLR$\geq$2.81, the outcome was significantly better in NLR$<$2.81 group. Along with NLR, the general clinical indicators for prognosis include age, gender, tumor size, tumor number, hepatitis B virus (HBV) infection, portal vein invasion, alanine aminotransferase (ALT), alkaline phosphatase (AKP), glutamic-pyruvic transaminase (AST), the Barcelona-clinic liver cancer (BCLC) stage, liver cirrhosis, and alpha fetoprotein (AFP) [32].

The clinical factors collected in our study was presented in Table 1, including sex (male or female), age ($<$60y or $\geq$60y) tumor size ($<$5; $\geq$5cm), tumor number(single or multiple); portal vein invasion (yes/no), liver cirrhosis (yes/no), HBV infection (yes/ no), NLR ($<$2.81or $\geq$2.81), ALT ($<$40 or $\geq$40U/L), AST ($<$50 or $\geq$50U/L), AKP ($<$125 or $\geq$125U/L), GGT ($<$45 or $\geq$45U/L), ALB ($\geq$40 or

<40U/L), TB (<20.5 or ≥20.5U/L), AFP (<9 or ≥9μg/L), CP Level (A or B), BCLC stage (0,A,B,C). Clinical data were evaluated through Chi-squared test which is a well-known method used to estimate the dependency between categorical variables [33], [34]. P value less than 0.05 was considered as significant differences. Seven clinical factors including age, tumor size, portal vein invasion, N/L ratio, TB, AFP, and BCLC stage were selected and further represented by a binary vector with nine elements of [c1, c2,c3, …, c8, c9], where [c1] represents age ([0]: <60, [1]: ≥60); [c2] represents tumor size ([0]: <5cm, [1]: ≥5cm); [c3] represents portal vein invasion ([0]: no, [1]: yes); [c4] represents N/L ratio ([0]: <2.8, [1]: ≥2.8); [c5] represents TB ([0]: <20.5, [1]: ≥20.5); [c6] represents AFP ([0]: <9, [1]: ≥9);[c7, c8, c9] represents BCLC stage ([0,0,0]: 0, [0,0,1]: A, [0,1,0]: B, [1,0,0]: C).

B. MULTI-PHASE CT IMAGES AND IMAGE PREPROCESSING

Contrast-enhanced CT has been widely used in noninvasive diagnosis of focal liver lesions. All CT images were obtained with Siemens SOMATON Definition AS scanner (Siemens Healthcare, Forchheim, Germany) and GE LightSpeed VCT scanner (GE Medical systems, Milwaukee, WI). Non-contrast enhanced (NC) phase was routinely performed at first. After injecting contrast agent, arterial (ART) phase, portal venous (PV) phase, and delay (DL) phase was scanned at 30s, 60s, 180s respectively). In this study, NC, ART, and PV phases were selected for the following experiment. The pixels size of these three phases is 512 × 512, and the thickness of each slice is 5 or 7 mm.

One abdominal radiologist with three years of experience manually marked the bounding box of HCC via ITK-SNAP (version 3.6.0, University of Pennsyvania, Philadelphia, USA) [35], and the ROI were revised by a radiologist with six years of experience. Example images covering three phases (NC, ART, PV) of the HCC patient are shown in Figure 1.

FIGURE 1. Example images of HCC over three-phases. The tumor is marked with a red frame (ROI).

For the image preprocessing, after the ROI images were obtained, three-phase images were first registered according to the center point. Following this, we took linear interpolation method to resize each ROI image. Finally, we combined

TABLE 1. Clinical variables of patients with ER and NER.

Clinical Variables	Total n=167	NER n=102	ER n=65	P-value (Chi-Square Test)
Sex				0.826
female	27	17	10	
male	140	85	55	
Age				**0.018**
<60	102	55	47	
⩾60	65	47	18	
Tumor size (mm)				**<0.001**
<50	102	74	28	
⩾50	65	28	37	
Tumor number				0.152
single	155	97	58	
multiple	12	5	7	
Portal vein invasion				**0.001**
no	143	95	48	
yes	24	7	17	
Liver cirrhosis				0.794
no	52	31	21	
yes	115	71	44	
HBV Infection				0.054
no	33	25	8	
yes	134	77	57	
N/L ratio				**0.039**
<2.8	106	71	35	
⩾2.8	61	31	30	
ALT (U/L)				0.417
<40	104	66	38	
⩾40	63	36	27	
AST (U/L)				0.500
<50	123	77	46	
⩾50	44	25	19	
AKP (U/L)				0.054
<125	131	85	46	
⩾125	36	17	19	
GGT (U/L)				0.065
<45	66	46	20	
⩾45	101	56	45	
ALB (g/L)				0.256
≥40	165	100	65	
<40	2	2	0	
TB (μmol/L)				**0.020**
<20.5	122	81	41	
⩾20.5	45	21	24	
AFP (μg/L)				**0.003**
<9	50	39	11	
⩾9	117	63	54	
CP Level				0.132
A	140	89	51	
B	27	13	14	
BCLC stage				**<0.001**
0	22	17	5	
A	117	78	39	
B	4	1	3	
C	24	6	18	

*P<0.05 means significant difference.

the resized images into a multi-channel image. Image preprocessing process is shown in Figure 2.

