

Article

Early Recurrence Prediction of Hepatocellular Carcinoma Using Deep Learning Frameworks with Multi-Task Pre-Training

Jian Song ^{1,2}, Haohua Dong ^{2,*†}, Youwen Chen ¹, Xianru Zhang ², Gan Zhan ², Rahul Kumar Jain ²  and Yen-Wei Chen ^{2,*} 

¹ School of Mathematical Sciences, Huaqiao University, Quanzhou 362021, China; songjian@hqu.edu.cn (J.S.); 1821121003@stu.hqu.edu.cn (Y.C.)

² College of Information Science and Engineering, Ritsumeikan University, Osaka 567-0817, Japan; dong.h@ctw.inc (H.D.); gr0609xf@ed.ritsumei.ac.jp (X.Z.); gr0502vs@ed.ritsumei.ac.jp (G.Z.); r-jain@fc.ritsumei.ac.jp (R.K.J.)

* Correspondence: chen@is.ritsumei.ac.jp

† Current address: CTW Inc., Tokyo 106-0032, Japan.

Abstract: Post-operative early recurrence (ER) of hepatocellular carcinoma (HCC) is a major cause of mortality. Predicting ER before treatment can guide treatment and follow-up protocols. Deep learning frameworks, known for their superior performance, are widely used in medical imaging. However, they face challenges due to limited annotated data. We propose a multi-task pre-training method using self-supervised learning with medical images for predicting the ER of HCC. This method involves two pretext tasks: phase shuffle, focusing on intra-image feature representation, and case discrimination, focusing on inter-image feature representation. The effectiveness and generalization of the proposed method are validated through two different experiments. In addition to predicting early recurrence, we also apply the proposed method to the classification of focal liver lesions. Both experiments show that the multi-task pre-training model outperforms existing pre-training (transfer learning) methods with natural images, single-task self-supervised pre-training, and DINOv2.

Keywords: early recurrence (ER); hepatocellular carcinoma (HCC); multi-task pre-training; deep neural network; self-supervised learning; phase shuffle prediction; multi-phase CT image



Citation: Song, J.; Dong, H.; Chen, Y.; Zhang, X.; Zhan, G.; Jain, R.K.; Chen, Y.-W. Early Recurrence Prediction of Hepatocellular Carcinoma Using Deep Learning Frameworks with Multi-Task Pre-Training. *Information* **2024**, *15*, 493. <https://doi.org/10.3390/info15080493>

Academic Editors: Haiyan Li and Dongming Zhou

Received: 1 July 2024

Revised: 6 August 2024

Accepted: 15 August 2024

Published: 17 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hepatocellular carcinoma (HCC) is a primary liver cancer with a high mortality rate. It is a prevalent malignancy, particularly in regions such as Asia, Africa, and Southern Europe, which have a high incidence of chronic hepatitis B and C virus infections [1,2]. The accurate differentiation of focal liver lesions (FLLs) is a crucial task for the diagnosis of liver cancers. On the other hand, as a main treatment option for HCC, surgical resection is the most commonly used method [3,4]. Patients with stage 1 HCC who undergo surgical treatment generally have the longest 5-year survival rate compared to other treatment options. However, the recurrence rate of HCC can reach 70–80% after surgical resection, leading to disease progression and reduced survival rates [5]. Despite advances in surgical techniques and other curative therapies, postoperative recurrence of HCC (intrahepatic or extrahepatic) remains a leading cause of patient mortality [6]. The peak time for hepatocellular carcinoma (HCC) recurrence after resection is typically within the first year, which is defined as “early recurrence” (ER) [7]. Time to recurrence is an independent survival factor in patients with HCC. ER is associated with a worse prognosis and lower overall survival (OS) rates compared to late recurrence [7,8]. Preoperative prediction of ER for patients with liver cancer can help physicians to select appropriate treatment modalities and optimize postoperative monitoring and surveillance. Therefore, alongside differentiating FLLs, early recurrence prediction for patients with HCC before radical surgical resection is also crucial for improving patient outcomes and survival rates.

Medical imaging plays a crucial role in the standard care of patients with FLLs and has evolved into a significant non-invasive technique for detecting and characterizing the malignancy of HCC [9,10]. In 2012, Lambin introduced the concept of radiomics, which uses machine learning techniques to extract numerous features from medical images to analyze disease and prognosis [11]. Radiomics enables personalized medicine via non-invasive tools, improving treatment and enabling patient-specific care [12,13]. Currently, challenges such as the absence of standardization and adequate validation of radiomics models are impeding the clinical implementation of radiomics-based technologies [14]. Furthermore, previous studies have relied on manually designed low- or mid-level image features for feature extraction, which may not capture the full range of information relevant to early recurrence. The manual tuning of these models can also introduce human bias.

In recent years, deep learning has been applied to the computer-aided diagnosis of various cancers [15,16], including the differentiation (or classification) of FLLs [17–19] and the prediction of ER in HCC [20,21]. Deep learning uses convolutional neural networks (CNNs) that can directly perform feature extraction and feature analysis on image inputs. Deep learning has an end-to-end model structure that can automatically extract relevant features from images, eliminating human bias and surpassing manually defined feature extraction to extract high-level semantic features. Although deep learning has demonstrated superior performance compared to radiomics approaches in various areas, the data-hungry nature of deep learning frameworks presents a significant challenge for medical image analysis, mainly due to the limited availability of annotated data samples.

Wang et al. have demonstrated that pre-trained deep learning models using ImageNet can significantly improve computer-aided diagnosis performance [19–21]. The enhancements will be constrained by the domain gap between natural images and medical images. Self-supervised learning has recently been proposed as a solution to address the domain gap problem in pre-training using ImageNet. Instance discrimination [22] and MOCO [23] used contrast learning to learn the difference between each instance to obtain a representation of the similarity among instances. In preliminary studies of this work, we proposed two multi-phase CT image-specific self-supervised pre-training methods for the classification of FLLs: case discrimination [24] and phase shuffle prediction [25]. The case discrimination method leverages the properties of 3D volumetric medical images for the classification of FLLs by focusing on feature representation between different CT images. Phase shuffle prediction involves shuffling the phase order of unannotated multi-phase CT images and predicting the sequence. This method aims to enhance the classification of FLLs by concentrating on the feature representation within the multi-phase CT images. The effectiveness of our two multi-phase CT image-specific self-supervised pre-training methods (case discrimination and phase shuffle prediction) has been demonstrated in the classification of FLLs, and the results have been presented at two international conferences [24,25].

