Title: Dynamic contrast-enhanced computed tomography diagnosis of primary liver cancers using transfer learning of pre-trained convolutional neural networks: Is registration of multi-phasic images necessary?

Authors: Akira Yamada (ORCID: 0000-0002-4199-203X), Kazuki Oyama, Sachie Fujita, Eriko Yoshizawa, Fumihito Ichinohe, Daisuke Komatsu, Yasunari Fujinaga

Affiliation: Shinshu University School of Medicine, Department of Radiology

Corresponding author: Akira Yamada
Address: 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan
E-mail: a_yamada@shinshu-u.ac.jp
Telephone number: +81-263-37-2650
Fax number: +81-263-37-3087

Conflict of interest:
The authors declare that they have no conflict of interest.

Abstract

Purpose: To evaluate the effect of image registration on the diagnostic performance of transfer learning (TL) using pre-trained convolutional neural networks (CNNs) and three-phasic dynamic contrast-enhanced computed tomography (DCE-CT) for primary liver cancers.

Methods: We retrospectively evaluated 215 consecutive patients with histologically proven primary liver cancers, including six early, 58 well-differentiated, 109 moderately-differentiated, and 29 poorly-differentiated hepatocellular carcinomas (HCCs), and 13 non-HCC malignant lesions containing cholangiocellular components. We performed TL using various pre-trained CNNs and preoperative three-phasic DCE-CT images. Three-phasic DCE-CT images were manually registered to correct respiratory motion. The registered DCE-CT images were then assigned to the three color channels of an input image for TL: pre-contrast, early phase, and delayed phase images for the blue, red, and green channels, respectively. To evaluate the effects of image registration, the registered input image was intentionally misaligned in the three color channels by pixel shifts, rotations, and skews with various degrees. The diagnostic performances (DP) of the pre-trained CNNs after TL in the test set were compared by three general radiologists (GRs) and two experienced abdominal radiologists (ARs). The effects of misalignment in the input image and the type of pre-trained CNN on the DP were statistically evaluated.

Results: The mean DPs for histological subtype classification and differentiation in primary malignant liver tumors on DCE-CT for GR and AR were 39.1%, and 47.9%, respectively.

The highest mean DPs for CNNs after TL with pixel shifts, rotations, and skew misalignments were 44.1%, 44.2%, and 43.7%, respectively. Two-way analysis of variance revealed that the DP is significantly affected by the type of pre-trained CNN ($P = 0.0001$), but not by misalignments in input images other than skew deformations.

Conclusion: TL using pre-trained CNNs is robust against misregistration of multi-phasic images, and comparable to experienced ARs in classifying primary liver cancers using three-phasic DCE-CT.

Key words:

Primary liver cancer; Dynamic contrast-enhanced computed tomography; Transfer learning; Convolutional neural network; Registration

Introduction

Diagnostic imaging of primary liver cancers is important, because primary liver cancers are often treated through imaging diagnosis only, without pathological diagnosis [1]. Furthermore, the therapeutic strategy can differ significantly depending on the pathological subtype. For example, it has been reported that macro-trabecular and compact types, which are common in poorly differentiated hepatocellular carcinoma (HCC), exhibit higher rates of recurrence after transarterial catheter embolization than after hepatectomy or radio frequency ablation [2].

A convolutional neural network (CNN) is a machine learning algorithm that has attracted considerable attention in diagnostic imaging, because it can perform as well as or better than humans in image classification tasks [3]. The advantage of transfer learning (TL) using pre-trained CNNs compared to usual deep learning algorithms utilizing untrained CNNs is that TL can achieve a high classification performance with a relatively small dataset [3]. The usefulness of TL with pre-trained CNNs for liver disease has been demonstrated by several studies [4,5]. However, the effect of misregistration of input images on the diagnostic performance of CNNs has not been fully investigated. In particular, this is an important issue for radiologists when multiple images are employed as input images for a CNN, such as in dynamic contrast-enhanced computed tomography (DCE-CT), because respiratory misregistration frequently occurs in DCE-CT imaging of the liver. Furthermore, manual registration is laborious and time-consuming for radiologists.