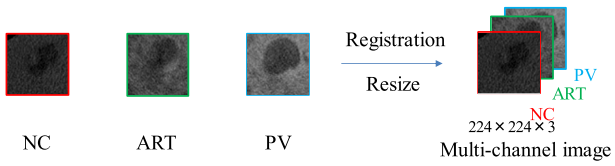


FIGURE 2. Image preprocessing.

III. METHODS

In our study, we built a deep learning model and four kinds of fusion models. The deep learning models only used multi-phase CT image information, while the fusion models used both multi-phase CT imaging and clinical data for the ER prediction.

A. DEEP RADIOMICS MODELS

In recent years, DCNN has been proven to be superior in the field of image classification. Generally, increasing the number of convolutional layers will enable the network to extract higher-level features. For example, VGG-Net [36] and GoogleNet [22] improve the classification accuracy of images by deepening the network. Since the gradient of DCNN is back-propagate, after increasing the depth of the network, the gradient of the upper layer will become smaller, leading the gradient to disappear. Residual network (ResNet) uses a shortcut connection [23], as shown in Figure 3.

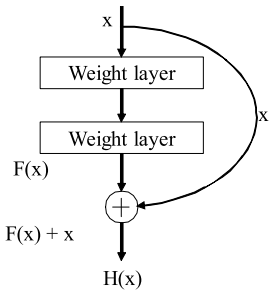


FIGURE 3. Shortcut connection.

Here, x represents the input, $\mathrm{H}(x)$ represents the output, then, the residual knowledge of the shortcut connection is $\mathrm{F}(x) = \mathrm{H}(x) - x$. The network learning residual knowledge $\mathrm{F}(x)$ is simpler than the direct learning output $\mathrm{H}(x)$, meaning it effectively resolves the problem of network degradation.

We built a deep learning model based on the structure of ResNet. This model consists of 49 convolutional layers and a fully connected layer. The convolutional layers extracted the high-level features of the image, and the fully connected layer performed the features selection and classification. Unlike conventional radiomics model, the deep learning model involves an end-to-end analysis from image input to prediction output.

However, due to the deepening of the network and the relatively small-scale data set, it became difficult to train the parameters in DCNN effectively. To reduce or alleviate overfitting of network, we used a fine-tuning approach. The basic principle of fine-tuning is that before training the network with the target dataset, the network is first trained in a large dataset (ImageNet) to extract relatively accurate shallow features. Various studies have demonstrated the effectiveness of fine-tuning for medical imaging [37], [38].

B. FUSION MODELS

While both CT images and clinical data are valuable for predicting ER of HCC, data conversion are needed before combining different types of data. In addition, we proposed several fusion models combining these two types of data aforementioned to make full use of the information and compare the performance between different models.

1) FUSION MODEL A

Fusion model A contained two pathways: a CT image pathway and a clinical data pathway. The image pathway consisted of one convolution layer, one pooling layer, and 16 residual blocks. The high-level features of the CT images were extracted through the image pathway. After the high-level features were flattened, images features and the clinical data were concatenated together to form a mixed-feature vector. Finally, the mixed-feature vector was classified by the fully connected layer. The network structure wasshown in Figure 4(a).

2) FUSION MODEL B

The number of image features extracted after the residual CNN was 2,048, while there were only 9 clinical features. To balance the huge gap of magnitude, we added a fully connected layer to the image pathway. The network structure is shown in Figure 4(b).

3) FUSION MODEL C

In fusion model A and B, only the image pathway was trained, while the clinical features are directly combined with the image features. In fusion model C, we added a convolutional layer to extract clinical information before concatenation. The network structure is shown in Figure 4(c).

4) FUSION MODEL D

Clinical data is textual data and is not suitable for convolution. We thus replaced the convolutional layer with the fully connected layer in model D. The network structure is

shown in Figure 4(d). In this case, both the image pathway and the clinical data pathway were given training weights.

C. LOSS FUNCTION OF FUSION MODELS

We used cross entropy as the loss function of fusion model D. Let N be the number of samples. $\mathbf{I}_j$ and $\mathbf{c}_j$ are the j-th CT image input data and clinical input data ($i = 1, 2, \ldots N$). We used b, k, and $\boldsymbol{W}$ to denote the bias term, number of