In this paper, we propose a multi-task pre-training framework by combining case discrimination [24] and phase shuffle prediction [25] for future improvements in computer-aided diagnosis performance. Multi-task pre-training aims to learn both feature representations within and between medical images, allowing for the extraction of comprehensive information relevant to FLLs.

The main contributions are summarized below:

- (1) We propose a simple but effective self-supervised learning method, which is called Phase Shuffle Prediction. The proposed phase shuffle prediction focuses on feature representation within multi-phase CT images.
- (2) To further enhance the pre-training performance of deep learning, we propose a novel self-supervised feature learning approach based on multi-tasking by combining the newly proposed phase shuffle prediction with our previously proposed case discrimination [24], focusing on the feature representation of different CT images. Through these two pretext tasks, it is possible to obtain a representation that encompasses information from both within and between images, allowing the extraction of comprehensive information relevant to liver cancers.

- (3) The effectiveness of our proposed method is demonstrated not only in the classification of FLLs, but also in the prediction of ER of HCC. To the best of our knowledge, this is the first application of self-supervised learning for predicting the ER of HCC using multi-phase CT imaging.

This paper involves methodological and experimental extensions and validations. We validated the effectiveness of the proposed method not only in the prediction of ER in HCC, but also in the classification of FLLs.

This paper is organized into four sections. Section 2 gives a brief review of related work. The proposed approach is described in detail in Section 3. The experiments and results are shown in Section 4. The last section includes our conclusions.

2. Related Work

2.1. Pre-Trained ImageNet Model

Transfer learning is a powerful technique for training a model with limited annotated data for a specific task. It reuses a pre-trained model, which is pre-trained on ImageNet datasets, for other image classification tasks. Typically, the network's weights are updated using the limited annotated target datasets available while retaining their original structure. The shape of the Fully Connected (FC) layer, which serves as the classifier, is modified to suit the target task classes and updated from scratch using the target dataset, whereas in some other fine-tuning approaches, the weights of deeper layers are updated using the target dataset while keeping the weights of some shallower layers frozen. Wang et al. demonstrated that the pre-trained ImageNet model can improve the prediction performance of ER [20,21]. However, pre-trained ImageNet models have limited representation for medical images due to the differences in domains, which may hinder their effectiveness for downstream tasks.

2.2. Self-Supervised Learning

Self-supervised learning is a novel approach to unsupervised learning that involves pre-training models using a target dataset with a predefined pretext task. This differs from pre-trained ImageNet models that are pre-trained on a different domain dataset. The pipeline of self-supervised learning has two steps:

- (1) Pre-training a deep neural network model on a pretext task with an unannotated target dataset.
- (2) Fine-tuning the pre-trained model for the main task with an annotated target dataset.

The design of the pretext task is a critical factor in self-supervised learning. For this task, several self-supervised learning methods with different pretext tasks have been proposed, such as solving jigsaw puzzles [26], rotation prediction [27], and phase shuffle prediction [25]. This kind of self-supervised learning can mine the internal features of each image. In addition to this, several other self-supervised methods, such as case discrimination [24], Moco [23], and SimCLR [28], based on contrastive learning, are also proposed. This kind of self-supervised learning can capture apparent visual similarity among categories. In our prior research, we have shown that the methods of phase shuffle prediction [25] and case discrimination [24] were effective in capturing either intra-image or inter-image features. In this paper, we propose a multi-phase CT image-specific self-supervised learning approach that fuses two kinds of pretext tasks (phase shuffle prediction [25] and case discrimination [24]) in order to achieve a representation both within and between images. We called it the multi-task pre-training model.

3. Methods

3.1. Overview of the Proposed Method

An overview of the proposed method is shown in Figure 1. It can be divided into two steps. The first step is the pre-training step using a pre-defined pretext task (Figure 1a). The second step is a fine-tuning step using the target task (i.e., ER prediction of HCC)

(Figure 1b). The main network is a CNN encoder, which is used for high-level semantic feature extraction. After pre-training (first step), the weights of the pre-trained model (CNN encoder) are used as initialization parameters and are fine-tuned using the original multi-phase CT images and their labels for the target task (i.e., the prediction of early recurrence) (second step). We use ResNet18 [29] as the CNN encoder, which has been widely used to perform image classification tasks in various previous works [30]. The network architecture is shown in Figure 2. In the pre-training step (Figure 1a), simple fully connected layers (FCs) are used as classifiers, while we use a Multi-layer Perceptron (MLP) as a classifier for the target task (Figure 1b). In this research, we propose a multi-task (dual-task) pre-training approach. Thus, there are two FCs in the pre-training step (Figure 1a). The encoder and two FCs are trained in the pre-training step using the pretext tasks. In the fine-tuning step (Figure 1b), the pre-trained encoder and the MLP are trained using the target task (fine-tuning).