The purpose of this study was to evaluate the effects of image registration on the diagnostic performance of TL using a pre-trained CNN and three-phasic DCE-CT for primary liver cancers.

Materials and Methods

Subjects

We retrospectively evaluated 215 consecutive patients (median age = 70 years; age range = 34–85 years; male:female = 165:52) with histologically proven primary liver cancers in a single institute (Shinshu University Hospital, Matsumoto, Japan) from 2005 to 2010, including six early (eHCC), 58 well-differentiated (wHCC), 109 moderately differentiated (mHCC), and 29 poorly differentiated (pHCC) HCCs, and 13 non-HCC malignant lesions containing cholangiocellular components (CCC). Written informed consent was obtained from all patients when preoperative DCE-CT was performed. The patients who did not undergo preoperative DCE-CT within 1 month before hepatectomy were excluded from the study.

DCE-CT protocol

Three-phasic DCE-CT (a pre- contrast phase and two phases after intravenous contrast agent injection) was performed at 40 (early phase) and 130 s (delayed phase) after injection, using a 64- row CT scanner. The scan parameters were as follows: the range was whole abdomen from the upper level of the diaphragm; the tube voltage was 120 kVp; the tube current was 500 mA; the matrix had 512 × 512 pixels; the field of view was 320 × 320 mm; the size of collimation was 0.625 mm; and the reconstruction thickness was 2.5 mm. A non-ionic iodinated contrast agent (Iopamiron 370 mg/mL; Bayer Healthcare, Berlin, Germany) was administered intravenously through a 22- gauge catheter in the median cubital vein. The total dose was 100 mL, and the rate of injection was 3 mL/s.

TL using pre-trained CNNs

TL was performed using various pre-trained CNNs (Alexnet, VGG-16, VGG-19, GoogLeNet, Inception-v3, ResNet-50, and ResNet-101) and preoperative three-phasic DCE-CT images at the maximal cross-sectional lesion area. In the image presentation in TL, three-phasic DCE-CT DICOM images were manually registered to correct respiratory motion by an abdominal radiologist (A.Y.) who has 18 years of diagnostic experience. The registered three-phasic DCE-CT DICOM images were then cropped at the hepatic lesion, and assigned to the three color channels of an input JPEG image for TL as follows: pre-contrast, early phase, and delayed phase images for the blue, red, and green channels, respectively. The window level and width for DICOM images were fixed as 80 and 350 Hounsfield units, respectively. The image size was transformed according to the utilized pre-trained CNN (227 × 227 pixels for Alexnet, 299 × 299 pixels for Inception-v3, and 224 × 224 pixels for the other CNNs). To evaluate the effect of registration, manually registered input images were intentionally misaligned to various degrees in the three color channels by pixel shifts (0, 1, 2, 4, 8, 16, and 32 pixels), rotations (0, 1, 2, 4, 8, 16, and 32 degrees), and skews (0%, 1%, 2%, 4%, 8%, 16%, and 32%) (**Fig. 1**). The image with 0 pixel shift, 0 degree rotation, and 0% skew represents the original registered image. The input images with specific degrees of misalignment were divided into training (70%) and test (30%) sets, such that the proportion of histological subtypes was the same in both sets. In the transfer learning procedure, the final three layers of the pre-trained CNNs, originally developed for the ImageNet dataset (1,000 classes), were replaced by a fully connected layer, a softmax layer, and a classification output layer (with five classes in this study, including eHCC, wHCC, mHCC, pHCC, and CCC). To learn faster in the new layers than in the transferred layers, the initial learning rate was set to a small value (0.0001). Meanwhile, the learning rate factor for the new fully connected layer was set to a large value (20). The mini-batch size was set to 10. A classification test was performed on the pre-trained CNNs after TL with 500 iterations of training using the five-fold cross validation method. The mean value of the obtained results was utilized for a statistical analysis. All the procedures were carried out using MATLAB software (2018a, MathWorks, Natick, MA, USA).