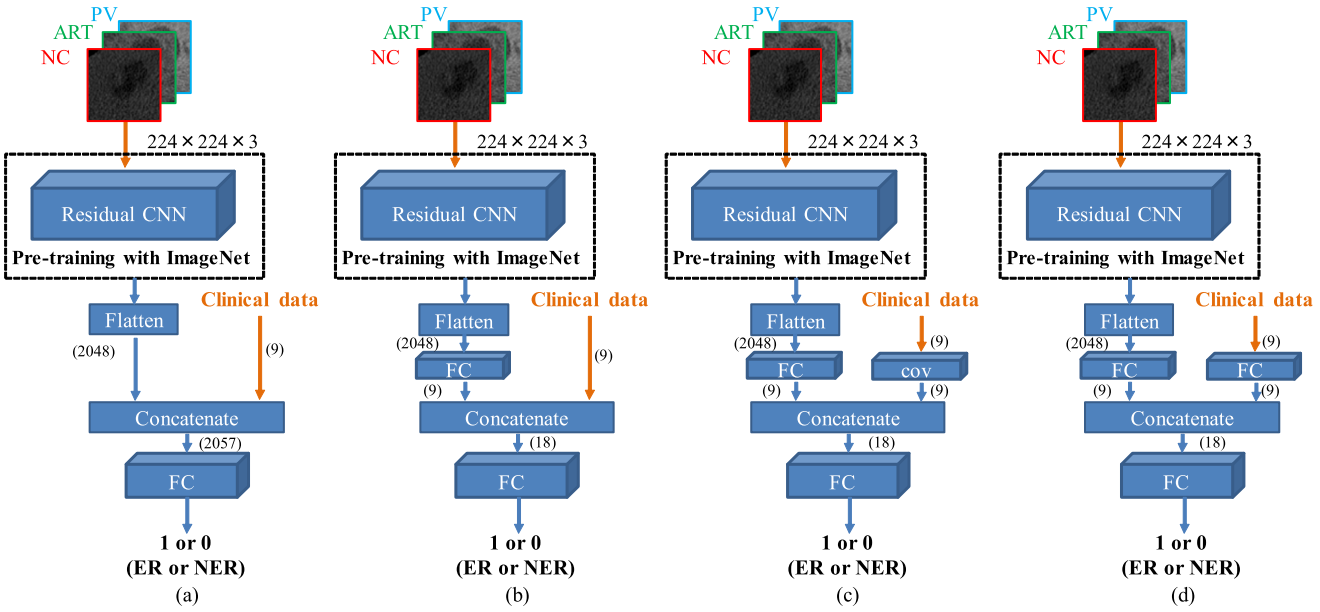


FIGURE 4. Four deep learning-based fusion models with clinical data and CT image data. (a) Model A, (b) model B, (c) model C, and (d) model D.

neurons, and weight of the last fully connected layer, respectively, while y_j denotes the label. $T(\mathbf{C}_j)$ and $S(\mathbf{I}_j)$ represent the output of clinical data training pathway and the image training pathway before the last fully connected layer; $\oplus$ denotes concatenate operation. The loss function is as follows:

$$\begin{aligned}\mathcal{L} &= Cross\ Entropy\left(S\left(\mathbf{I_j}\right) \oplus T\left(\mathbf{c_j}\right)\right) \\ &= -\frac{1}{N}\sum_{j=1}^{N}\left[y_j \ln p_{j1} + \left(1-y_j\right)\ln\left(p_{j2}\right)\right] \qquad (1)\end{aligned}$$

$$p_{jk} = \frac{e^{z_k}}{\Sigma_{k=1}^{2} e^{2k}} \qquad (2)$$

$$z_k = \boldsymbol{W}_k^T \cdot \left[S\left(\boldsymbol{I}_j\right) \oplus T\left(\boldsymbol{c}_j\right)\right] + b_k, k = 1, 2 \qquad (3)$$

D. JOINT LOSS FUNCTION

We combined the two types of data as input to the fusion models. However, we only calculated the loss after the combination of the final fully connected layer. In fusion models, we added a softmax layer to the image pathway and clinical data pathway before concatenation. With this, we could calculate the loss of both pathways.

Let p_{j1}^{I} represent the output possibility of the image pathway, p_{j1}^{c} represent the output possibility of the clinical data pathway, and p_{j1}^{con} represent the output possibility of the concatenate pathway. The joint loss is thus given by:

$$\begin{aligned}\mathcal{L} &= \frac{1}{4}\mathcal{L}_{image} + \frac{1}{4}\mathcal{L}_{clinical} + \frac{1}{2}\mathcal{L}_{fused} \\ &= \frac{1}{4}cross\ Entropy\left(S\left(\mathbf{I}_j\right)\right) + \frac{1}{4}Cross\ Entropy\left(T\left(\mathbf{I}_j\right)\right) \\ &\quad + \frac{1}{2}Cross\ Entropy\left(S\left(\mathbf{I}_j\right) \oplus T\left(\mathbf{c}_j\right)\right) \qquad (4)\end{aligned}$$

where $\mathcal{L}_{image}$ is a loss for the image pathway using image feature only, $\mathcal{L}_{clinical}$ is a loss for the clinical data pathway using clinical data only, $\mathcal{L}_{fused}$ is a loss for the fused features. Note that the single loss function (Eq. (1)) is the same as $\mathcal{L}_{fused}$.

IV. EXPERIMENTS

Our study was implemented in Python using the Tensorflow and Keras frameworks. All experiments were performed on a machine with the following specifications: Intel(R) Core (TM) i7-3930K CPU @ 3.20GHz 64-bit, GPU NVIDIA GeForce GTX 1080 Ti, 16GB of RAM, and the Windows 10 professional version 1830 system.