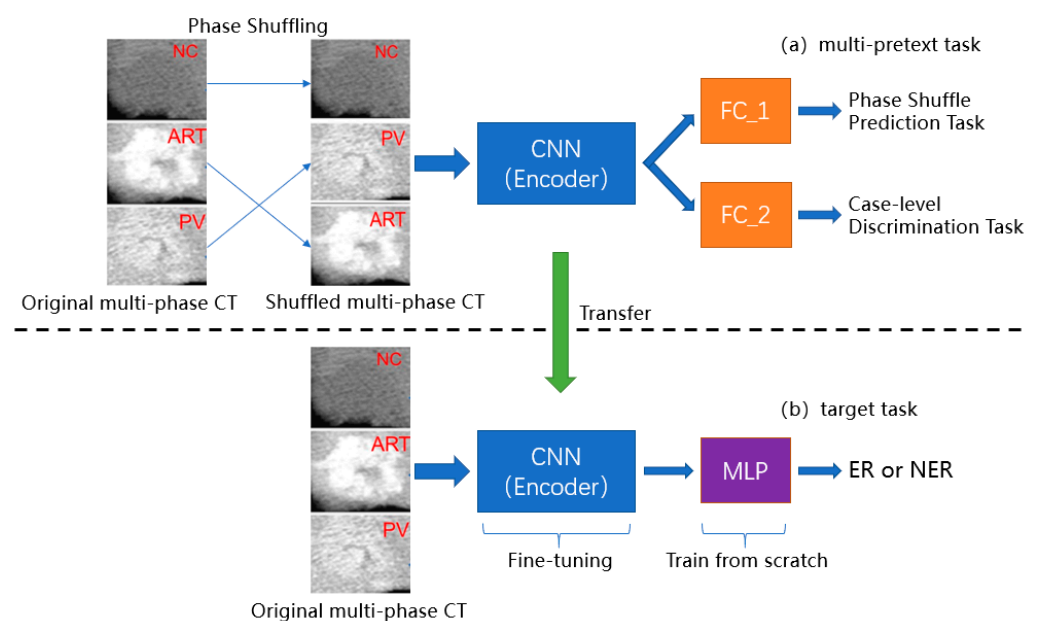


Figure 1. Overview of proposed method: (a) pretext task; (b) target task.

3.2. Multi-Task Pre-Training

In this paper, we propose a multi-phase CT image-specific self-supervised learning approach that fuses two kinds of pretext tasks (phase shuffle prediction [25] and case discrimination [24]) to achieve a representation both within and between images.

3.2.1. Phase Shuffle Prediction Task

The first pretext task is the phase shuffle prediction task [25], the goal of which is to predict the order of the shuffled phase of multi-phase CT images. The aim is to learn the representation within images (intra-image). The phase order of multi-phase CT images is randomly shuffled as shown in Figure 1a. The original order is NC, ART, and PV. In the case of Figure 1a, the phase order is shuffled to NC, PV, and ART. The possible number of the phase order is $K!$, where K represents the number of phases ($K = 3$ in this research). In the kind of self-supervised learning that captures inter-image features, the goal of the pretext task is to predict the order of phases, which is a six-class classification problem (i.e., class 1: NC, ART, PV; class 2: NC, PV, ART; class 3: ART, NC, PV; class 4: ART, PV, NC; class 5: PV, NC, ART; class 6: PV, ART, NC) [25].

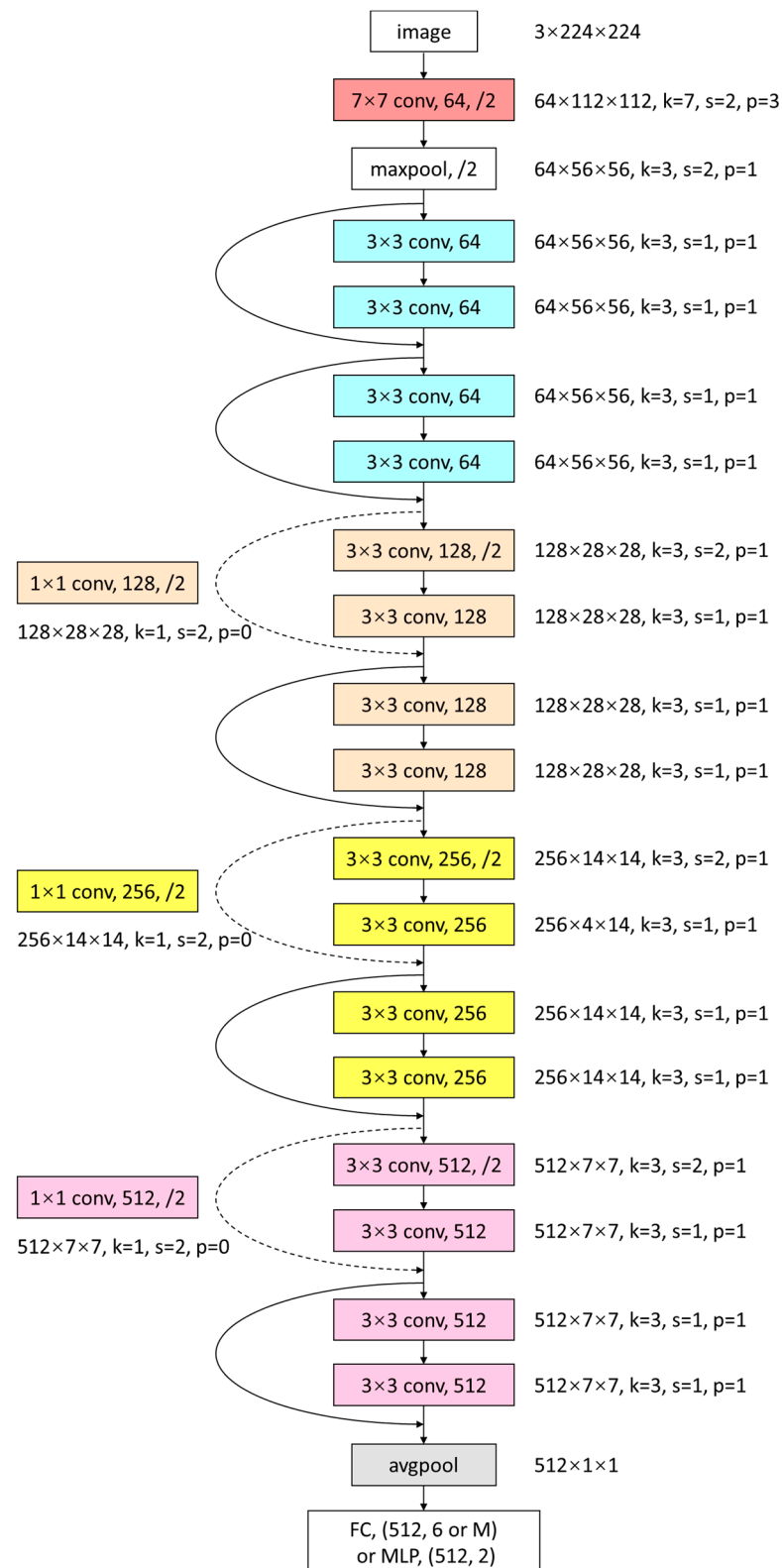


Figure 2. The architecture of the ResNet18 network. Different colors are used to mark the five stages of convolutional layers in ResNet18.

These shuffled images of three different CT phases are treated as a color image and are fed into the convolutional neural network (CNN) for feature extraction. The input size is $3 \times 224 \times 224$ and the feature size is 512×1 . The classifier of FC1 is used to predict

the order of the shuffled phases (i.e., six-class classification). The output layer of FC_1 has 6 neurons.