Statistical analysis

The diagnostic performances (DP = [number of correctly classified cases] / [total number of cases] × 100) of the pre-trained CNNs after TL in the test set were compared by three general radiologists (GRs) and two experienced abdominal radiologists (ARs). The observer agreement was tested by weighted kappa. The effects of misalignments (pixel shifts, rotations, and skews) in the input image and the type of pre-trained CNN on DP were statistically evaluated by two-way analysis of variance (ANOVA) and a multiple comparison test using Turkey's honest significant difference criterion. A probability value of less than 0.05 or no overlapping in a 95% confidence interval were regarded as statistically significant. All the procedures were carried out using MATLAB software (2018a, MathWorks, Natick, MA, USA).

Results

The mean DPs for the classification of histological subtype and differentiation in primary malignant liver tumors on DCE-CT for GR and AR were 39.1% and 47.9%, respectively. The mean weighted kappa between observers was 0.92 (range 0.90-0.95).

Two-way ANOVA revealed that the type of pre-trained CNN ($P < 0.0001$) had a significant effect on the DP, whereas the degree of misalignment in input images for TL ($P = 0.17$) when pixel-shift misalignment was applied (**Table 1**) did not. A multi-comparison revealed that GoogLeNet exhibited the highest mean DP (44.1%) using input images misaligned by pixel shift. Statistical significance was observed between GoogLeNet and some other pre-trained CNNs (VGG-16, VGG-19, ResNet-50, and ResNet-101) (**Fig. 2**).

Significant effects on the DP were observed for the type of pre-trained CNN ($P < 0.0001$) and degree of misalignment of input images for TL ($P = 0.001$) when a rotation misalignment was applied (**Table 2**). However, a multi-comparison revealed that there was no significant difference in the DPs of CNNs between registered and misaligned input images (**Fig. 3**). GoogLeNet exhibited the highest mean DP (44.2%) using input images misaligned by rotation. Statistical significance was observed between GoogLeNet and some other pre-trained CNNs (Alexnet, VGG-16, VGG-19, and ResNet-50) (**Fig. 3**).

Two-way ANOVA revealed that the type of pre-trained CNN ($P < 0.0001$) and the degree of misalignment in input images for TL ($P < 0.0001$) had a significant effect on the DP when skew misalignment was applied (**Table 3**). There was a significant decrease in the DPs of CNNs when the skew ratios in input images were 4% and 8% (**Fig. 4**). Inception-v3 and GoogLeNet exhibited higher mean DPs (43.7% and 43.4%) even if skew misalignment was applied. Statistical significance was observed between these two pre-trained CNNs and the others (Alexnet, VGG-16, VGG-19, ResNet-50, and ResNet-101) (**Fig. 4**).

Discussion

Our results demonstrate the high diagnostic performance of TL using a pre-trained CNN, which is comparable to experienced ARs in classifying primary liver cancers using three-phasic DCE-CT. Our results also clarify that TL using particular pre-trained CNNs (GoogLeNet and Inception-v3) was robust against misregistration of DCE-CT images, even if the pre-trained CNNs were trained using RGB images without misalignment in the color channels [6]. One of the common features between GoogLeNet and Inception-v3 is the inception architecture, which enables efficient parameter reduction and allows for training high-quality networks on relatively modest-sized training sets [7,8]. This architecture may relate to the robustness against misregistration and higher diagnostic performance of TL using multi-phasic DCE-CT images. However, further study is required to confirm this.

Our findings in this study can accelerate the application of TL using pre-trained CNNs, not only in dynamic contrast-enhanced study, but also for multi-parametric imaging, such as magnetic resonance imaging. This is because this approach can be more easily applied in a clinical setting, without time-consuming registration procedures and using smaller training datasets compared to conventional deep learning algorithms using untrained CNNs [3]. However, special caution should be exercised when applied this approach to hollow organs, such as the heart or alimentary tract, which are frequently accompanied by skew-type deformations, because some CNNs were not robust to skew misregistration.