We used ten-fold cross validation to perform our experiments. We first randomly divided 167 patients into 10 groups, each group containing six or seven ERs and ten or eleven NERs. In the ten-fold cross-validation process, we chose 1 set of data as the test group, and the remaining 9 sets as the training samples. The average results of ten experiments is used as final score of the model.

The number of CT slices that containing tumor differed from patient to patient because of the varieties of tumor size and location. A total of 765 annotated slices were utilized in our experiments. Table 2 summarized the number of training images and test images (CT slice images) for each experiment.

In our experiments, the conventional radiomics model consists of the following three steps: First we extracted the radiomics features; Then we used LASSO for feature selection; Finally, we used random forest (RF) as the classifier. Different with the radiomics model, the deep learning model is an end-to-end structure. In the training stage, we first pre-trained our network with ImageNet. Next, we removed

TABLE 2. The distribution of the 10-fold cross-validation dataset.

Experiment	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Training	695 (150)	681 (150)	683 (150)	691 (150)	694 (150)	676 (150)	700 (150)	694 (151)	680 (151)	691 (151)
Testing	70 (17)	84 (17)	82 (17)	74 (17)	71 (17)	89 (17)	65 (17)	71 (16)	85 (16)	74 (16)
Total	765 (167)	765 (167)	765 (167)	765 (167)	765 (167)	765 (167)	765 (167)	765 (167)	765 (167)	765 (167)

E1~10 is 10 sets of experiments.
Outside the brackets is the number of slices. The bracket indicates the number of cases (patients).

the last layer of the pre-trained network and used a fully connected layer of 2 neurons instead. Finally, we used target training set retrained the network and updated the pretrained parameters and weights. The parameters used in our training process are as follows: Batch size is 16, training epoch is 50, the learning rate of retraining is 0.0001, and the loss function is shown in section III.C. In the deep fusion model, the training parameters are the same, and the loss function is shown in section III.D.

This section presented the experimental results. Accuracy and the area under the receiver operating characteristic curve (AUC) were calculated to evaluate the predictive performance of the models.

A. DEEP LEARNING MODEL VS. CONVENTIONAL RADIOMICS MODEL

We compared the deep learning model with the conventional radiomics model. In the conventional radiomics model, we extracted the features as in [39]. We used the LASSO method for feature selection and RF for classification. The comparison results are shown in Table 3 . The result demonstrated that deep learning model is superior than the conventional radiomics model.

TABLE 3. The accuracy and AUC of the conventional radiomics model and deep Learning model.

Model	Conventional radiomics model [41]	Deep learning model
Average accuracy	67.04%±4.9	69.52%±5.1
Average AUC	0.6400±0.03	0.7233±0.06

B. EFFECTS OF FUSION MODEL

We compared the fusion models with the clinical model and the deep learning model. The clinical model only used the clinical data as mentioned in second II based on random forest (RF) method. The prediction results of the fusion models (A-D) outlined in section III are shown in Table 4, while the ROC curves were shown in Figure 5. The results showed that the proposed fusion model D can effectively combine CT images and clinical data, with the accuracy and AUC of model D reaching to 78.66% and 0.8248, respectively, which was better than either deep learning model based on CT images only or clinical model.

TABLE 4. The accuracy and AUC of deep Learning model, clinical model and fusion model A, B, C, D.

Model	Average accuracy	Average AUC
Deep learning model	69.52%±5.1	0.7233±0.06
Clinical model (RF)	76.03%±10.0	0.7532±0.13
Fusion model A	74.80%±4.1	0.7880±0.05
Fusion model B	78.05%±3.2	0.8045±0.08
Fusion model C	75.44%±7.0	0.7649±0.08
Fusion model D	78.66%±3.9	0.8248±0.07

Random forest performs best in clinical models

C. EFFECTS OF JOINT LOSS FUNCTION

The experimental results of the joint loss model as proposed in section III were shown in TABLE 5. The results showed that the joint loss model improved the prediction accuracy of fusion model D from 78.66% to 80.49% and slightly improved AUC from 0.8248 to 0.8331.

TABLE 5. The accuracy and AUC of the fusion model D with joint loss.

Model	Average accuracy	Average AUC
Fusion model D without joint loss	78.66%±3.9	0.8248±0.07
Fusion model D with joint loss	80.49%±4.3	0.8331±0.03

V. DISCUSSION

Our work focused on deep learning methods in predicting early recurrence of HCC. We proposed a deep learning model and four fusion models combining clinical and image data with different structures and compared its performance with clinical model and conventional radiomics model respectively.