3.2.2. Case-Level Discrimination Task

The second pretext task is the case-level discrimination task [24], the goal of which is to learn the difference between each case (3D volume) to obtain a representation between cases (inter-image). In the traditional instance-level discrimination task [22], each image (instance) has its own pseudo-label. Suppose we have M images; the pretext task involves M -class classification. The model will be pre-trained using one-positive and $M-1$ negative samples. On the other hand, medical images from CT or MRI scanners capture 3D volumes of the body, which are usually reconstructed into a series of 2D images, each representing a thin slice of the body. Typically, a case of one patient consists of multiple such slices but with one single lesion. A tumor that spreads across multiple slices often shares similar features such as curves and edges. To capture the similarities within a case, these slices can be merged into the same pseudo-label using self-supervised learning [24]. Thus, the pretext task is called case-level classification, which was proposed in our previous work [24]. Suppose the number of cases used for pre-training is M ($M = 167$ in this research); then, the case-level discrimination task is an M -class classification. The classifier of FC_2 in Figure 1a is used for the case-level discrimination task.

3.2.3. Loss Function for Pre-Training

We pre-train a deep neural network model on fusing these two kinds of self-supervised methods together in order to explore a robust representation both within and between images.

The proposed fusion model incorporates a softmax layer in both the phase shuffle prediction and case-level classification paths, with cross-entropy serving as the loss function. Given the multi-task nature of the model, two losses are obtained, as depicted in Figure 1. Accordingly, we formulate the loss function of the overall model as follows:

$$\begin{aligned} L &= \alpha \times L_{case} + (1 - \alpha) \times L_{phase} \\ &= \alpha \times Cross\ Entropy(C(I_j)) + (1 - \alpha) \times Cross\ Entropy(P(I_j)) \end{aligned} \quad (1)$$

where L_{case} is the loss from case discrimination and L_{phase} is the loss from phase shuffle prediction. I_j represents the j th CT image input data. $C(I_j)$ and $P(I_j)$ are the outputs of the two pathways.

α is a weight used to balance two tasks, which is a hyper-parameter. After conducting the experiment, we observed that the training difficulty for case discrimination was slightly higher compared to the phase shuffle prediction. In this study, the optimum value of α was found to be 0.6.

3.3. Target Task (Fine-Tuning)

In the fine-tuning stage, non-linear MLP with one hidden layer (h-Dimension) is used instead of an FC layer in the classifier, as shown in Figure 1b. After pre-training, the weights of the pre-trained model (CNN encoder) are used as the initialization parameters and are fine-tuned using original multi-phase CT images and their labels for the target task (i.e., prediction of ER). The MLP is also trained together with the pre-trained CNN encoder using the target task in an end-to-end manner. Each sample has a label of ER or NER, which are provided by doctors. Thus, the target task (i.e., the prediction of ER) is a two-class classification problem. The MLP has an output layer with two neurons. Cross-entropy is used as the loss function for the target task (fine-tuning).

4. Experiment

In order to validate the effectiveness of the proposed method, we first apply the proposed method to the prediction of the ER of HCC (Task 1). Then, we also apply the proposed method to the classification of focal liver lesions (FLLs) (Task 2) to validate the

generalization of the proposed method. To this end, our experiments are organized into two parts.

4.1. Task 1: Prediction of Early Recurrence

In this section, we focused on a challenging task, predicting the ER of HCC, examining whether the proposed method can work well for prediction tasks. We used ResNet18 as our backbone network to validate the effectiveness of our proposed method for the task of prediction.

4.1.1. Data

The medical images used in this study were collected from Run Run Shaw Hospital, Zhejiang University, China. This retrospective study initially included 331 consecutive HCC patients who underwent hepatectomy between 2012 and 2016. The patient selection process followed the following criteria: (1) confirmation of postoperative HCC; (2) availability of a contrast-enhanced CT scan taken within one month prior to surgery; (3) follow-up for at least one year postoperatively; (4) no history of preoperative HCC treatment; and (5) negative surgical margins indicating complete tumor resection. The peak time for HCC recurrence was found to be 1 year after the resection, which was referred to as “early recurrence” (ER) [7]. A total of 167 HCC patients (140 men and 27 women) were included in this study. Out of the 167 included HCC patients, 65 patients (38.9%) were classified as early recurrence (ER), while the remaining 102 patients (61.1%) did not experience any recurrence, and were thus classified as non-early recurrence (NER). Therefore, the patients were categorized into two groups: ER and NER.

The number of CT slices containing tumors varied across patients due to differences in tumor sizes and locations. For our datasets, we selected the central slice (with the largest tumor cross-section) as well as its adjacent slices. A total of 765 labeled slices were used in our experiments, with slice thickness ranging from 5 to 7 mm and an in-plane resolution of 0.57–0.59 mm. Each CT image had three phases (i.e., NC, ART, PV) with a pixel size of 512×512 . The region of interest (ROI) for each lesion was manually annotated by experienced radiologists. For the experiments, we utilized 2D region of interest (ROI) slice images, which were resized to 224×224 . The three phase images were treated as a color image with three channels. Thus, the input image size was $3 \times 224 \times 224$.

Different stages of the tumor and liver may exhibit different characteristics, indicating that multi-phase CT could provide more information. An example of a contrast-enhanced CT scan of a patient before surgery is shown in Figure 3.

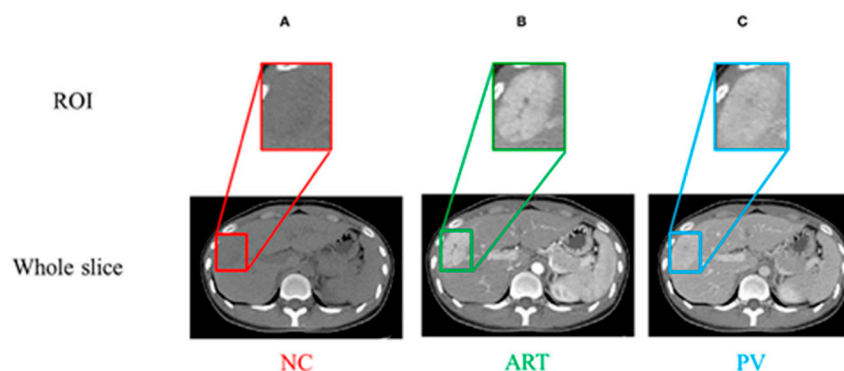


Figure 3. (A–C) NC, ART, and PV phases, respectively. The region of interest (ROI) is the bounding box of the tumor.