In conclusion, TL using pre-trained CNNs is robust against misregistrations, and comparable to experienced ARs in the classification of primary liver cancers using three-phasic DCE-CT. Therefore, there is no need for the correction of misregistrations for TL using pre-trained CNNs.

References

1) Torzilli G, Minagawa M, Takayama T, Inoue K, Hui AM, Kubota K, Ohtomo K, Makuuchi M. (1999) Accurate preoperative evaluation of liver mass lesions without fine-needle biopsy. Hepatology. 30:889-93.

2) Okabe H, Yoshizumi T, Yamashita YI, Imai K, Hayashi H, Nakagawa S, Itoh S, Harimoto N, Ikegami T, Uchiyama H, Beppu T, Aishima S, Shirabe K, Baba H, Maehara Y. (2018) Histological architectural classification determines recurrence pattern and prognosis after curative hepatectomy in patients with hepatocellular carcinoma. PLoS One. 13:e0203856. https://doi.org/10.1371/journal.pone.0203856.

3) Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. (2018) Deep learning with convolutional neural network in radiology. Jpn J Radiol. 36:257-272. https://doi.org10.1007/s11604-018-0726-3.

4) Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michałowski Ł, Paluszkiewicz R, Ziarkiewicz-Wróblewska B, Zieniewicz K, Sobieraj P, Nowicki A. (2018) Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. Int J Comput Assist Radiol Surg. 13:1895-1903. https://doi.org/10.1007/s11548-018-1843-2.

5) Yu Y, Wang J, Ng CW, Ma Y, Mo S, Fong ELS, Xing J, Song Z, Xie Y, Si K, Wee A, Welsch RE, So PTC, Yu H. (2018) Deep learning enables automated scoring of liver fibrosis stages. Sci Rep. 8:16016. https://doi.org/10.1038/s41598-018-34300-2.

6) Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. (2009) ImageNet: A large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2009.5206848.

7) Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. (2015) Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2015.7298594.

8) Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. (2016) Rethinking the inception architecture for computer vision. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2016.308.

Figure Captions

Fig. 1 Illustration of input image preparation from three-phasic DCE-CT for transfer learning using a pre-trained CNN. Three-phasic DCE-CT images were manually registered to correct respiratory motion (misaligned value = 0). The registered three-phasic DCE-CT images were then assigned into the three color channels of an input image for transfer learning as follows: pre-contrast, early phase, and delayed phase images for the blue, red, and green channels, respectively. The manually registered input images were intentionally misaligned in the three color channels by pixel shifts, rotations, and skews with various misaligned values, to generate misaligned input images for transfer learning. DCE-CT = dynamic contrast-enhanced computed tomography, CNN = convolutional neural network

Fig. 2 Multi-comparison of diagnostic performances of CNNs according to pixel-shift values in misaligned input images and the type of pre-trained CNN. Circles and bars indicate the mean values and 95% confidence intervals, respectively. There was no significant difference between diagnostic performances of CNNs for registered and misaligned input images. GoogLeNet exhibited the highest mean diagnostic performance (44.1%) using input images by misaligned pixel shift. Statistical significance was observed between GoogLeNet and some other pre-trained CNNs (VGG-16, VGG-19, ResNet-50, and ResNet-101). CNN = convolutional neural network

Fig. 3 Multi-comparison of diagnostic performances of CNNs according to the rotation value in misaligned input images and the type of pre-trained CNN. Circles and bars indicate the mean values and 95% confidence intervals, respectively. There was no significant difference between the diagnostic performances of CNNs for registered and misaligned input images. GoogLeNet exhibited the highest mean diagnostic performance (44.2%) using input images misaligned by rotation. Statistical significance was observed between GoogLeNet and some other pre-trained CNNs (Alexnet, VGG-16, VGG-19, and ResNet-50). CNN = convolutional neural network