Eran Segal *et al.* established a correlation map between 28 CT imaging traits and 78% gene expression profiles of HCC [40] and proved that imaging features could reflect biological behavior. Compared with qualitative semantic features, many studies have confirmed that quantitative radiomics feature can capture more potential information and successfully establish models to differentiate benign and

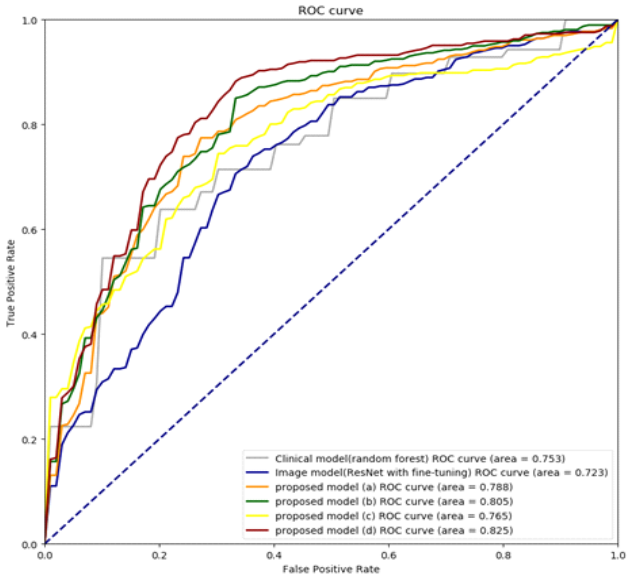


FIGURE 5. ROC curve of the models.

malignant properties, evaluate treatment responses as well as predict clinical outcomes [16], [19], [41], [42]. Besides, deep learning can mine deeper features from images. In this paper, the deep learning model (accuracy: 69.52%, AUC: 0.7233) performed better than conventional radiomics model (accuracy: 67.04%, AUC: 0.6400) in the prediction of ER of HCC. Which also indicated that the high-level features representation of deep learning is superior to the hand-crafted low-level and mid-level features.

Zhou *et al.* proposed a CNN based radiomics approach to extract high-level features from shear-wave elastography (SWE) images to classify malignant and benign breast tumor and demonstrated that deeper layers network configurations could achieved better performance. Two other previous studies showed that fusing multiparametric MRI with radiomics and deep feature performed better than single sequence or single feature type model in the prediction of lymph-vascular space invasion (LVSI) of cervical cancer. However, neither of them took clinical data into account in their models.

We built a clinical model based on clinical factors only to predict ER of HCC. Many studies have confirmed that clinical factors, such as HBV infection status, tumor size larger than 5 cm, the presence of liver cirrhosis is associated with ER of HCC [43], [44], [19], [20], [29]. In our study, age older than 40, tumor size larger than 5cm, N/L ratio$\geq$2.81, TB ($>$20.5U/L), AFP ($\geq 9\mu$g/L), the presence of portal vein invasion, and advanced BCLC stage were relevant with ER. In which, portal vein thrombosis and NLR$\geq$2.81 were consistence with Okamura *et al.* [31], Yang *et al.* [32].

In our experiment, the ER prediction ability of the clinical model (accuracy: 76.03%, AUC: 0.7532) was stronger than that of the deep learning model (accuracy: 69.03%, AUC: 0.7233), which showed that clinical variables were vital in predicting HCC prognosis as well as image features. But how to fuse two different data effectively is in little consensus. Therefore, we proposed four different fusion models to integrate clinical data and deep image features together. The results in Table 4 showed that the performance of fusion model A was better than the deep learning model without clinical data, but worse than the clinical model. The reason might be that the number of image features extracted by the convolutional layer was much larger than clinical data, combining clinical data with image features directly may decrease the performance of the model. Fusion model B was designed to reduce the dimensionality of image features at first and then combined with clinical data. According to the experimental results, the performance of Model B was significantly improved, further indicating that it is important to balance the number of image features and clinical data. In addition, we proposed fusion models C and D, in which clinical data was trained through convolutional layers or fully connected layer as same as medical images. The performance of fusion model C is worse than fusion model B while fusion model D is better than model B. The results suggested that the training of fully connected layer before the concatenation of clinical data and image features can improve the prediction performance, but clinical data is not suitable to be trained through convolutional layer.

Since our fusion models added clinical data as the second input data, the network structure can be divided into two branches. We calculated the value of the loss function of the two branches separately and added the value of the loss function after the concatenation. For fusion model D, we found that joint loss function can further improve the accuracy and the AUC of the prediction of the ER of HCC (see Figure 5).

In this paper, compared to those proposed models in previous studies (clinical model, conventional radiomics model), our proposed fusion models achieve better performance, indicating the potential value of combining multi-phase CT imaging with clinical data in ER prediction of HCC.

VI. CONCLUSION

In this paper, we proposed a deep learning model and four kinds of fusion models for ER prediction of HCC. Our proposed fusion model D, which incorporated multi-phase CT imaging and clinical data, achieved highest prediction accuracy of 78.66% and an AUC of 0.8248, which is significantly superior to both clinical model and the conventional radiomics model. Furthermore, fusion model D demonstrated a good performance with joint loss function, achieving an accuracy of 80.49% and an AUC of 0.8331.

ACKNOWLEDGMENT

(Weibin Wang and Qingqing Chen are co-first authors.)