We used 10-fold cross-validation as our evaluation method. The accuracy and the area under the curve (AUC) for the receiver operating characteristics were calculated to evaluate the prediction performance of the model. We randomly divided 167 patients into 10 groups; each group contained 6 or 7 ERs and 10 or 11 NERs. During the 10-fold cross-validation,

we selected one group as the test dataset and the remaining nine groups were used as the training dataset. The mean value was calculated for the results obtained from the ten sets of experiments. This mean value was used as the final score of the model. The number of CT slices containing a tumor varied among patients due to differences in tumor size and location. For our dataset, we selected the central slice (which has the largest tumor cross-section) along with its adjacent slices. In total, 765 labeled slices were utilized in our experiments. Table 1 summarizes the number of training images and test images (CT slice images) for each experiment. The numbers in brackets indicate the number of cases (3D CT volumes).

Table 1. Data distribution for 10-fold cross-validation.

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)

4.1.2. Implementations

For the pre-training, we trained our network for 1000 epochs using a batch size of 256. We use Adam as our optimizer with a learning rate of 0.05. For the training of the target task, we fine-tune our network for 200 epochs using a batch size of 256. The learning rate was also 0.05. The training setup is shown in Table 2.

Table 2. Computation environment.

GPU	NVIDIA GeForce RTX 3090
CPU	Intel® X® Platinum 8358P
OS	Ubuntu 20.04
Deep learning Framework	PyTorch2.0

4.1.3. Results

We conducted ablation experiments on our datasets for predicting the ER of HCC, which demonstrated the effectiveness of each component in our proposed model. The results of the ablation experiments are summarized in Table 3. To assess the predictive performance of the model, we calculated both the accuracy and the AUC. ResNet18 without pre-training was employed as the baseline (Model 1). When trained from scratch without weight initialization using pre-trained models, the prediction accuracy was $67.44\% \pm 5.29$, with an AUC of 0.666 ± 0.06 . Two self-supervised methods surpassed the performance of training from scratch, reaffirming the effectiveness of self-supervised learning. We first validated the effectiveness of case discrimination (Model 2), which led to an accuracy improvement of around 4.5% and an AUC improvement of approximately 0.05 compared to the baseline (Model 1). Subsequently, we evaluated the impact of phase shuffle (Model 3), resulting in an accuracy improvement of roughly 3% and an AUC increase of about 0.03. This indicates that self-supervised learning based on the two pretext tasks can effectively learn features within and between images, resulting in better performance compared to deep learning models trained from scratch. Finally, the proposed multi-tasking pre-training model further improved the accuracy by approximately 7.2% and enhanced the AUC by about 0.07 in comparison to the baseline (Model 1). The proposed multi-task pre-training model also significantly surpassed the single pre-training models (either case discrimination or phase shuffle prediction).

Table 3. Ablation study on predicting ER.

Model	Pre-Training		ACC (%)	AUC
	Case Discrimination	Phase Shuffle Prediction		
Model 1			67.44% ± 5.29	0.666 ± 0.06
Model 2	✓		71.98% ± 2.64	0.715 ± 0.03
Model 3		✓	70.15% ± 3.71	0.694 ± 0.04
Proposed	✓	✓	74.65% ± 3.30	0.739 ± 0.04

Bold indicates the highest values.

We also compared the proposed method with other different existing pre-training methods, including pre-training with ImageNet [21], self-supervised learning methods with rotation prediction [27], phase shuffle prediction [25], instance modeling [22], case discrimination [24], and DINOv2 [31]. Note that DINOv2 is a state-of-the-art self-supervision method, which was proposed in 2023. The comparison results are summarized in Table 4.

Table 4. Comparison with other transfer learning methods for predicting early recurrence.

Models	ER	NER	Average	AUC
Fine-tuning (ImageNet) [21]	57.45% ± 11.35	80.45% ± 10.96	69.34% ± 3.43	0.695 ± 0.04
Self-supervised (rotation) [27]	60.15% ± 13.60	76.02% ± 9.29	68.53% ± 3.75	0.684 ± 0.05
Self-supervised (phase shuffle prediction) [25]	55.22% ± 11.84	83.22% ± 13.98	70.15% ± 3.71	0.694 ± 0.04
Self-supervised (instance-level) [22]	56.83% ± 18.53	79.68% ± 11.33	69.17% ± 4.02	0.683 ± 0.05
Self-supervised (case-level) [24]	59.88% ± 11.98	82.56% ± 10.01	71.98% ± 2.64	0.715 ± 0.03
Self-supervised (DINOv2) [31]	58.68% ± 11.73	83.31% ± 10.30	71.79% ± 2.86	0.711 ± 0.04
Self-supervised multi-task pre-training model (proposed)	65.64% ± 11.18	82.17% ± 9.89	74.65% ± 3.30	0.739 ± 0.04

Bold indicates the highest values.

As shown in Table 4, self-supervised learning methods using medical images achieved better results than the pre-training model using ImageNet. For methods learning representations within images, our proposed phase shuffle prediction method achieved better results than the rotation prediction method. For the methods learning representations between images, our proposed case-level discrimination achieved better results than the instance-level discrimination. The proposed multi-task prediction model, which learns both intra-image and inter-image representations, achieved superior results compared to single-task models (i.e., phase shuffle prediction [25] and case-level discrimination [24]). Compared with the currently proposed self-supervised method (i.e., DINOv2), our proposed method still showed superior results. Compared to case-level discrimination [24], which had the highest accuracy among the existing methods, the proposed method improved both the accuracy and AUC by approximately 2.5%. The effectiveness of our proposed method was demonstrated.

4.2. Task2: Classification of Focal Liver Lesions

We applied the proposed method to FLL classification to validate its generalization. This section covers the dataset and implementation details, an ablation study to assess each component's effectiveness, and a comparative analysis with the baseline models.