Fig. 4 Multi-comparison of diagnostic performances of CNNs according to skew values in misaligned input images and the type of pre-trained CNN. Circles and bars indicate the mean values and 95% confidence intervals, respectively. There was a significant decrease in the diagnostic performances of CNNs when skew values in misaligned input images were 4% and

8%. Inception-v3 and GoogLeNet exhibited higher mean diagnostic performances (43.7% and 43.4%) using input images misaligned by skewing. Statistical significance was observed between these two pre-trained CNNs and the others (Alexnet, VGG-16, VGG-19, ResNet-50, and ResNet-101). CNN = convolutional neural network

Figure1

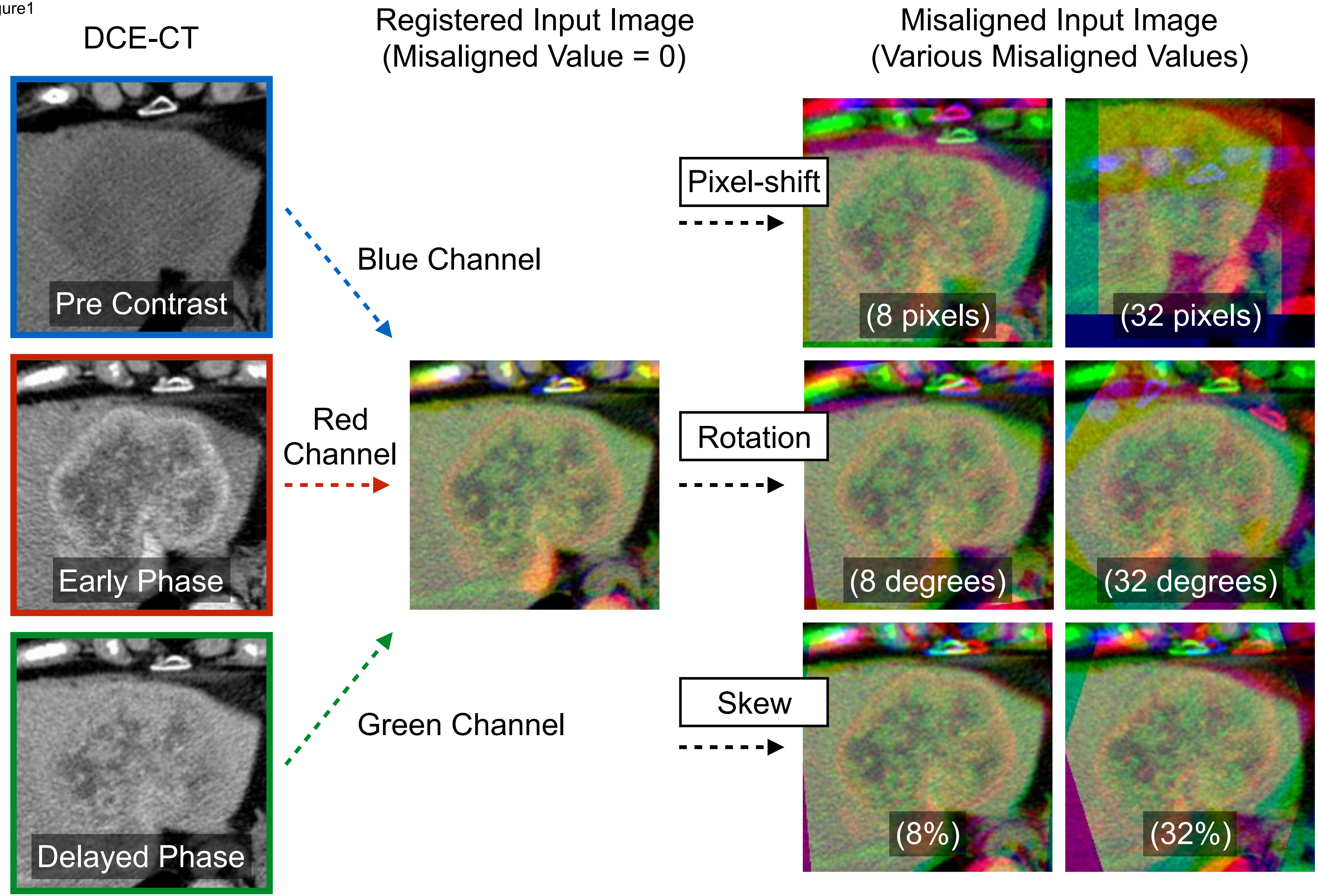

DCE-CT
Registered Input Image
(Misaligned Value = 0)
Misaligned Input Image
(Various Misaligned Values)
Pre Contrast
Early Phase
Delayed Phase
Blue Channel
Red
Channel
Green Channel
Pixel-shift
Rotation
Skew
(8 pixels)
(32 pixels)
(8 degrees)
(32 degrees)
(8%)
(32%)