REFERENCES

[1] K. M. Elsayes, A. Z. Kielar, M. M. Agrons, J. Szklaruk, A. Tang, M. R. Bashir, D. G. Mitchell, R. K. Do, K. J. Fowler, V. Chernyak, and C. B. Sirlin, "Liver imaging reporting and data system: An expert consensus statement," *J. Hepatocellular Carcinoma*, vol. 4, pp. 29–39, Feb. 2017.

[2] R. X. Zhu, W.-K. Seto, C.-L. Lai, and M.-F. Yuen, "Epidemiology of hepatocellular carcinoma in the Asia–Pacific region," *Gut Liver*, vol. 10, no. 3, pp. 332–339, May 2016.

[3] M. B. Thomas and A. X. Zhu, "Hepatocellular carcinoma: The need for progress," *J. Clin. Oncol.*, vol. 23, no. 13, pp. 2892–2899, May 2005.
[4] W. Y. Lau, *Hepatocellular Carcinoma*. Singapore: World Scientific, 2008.
[5] T. Yang, C. Lin, J. Zhai, S. Shi, M. Zhu, N. Zhu, J.-H. Lu, G.-S. Yang, and M.-C. Wu, "Surgical resection for advanced hepatocellular carcinoma according to Barcelona clinic liver cancer (BCLC) staging," *J. Cancer Res. Clin. Oncol.*, vol. 138, no. 7, pp. 1121–1129, Jul. 2012.
[6] S. A. Shah, S. P. Cleary, A. C. Wei, I. Yang, B. R. Taylor, A. W. Hemming, B. Langer, D. R. Grant, P. D. Greig, and S. Gallinger, "Recurrence after liver resection for hepatocellular carcinoma: Risk factors, treatment, and outcomes," *Surgery*, vol. 141, no. 3, pp. 330–339, Mar. 2007.
[7] S. Ibrahim, A. Roychowdhury, and T. K. Hean, "Risk factors for intrahepatic recurrence after hepatectomy for hepatocellular carcinoma," *Amer. J. Surg.*, vol. 194, no. 1, pp. 17–22, Jul. 2007.
[8] Z. Cheng, P. Yang, S. Qu, J. Zhou, J. Yang, X. Yang, Y. Xia, J. Li, K. Wang, Z. Yan, D. Wu, B. Zhang, N. Hüser, and F. Shen, "Risk factors and management for early and late intrahepatic recurrence of solitary hepatocellular carcinoma after curative resection," *HPB*, vol. 17, no. 5, pp. 422–427, May 2015.
[9] F. Hirokawa, M. Hayashi, Y. Miyamoto, M. Asakuma, T. Shimizu, K. Komeda, Y. Inoue, and K. Uchiyama, "Outcomes and predictors of microvascular invasion of solitary hepatocellular carcinoma," *Hepatol. Res.*, vol. 44, no. 8, pp. 846–853, Aug. 2014.
[10] Y. S. Choi, H. Rhee, J.-Y. Choi, Y. E. Chung, Y. N. Park, K. W. Kim, and M.-J. Kim, "Histological characteristics of small hepatocellular carcinomas showing atypical enhancement patterns on gadoxetic acid-enhanced MR imaging," *J. Magn. Reson. Imag.*, vol. 37, no. 6, pp. 1384–1391, Jun. 2013.
[11] R. K. Sterling, E. C. Wright, T. R. Morgan, L. B. Seeff, J. C. Hoefs A. M. Di Bisceglie, J. L. Dienstag, and A. S. Lok, "Frequency of elevated hepatocellular carcinoma (HCC) biomarkers in patients with advanced hepatitis C," *Amer. J. Gastroenterol.*, vol. 107, no. 1, p. 64, 2012.
[12] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. W. L. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.
[13] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016.
[14] N. M. Braman, M. Etesami, P. Prasanna, C. Dubchuk, H. Gilmore, P. Tiwari, D. Plecha, and A. Madabhushi, "Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI," *Breast Cancer Res.*, vol. 19, no. 1, p. 57, Dec. 2017.
[15] M. D. Kuo and N. Jamshidi, "Behind the numbers: Decoding molecular phenotypes with radiogenomics—Guiding principles and technical considerations," *Radiology*, vol. 270, no. 2, pp. 320–325, 2014.
[16] Y. Huang, C. Liang, L. He, J. Tian, C. Liang, X. Chen, Z. Ma, and Z. Liu, "Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer," *J. Clin. Oncol.*, vol. 34, no. 18, pp. 2157–2164, 2016.
[17] P. Kickingereder, S. Burth, A. Wick, M. Götz, O. Eidel, H.-P. Schlemmer, K. H. Maier-Hein, W. Wick, M. Bendszus, A. Radbruch, and D. Bonekamp, "Radiomic profiling of glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models," *Radiology*, vol. 280, no. 3, pp. 880–889, Sep. 2016.
[18] X. Ma, J. Wei, D. Gu, Y. Zhu, B. Feng, M. Liang, S. Wang, X. Zhao, and J. Tian, "Preoperative radiomics nomogram for microvascular invasion prediction in hepatocellular carcinoma using contrast-enhanced CT," *Eur. Radiol.*, vol. 29, no. 7, pp. 3595–3605, Jul. 2019.