4.2.1. Data

The effectiveness of the proposed method was confirmed through the utilization of our proprietary Multi-Phase CT dataset of Focal Liver Lesions (MPCT-FLLs) [24,25]. In our experiments, we employed four distinct lesion types (Cyst, FNH, HCC, and HEM) that were collected by Sir Run Run Shaw Hospital, Zhejiang University, spanning the years 2015 to 2017. In total, our dataset comprised 85 CT volumes, which included 489 slice images. For each volume, a selection of slices centered on the lesion was made. The slice thickness ranged from 5 to 7 mm, and the in-plane resolution was between 0.57 and 0.59 mm. The size of the 2D slice image was 512×512 . Each CT image consisted of three phases (i.e., NC, ART, and PV). Experienced radiologists annotated the region of interest (ROI) for each lesion. The 2D ROI slice images were employed for the experiments, with each ROI resized to 128×128 . Treating the three-phase images as a color image with three channels, the input image size was $3 \times 128 \times 128$. Figure 4 illustrates the evolution patterns of FLLs as observed in the multiphase CT scans (NC, ART, and PV).

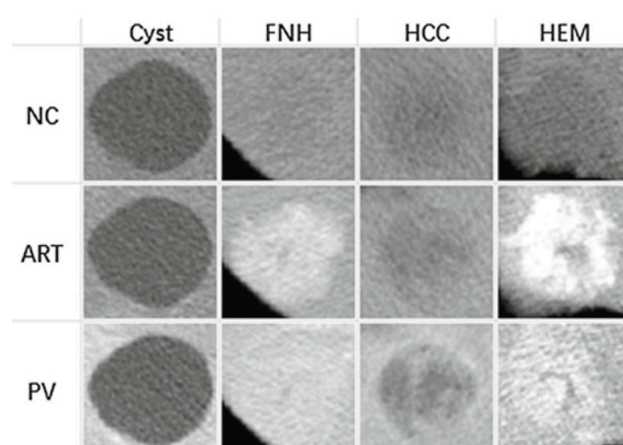


Figure 4. Evolution patterns of four FLLs in multi-phase CT.

The dataset was partitioned into 5 groups for the purpose of conducting 5-fold cross-validation. The data distribution is detailed in Table 5. For each fold, one group was designated as the test dataset, while the remaining four groups were employed as the training dataset.

Table 5. Data distribution for 5-fold cross-validation.

Type	Cyst	FNH	HCC	HEM	Total
Group 1: case (slice)	5 (29)	4 (15)	4 (30)	4 (21)	17 (95)
Group 2: case (slice)	6 (31)	3 (17)	4 (29)	4 (33)	17 (110)
Group 3: case (slice)	6 (37)	3 (7)	4 (36)	4 (17)	17 (97)
Group 4: case (slice)	6 (24)	3 (17)	4 (35)	4 (19)	17 (95)
Group 5: case (slice)	7 (28)	3 (20)	3 (32)	4 (12)	17 (92)
Total: case (slice)	30 (149)	16 (76)	19 (162)	20 (102)	85 (489)

4.2.2. Implementations

For the pre-training phase, we conducted training for 1000 epochs with a batch size of 128. Stochastic Gradient Descent with momentum was employed as the optimizer, and the learning rate was set to 0.01. Then, for the target task training, the network was fine-tuned for 200 epochs with a batch size of 256, and the learning rate remained at 0.01. Details of the development environment can be found in Table 2.

4.2.3. Results

We conducted ablation experiments on our datasets for the classification of FLLs, which demonstrated the effectiveness of each component in our proposed model. The results of the ablation experiments are summarized in Table 6. To assess the predictive performance of the model, we calculated both the accuracy and the AUC. ResNet18 without pre-training was employed as the baseline (Model 1). When trained from scratch without weight initialization using pre-trained models, the prediction accuracy was $80.84\% \pm 2.91$, with an AUC of 0.709 ± 0.07 . Two self-supervised methods outperformed this baseline, demonstrating their effectiveness. Case discrimination (Model 2) improved accuracy by 6.2% and the AUC by 0.05, while phase shuffle (Model 3) led to accuracy gains of 4% and an AUC increase of 0.04. These results show that self-supervised learning with these pretext tasks effectively captures features and outperforms models trained from scratch. The multi-task pre-training model improved accuracy by 7.2% and the AUC by 0.08 compared to the baseline (Model 1), significantly outperforming single pre-training models (case discrimination or phase shuffle).

Table 6. Ablation study on classification of FLLs.

Model	Pre-Training		ACC (%)	AUC
	Case Discrimination	Phase Shuffle Prediction		
Model 1			$80.84\% \pm 2.91$	0.709 ± 0.07
Model 2	✓		$87.04\% \pm 2.27$	0.760 ± 0.05
Model 3		✓	$84.82\% \pm 1.99$	0.746 ± 0.06
Proposed	✓	✓	$88.06\% \pm 4.72$	0.791 ± 0.04

Bold indicates the highest values.

We also compared the proposed method with other different existing pre-training methods, including pre-training with ImageNet [21], self-supervised learning methods with rotation prediction [27], phase shuffle prediction [25], instance modeling [22], and case discrimination [24]. The comparison results are summarized in Table 7.

Table 7. Comparison with other transfer learning methods for classification of FLLs.

Models	Cyst	FNH	HCC	HEM	Average	AUC
Fine-tuning (ImageNet) [21]	$95.56\% \pm 2.84$	$83.53\% \pm 13.62$	$80.99\% \pm 11.69$	56.56 ± 32.61	$81.26\% \pm 1.20$	0.721 ± 0.06
Self-supervised (rotation) [27]	$96.66\% \pm 2.89$	$88.44\% \pm 7.51$	$78.81\% \pm 13.81$	60.30 ± 17.85	$81.84\% \pm 1.72$	0.713 ± 0.05
Self-supervised (phase shuffle prediction) [25]	$98.27\% \pm 1.42$	$86.22\% \pm 12.27$	$82.90\% \pm 5.84$	63.72 ± 12.68	$84.82\% \pm 1.99$	0.746 ± 0.06
Self-supervised (instance-level) [22]	$93.75\% \pm 1.56$	$87.46\% \pm 5.69$	$85.04\% \pm 5.58$	69.05 ± 11.99	$82.82\% \pm 3.98$	0.759 ± 0.05
Self-supervised (case-level) [24]	$90.29\% \pm 1.43$	$88.82\% \pm 7.21$	$88.74\% \pm 14.85$	80.15 ± 16.28	$87.04\% \pm 2.27$	0.760 ± 0.05
Self-supervised multi-task pre-training model (proposed)	$97.21\% \pm 4.66$	$93.00\% \pm 7.80$	$90.55\% \pm 11.58$	66.49 ± 8.21	$88.06\% \pm 4.72$	0.791 ± 0.04

Table 7 shows that self-supervised learning with medical images outperformed ImageNet pre-training. Among the methods of learning representations within images, our proposed phase shuffle prediction was more effective than rotation prediction. For representations between images, our proposed case-level discrimination surpassed instance-level discrimination. The proposed multi-task prediction model, which learns both representations within and between images, achieved the best results. The classification accuracy was $88.06\% \pm 4.72$ and the AUC was 0.791 ± 0.04 . The effectiveness and generalization of our proposed method are demonstrated.

5. Conclusions

In this paper, we propose a multi-task prediction model for predicting the early recurrence of hepatocellular carcinoma in multi-phase CT images before radical surgical resection by combining two types of our previously proposed self-supervised methods: phase shuffle prediction and case-level discrimination. The effectiveness and generalization of the proposed method were validated through two different experiments (predicting early recurrence and the classification of focal liver lesions). Both experiments demonstrated that the multi-task pre-training model outperforms existing pre-training (transfer learning) methods with natural images, single-task self-supervised pre-training, and DINOv2.

The strength of the proposed method is in its use of two multi-phase CT image-specific tasks (phase shuffle prediction and case-level discrimination) for self-supervised pre-training. Phase shuffle prediction is used to learn intra-image representations, and case-level discrimination is used to learn inter-image representations. The proposed method effectively learns both intra-image and inter-image feature representations simultaneously, resulting in higher prediction performance. Note that our previous proposed phase shuffle prediction and case-level discrimination can only learn intra-image or inter-image representations separately. However, the proposed method has a limitation: it is specific to multi-phase CT images and cannot be applied to non-multi-phase CT images, even though multi-phase CT images are widely used for liver cancer diagnosis. Developing self-supervised methods applicable to all medical images is part of our future work.

Author Contributions: Conceptualization, Y.-W.C. and H.D.; methodology, H.D. and J.S.; software, J.S. and H.D.; validation, J.S., Y.C., X.Z., G.Z. and R.K.J.; formal analysis, J.S.; investigation, J.S. and H.D.; resources, H.D.; data curation, Y.-W.C. and H.D.; writing—original draft preparation, J.S.; writing—review and editing, Y.-W.C. and J.S.; visualization, J.S. and H.D.; supervision, Y.-W.C.; project administration, J.S. and Y.-W.C.; funding acquisition, J.S. and Y.-W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Natural Science Foundation of Xiamen City, Fujian Province, China, under Grant No. 3502Z20227199, and in part by the Grant-in-Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under Grant Nos. 20KK0234 and No. 21H03470.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Ritsumeikan University under Approval No. BKC-LSMH-2021-037.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study. Written informed consent was obtained from the patients to publish this paper.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: The authors would like to thank Hongjie Hu of Run Run Shaw Hospital, Zhejiang University, China, for providing us with the medical dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Elsayes, K.M.; Kielar, A.Z.; Agrons, M.M.; Szklaruk, J.; Tang, A.; Bashir, M.R.; Mitchell, D.G.; Do, R.K.; Fowler, K.J.; Chernyak, V.; et al. Liver Imaging Reporting and Data System: An expert consensus statement. *J. Hepatocell. Carcinoma* **2017**, *4*, 29–39. [[CrossRef](#)] [[PubMed](#)]
2. Zhu, R.X.; Seto, W.K.; Lai, C.L.; Yuen, M.F. Epidemiology of hepatocellular carcinoma in the Asia-Pacific region. *Gut Liver* **2016**, *10*, 332–339. [[CrossRef](#)] [[PubMed](#)]
3. Thomas, M.B.; Zhu, A.X. Hepatocellular carcinoma: The need for progress. *J. Clin. Oncol.* **2005**, *23*, 2892–2899. [[CrossRef](#)] [[PubMed](#)]
4. Yang, T.; Lin, C.; Zhai, J.; Shi, S.; Zhu, M.; Zhu, N.; Lu, J.-H.; Yang, G.-S.; Wu, M.-C. Surgical resection for advanced hepatocellular carcinoma according to Barcelona Clinic Liver Cancer (BCLC) staging. *J. Cancer Res. Clin. Oncol.* **2012**, *138*, 1121–1129. [[CrossRef](#)] [[PubMed](#)]