[19] Y. Zhou, L. He, Y. Huang, S. Chen, P. Wu, W. Ye, Z. Liu, and C. Liang, "CT-based radiomics signature: A potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma," *Abdominal Radiol.*, vol. 42, no. 6, pp. 1695–1704, Jun. 2017.
[20] P. Ning, F. Gao, J. Hai, M. Wu, J. Chen, S. Zhu, M. Wang, and D. Shi, "Application of CT radiomics in prediction of early recurrence in hepatocellular carcinoma," *Abdominal Radiol.*, vol. 45, no. 1, pp. 64–72, Jan. 2020.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[24] L. Bi, J. Kim, A. Kumar, and D. Feng, "Automatic liver lesion detection using cascaded deep residual networks," 2017, *arXiv:1704.02703*. [Online]. Available: http://arxiv.org/abs/1704.02703
[25] D. Liang, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Lwamoto, X. Han, and Y.-W. Chen, "Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2018, pp. 666–675.
[26] D. Liang, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Lwamoto, X. Han, and Y.-W. Chen, "Residual convolutional neural networks with global and local pathways for classification of focal liver lesions," in *Proc. Pacific Rim Int. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2018, pp. 617–628.
[27] Z. Li, Y. Wang, J. Yu, Y. Guo, and W. Cao, "Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, Dec. 2017.
[28] J. Lao, Y. Chen, Z.-C. Li, Q. Li, J. Zhang, J. Liu, and G. Zhai, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Dec. 2017.
[29] N. Portolani, A. Coniglio, S. Ghidoni, M. Giovanelli, A. Benetti, G. A. Tiberio, and S. M. Giulini, "Early and late recurrence after liver resection for hepatocellular carcinoma: Prognostic and therapeutic implications," *Ann. Surg.*, vol. 243, no. 2, p. 229, 2006.
[30] P.-E. Chang, W.-C. Ong, H.-F. Lui, and C.-K. Tan, "Is the prognosis of young patients with hepatocellular carcinoma poorer than the prognosis of older patients? A comparative analysis of clinical characteristics, prognostic features, and survival outcome," *J. Gastroenterol.*, vol. 43, no. 11, pp. 881–888, Nov. 2008.
[31] Y. Okamura, R. Ashida, T. Ito, T. Sugiura, K. Mori, and K. Uesaka, "Preoperative neutrophil to lymphocyte ratio and prognostic nutritional index predict overall survival after hepatectomy for hepatocellular carcinoma," *World J. Surg.*, vol. 39, no. 6, pp. 1501–1509, Jun. 2015.
[32] H.-J. Yang, Z. Guo, Y.-T. Yang, J.-H. Jiang, Y.-P. Qi, J.-J. Li, L.-Q. Li, and B.-D. Xiang, "Blood neutrophil-lymphocyte ratio predicts survival after hepatectomy for hepatocellular carcinoma: A propensity score-based analysis," *World J. Gastroenterol.*, vol. 22, no. 21, p. 5088, 2016.
[33] H. O. Lancaster, *The Chi-Squared Distribution*. Hoboken, NJ, USA: Wiley, 1969.
[34] M. L. McHugh, "The chi-square test of independence," *Biochemia Medica*, vol. 23, no. 2, pp. 143–149, 2013.
[35] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.
[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[37] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.
[38] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.
[39] P. Aonpong, Q. Chen, Y. Iwamoto, L. Lin, H. Hu, Q. Zhang, and Y.-W. Chen, "Comparison of machine learning-based radiomics models for early recurrence prediction of hepatocellular carcinoma," *J. Image Graph.*, vol. 7, no. 4, pp. 117–125, 2019.
[40] E. Segal, C. B. Sirlin, C. Ooi, A. S. Adler, J. Gollub, X. Chen, B. K. Chan, G. R. Matcuk, C. T. Barry, H. Y. Chang, and M. D. Kuo, "Decoding global gene expression programs in liver cancer by noninvasive imaging," *Nature Biotechnol.*, vol. 25, no. 6, pp. 675–685, 2007.
[41] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine learning methods for quantitative radiomic biomarkers," *Sci. Rep.*, vol. 5, no. 1, p. 13087, Oct. 2015.