5. Portolani, N.; Coniglio, A.; Ghidoni, S.; Giovanelli, M.; Benetti, A.; Tiberio, G.A.M.; Giulini, S.M. Early and late recurrence after liver resection for hepatocellular carcinoma: Prognostic and therapeutic implications. *Ann. Surg.* **2006**, *243*, 229–235. [[CrossRef](#)] [[PubMed](#)]
6. Shah, S.A.; Cleary, S.P.; Wei, A.C.; Yang, I.; Taylor, B.R.; Hemming, A.W.; Langer, B.; Grant, D.R.; Greig, P.D.; Gallinger, S. Recurrence after liver resection for hepatocellular carcinoma: Risk factors, treatment, and outcomes. *Surgery* **2007**, *141*, 330–339. [[CrossRef](#)] [[PubMed](#)]
7. Feng, J.; Chen, J.; Zhu, R.; Yu, L.; Zhang, Y.; Feng, D.; Kong, H.; Song, C.; Xia, H.; Wu, J.; et al. Prediction of early recurrence of hepatocellular carcinoma within the Milan criteria after radical resection. *Oncotarget* **2017**, *8*, 63299–63310. [[CrossRef](#)] [[PubMed](#)]
8. Cheng, Z.; Yang, P.; Qu, S.; Zhou, J.; Yang, J.; Yang, X.; Xia, Y.; Li, J.; Wang, K.; Yan, Z.; et al. Risk factors and management for early and late intrahepatic recurrence of solitary hepatocellular carcinoma after curative resection. *HPB* **2015**, *17*, 422–427. [[CrossRef](#)] [[PubMed](#)]
9. Hirokawa, F.; Hayashi, M.; Miyamoto, Y.; Asakuma, M.; Shimizu, T.; Komeda, K.; Inoue, Y.; Uchiyama, K. Outcomes and predictors of microvascular invasion of solitary hepatocellular carcinoma. *Hepatol. Res.* **2014**, *44*, 846–853. [[CrossRef](#)] [[PubMed](#)]
10. Sterling, R.K.; Wright, E.C.; Morgan, T.R.; Seeff, L.B.; Hoefs, J.C.; Di Bisceglie, A.M.; Dienstag, J.L.; Lok, A.S. Frequency of elevated hepatocellular carcinoma (HCC) biomarkers in patients with advanced hepatitis C. *Am. J. Gastroenterol.* **2012**, *107*, 64. [[CrossRef](#)] [[PubMed](#)]
11. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.P.M.; Granton, P.; Zegers, C.M.L.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [[CrossRef](#)] [[PubMed](#)]
12. Scapicchio, C.; Gabelloni, M.; Barucci, A.; Cioni, D.; Saba, L.; Neri, E. A deep look into radiomics. *La Radiol. Medica* **2021**, *126*, 1296–1311. [[CrossRef](#)] [[PubMed](#)]
13. Coppola, F.; Giannini, V.; Gabelloni, M.; Panic, J.; Defeudis, A.; Monaco, S.L.; Cattabriga, A.; Cocozza, M.A.; Pastore, L.V.; Polici, M.; et al. Radiomics and magnetic resonance imaging of rectal cancer: From engineering to clinical practice. *Diagnostics* **2021**, *11*, 756. [[CrossRef](#)] [[PubMed](#)]
14. Gabelloni, M.; Faggioni, L.; Borgheresi, R.; Restante, G.; Shortrede, J.; Tumminello, L.; Scapicchio, C.; Coppola, F.; Cioni, D.; Gómez-Rico, I.; et al. Bridging gaps between images and data: A systematic update on imaging biobanks. *Eur. Radiol.* **2022**, *32*, 3173–3186. [[CrossRef](#)] [[PubMed](#)]
15. Afshar, P.; Mohammadi, A.; Plataniotis, K.N.; Oikonomou, A.; Benali, H. From handcrafted to deep-learning-based cancer radiomics: Challenges and opportunities. *IEEE Signal Process. Mag.* **2019**, *36*, 132–160. [[CrossRef](#)]
16. Chen, Y.W.; Jain, L.C. *Deep Learning in Healthcare*; Springer: Berlin/Heidelberg, Germany, 2020.
17. Yasaka, K.; Akai, H.; Abe, O.; Kiryu, S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study. *Radiology* **2017**, *286*, 887–896. [[CrossRef](#)] [[PubMed](#)]
18. Liang, D.; Lin, L.; Hu, H.; Zhang, Q.; Chen, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W. Combining Convolutional and Recurrent Neural Networks for Classification of Focal Liver Lesions in Multi-Phase CT Images. In Proceedings of the Medical Image Computing and Computer Assisted Intervention (MICCAI), Granada, Spain, 16–20 September 2018; pp. 666–675.
19. Wang, W.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Chen, Q.; Liang, D.; Lin, D.; Hu, H.; Zhang, Q. Classification of Focal Liver Lesions Using Deep Learning with Fine-tuning. In Proceedings of the Digital Medicine and Image Processing (DMIP2018), Okinawa, Japan, 12–14 November 2018; pp. 56–60.
20. Wang, W.; Chen, Q.; Iwamoto, Y.; Han, X.; Zhang, Q.; Hu, H.; Lin, L.; Chen, Y.-W. Deep Learning-Based Radiomics Models for Early Recurrence Prediction of Hepatocellular Carcinoma with Multi-phase CT Images and Clinical Data. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4881–4884. [[CrossRef](#)]
21. Wang, W.; Chen, Q.; Iwamoto, Y.; Aonpong, P.; Lin, L.; Hu, H.; Zhang, Q.; Chen, Y.-W. Deep Fusion Models of Multi-Phase CT and Selected Clinical Data for Preoperative Prediction of Early Recurrence in Hepatocellular Carcinoma. *IEEE Access* **2020**, *8*, 139212–139220. [[CrossRef](#)]
22. Wu, Z.; Xiong, Y.; Yu, S.X.; Lin, D. Unsupervised feature learning via nonparametric instance discrimination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3733–3742.
23. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738.
24. Dong, H.; Iwamoto, Y.; Han, X.; Lin, L.; Hu, H.; Cai, X.; Chen, Y.W. Case Discrimination: Self-supervised Feature Learning for the classification of Focal Liver Lesions. In *Innovation in Medicine and Healthcare, Smart Innovation, Systems and Technology*; Chen, Y.W., Tanaka, S., Eds.; Springer: Singapore, 2021; pp. 241–249.
25. Song, J.; Dong, H.; Chen, Y.; Lin, L.; Hu, H.; Chen, Y.W. Deep Neural Network-Based Classification of Focal Liver Lesions Using Phase-Shuffle Prediction Pre-training. In Proceedings of the Innovation in Medicine and Healthcare, Smart Innovation, Systems and Technologies, Rome, Italy, 14–16 June 2023. [[CrossRef](#)]
26. Noroozi, M.; Favaro, P. Unsupervised learning of visual representations by solving jigsaw puzzles. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 69–84.

27. Spyros, G.; Praveer, S.; Nikos, K. Unsupervised representation learning by predicting image rotations. In Proceedings of the ICLR2018, Vancouver, BC, Canada, 30 April–3 May 2018.
28. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the International Conference on Machine Learning, Virtual Event, 13–18 July 2020; pp. 1597–1607.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Xu, Y.; Cai, M.; Lin, L.; Zhang, Y.; Hu, H.; Peng, Z.; Zhang, Q.; Chen, Q.; Mao, X.W.; Iwamoto, Y.; et al. PA-ResSeg: A Phase Attention Residual Network for Liver Tumor Segmentation from Multi-phase CT Images. *Med. Phys.* **2021**, *48*, 3752–3766. [[CrossRef](#)]
31. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. DINOv2: Learning Robust Visual Features without Supervision. *arXiv* **2023**, arXiv:2304.07193. Available online: <https://arxiv.org/abs/2304.07193> (accessed on 2 February 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.