[42] M. Avanzo, J. Stancanello, and I. El Naqa, "Beyond imaging: The promise of radiomics," *Phys. Medica*, vol. 38, pp. 122–139, Jun. 2017.
[43] W.-H. Ho, K.-T. Lee, H.-Y. Chen, T.-W. Ho, and H.-C. Chiu, "Disease-free survival after hepatic resection in hepatocellular carcinoma patients: A prediction approach using artificial neural network," *PLoS ONE*, vol. 7, no. 1, Jan. 2012, Art. no. e29179.
[44] J. H. Shim, M.-J. Jun, S. Han, Y.-J. Lee, S.-G. Lee, K. M. Kim, Y.-S. Lim, and H. C. Lee, "Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma," *Ann. Surg.*, vol. 261, no. 5, pp. 939–946, May 2015.
[45] S. Singh, S. Sinha, and S. Rawat, "Extraction of influential markers affecting HCC survival chances and its prediction," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2019, pp. 1–6.

WEIBIN WANG was born in Guangxi, China, in 1994. He received the bachelor's degree from Northeast University, Shenyang, China, in 2017, and the master's degree in information science and engineering from Ritsumeikan University, Kusatsu, Shiga, Japan, in 2019, where he is currently pursuing the Ph.D. degree with the Department of Information Science and Engineering. His research interests include medical image processing and deep learning.

QINGQING CHEN was born in Wenzhou, China, in 1993. She received the bachelor's degree in medicine from Wenzhou Medical University, in 2016. She is currently pursuing the Ph.D. degree with the Sir Run Run Shaw Hosoital, Zhejiang University School of Medicine, China. She has participated and presented with the Radiology Society of North America (RSNA), Chicago, in 2017. Her main research interest includes artificial intelligence on liver lesion imaging, such as classification, li-rads category, prognosis, and so on. She received the Short-Term Exchange from the Department of Information Science and Engineering, Ritsumeikan University, for a period of three months.

YUTARO IWAMOTO (Member, IEEE) received the B.E., M.E., and D.E. degrees from Ritsumeikan University, Kusatsu, Japan, in 2011, 2013, and 2017, respectively.

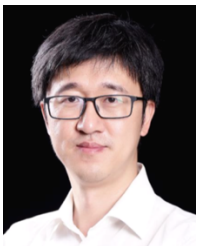
He is currently an Assistant Professor with Ritsumeikan University. His current research interests include image restoration, segmentation, classification of medical imaging, and deep learning.

PANYANAT AONPONG was born in Suphan Buri, Thailand, in 1993. He received the Bachelor of Engineering degree in computer engineering from the King Mongkut's Institute of Technology, Bangkok, Thailand, in 2015, and the Master of Engineering degree in information communication technology for embedded systems from Kasetsart University, Bangkok, in 2017. He is currently pursuing the Ph.D. degree with the Department of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. He has work experience in satellite imagery with the Geoinformatics Information and Space Technology Development Agency (GISTDA) and the Thai Government Organization in satellites and images. His research interest includes computer image processing.

LANFEN LIN (Member, IEEE) was born in Pingyang, China, in 1969. She received the B.S. and Ph.D. degrees in aircraft manufacture engineering from Northwestern Polytechnical University, in 1990 and 1995, respectively.

She held a postdoctoral position with the College of Computer Science and Technology, Zhejiang University, China, from January 1996 to December 1997, where she is currently a Full Professor and the Vice Director of the Artificial Intelligence Institute. She Leads the Research Group to exploring AI technologies for medical imaging. She is also a Leader of more than 40 projects, include subproject of the Major State Basic Research Program (973), China, and the National High Technology Research and Development Program of China. She has published more than 140 research papers in the journals or international conference proceedings. Her research interests include big data analysis, data mining, knowledge management, and so on. She is a member of the Advisory Expert Group for Manufacturing Industry Information, Zhejiang.

HONGJIE HU received the master's degree in medicine, in 1992, and the Ph.D. degree in medicine from Zhejiang Medical University, in 1998. He went for further education with the Loma Linda University Medical Center, in 2000, the Myo Clinic, in 2008, and the Cleveland University Clinic, USA. He has rich clinical experience in cardiothoracic diagnosis and interventional radiotherapy. He has engaged in medical imaging for over 30 years. He has been a Chief Physician and a Ph.D. Supervisor with the Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, since 2015.



QIAOWEI ZHANG was born in China, in 1973. He received the Ph.D. degree from Zhejiang University, Zhejiang, China, in 2005. He was a Resident and a Fellow Doctor with the Sir Run Run Shaw Hospital, Hangzhou, China. He received the Short Term Fellowship/Research Training with the Loma Linda University Medical Center, USA, and the Grosshadern Hospital, LMU, Germany. He is currently an Attending Doctor with the Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, and an Assistant Director of the Radiology Department.

YEN-WEI CHEN (Member, IEEE) was born in Hangzhou, China, in 1962. He received the B.E. degree from Kobe University, Kobe, Japan, in 1985, and the M.E. and D.E. degrees from Osaka University, Osaka, Japan, in 1987 and 1990, respectively.

He was a Research Fellow with the Institute of Laser Technology, Osaka, from 1991 to 1994. From October 1994 to March 2004, he was an Associate Professor and a Professor with the Department of Electrical and Electronic Engineering, The University of the Ryukyus, Okinawa, Japan. He is currently a Professor with the College of Information Science and Engineering, Ritsumeikan University, Japan. He is also a Visiting Professor with the College of Computer Science and Technology, Zhejiang University, and the Zhejiang Laboratory, Hangzhou, China. He is a principal investigator of several projects in bio-medical engineering and image analysis, funded by the Japanese Government. He has published more than 300 research articles. His research interests include medical image analysis and pattern recognition.

Prof. Chen received many distinguished awards.