Figure2

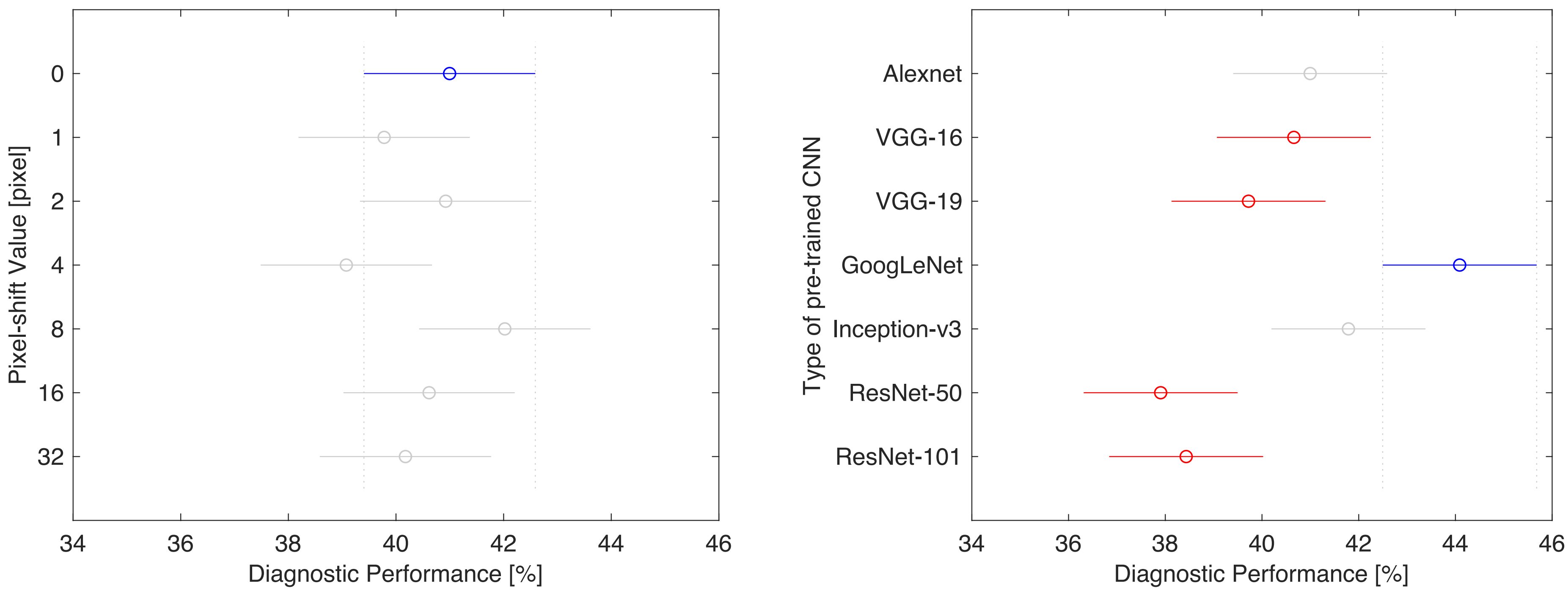
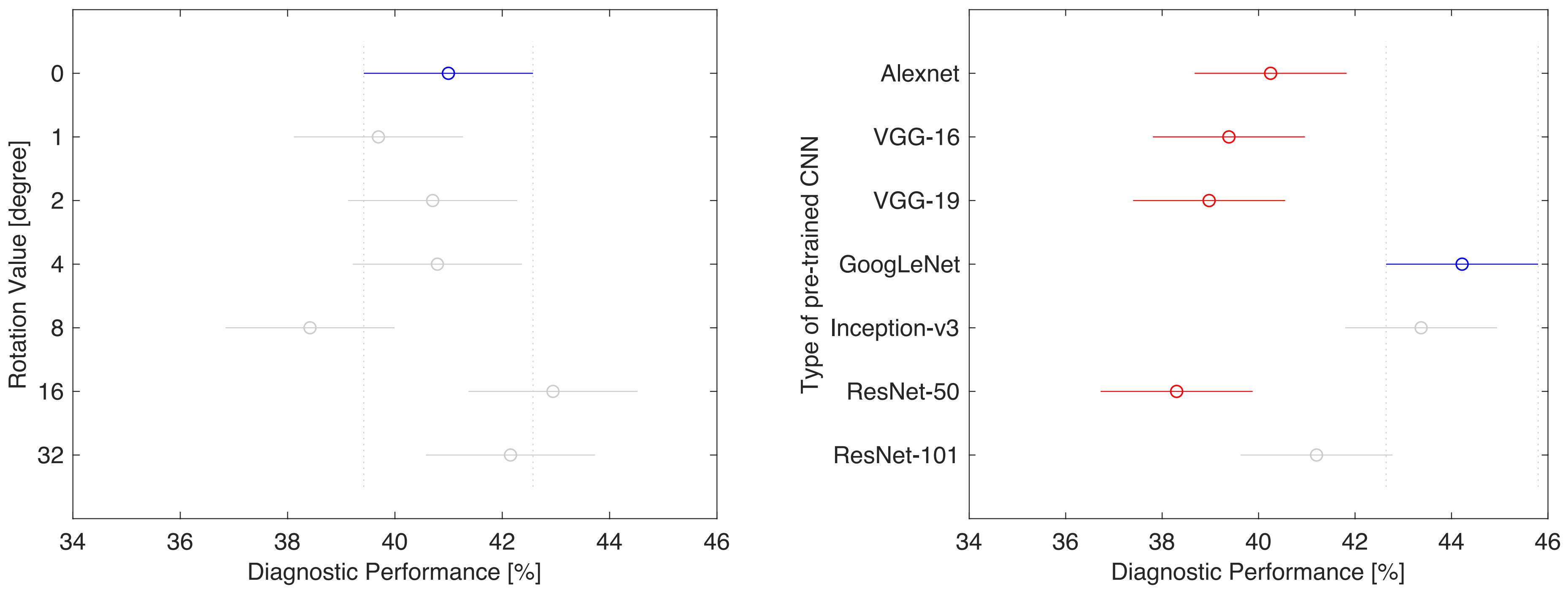

Pixel-shift Value [pixel]
0
1
2
4
8
16
32
34
36
38
40
42
44
46
Diagnostic Performance [%]
Type of pre-trained CNN
Alexnet
VGG-16
VGG-19
GoogLeNet
Inception-v3
ResNet-50
ResNet-101
34
36
38
40
42
44
46
Diagnostic Performance [%]

Figure3



Rotation Value [degree]
0
1
2
4
8
16
32
34
36
38
40
42
44
46
Diagnostic Performance [%]
Type of pre-trained CNN
Alexnet
VGG-16
VGG-19
GoogLeNet
Inception-v3
ResNet-50
ResNet-101
34
36
38
40
42
44
46
Diagnostic Performance [%]

Figure4

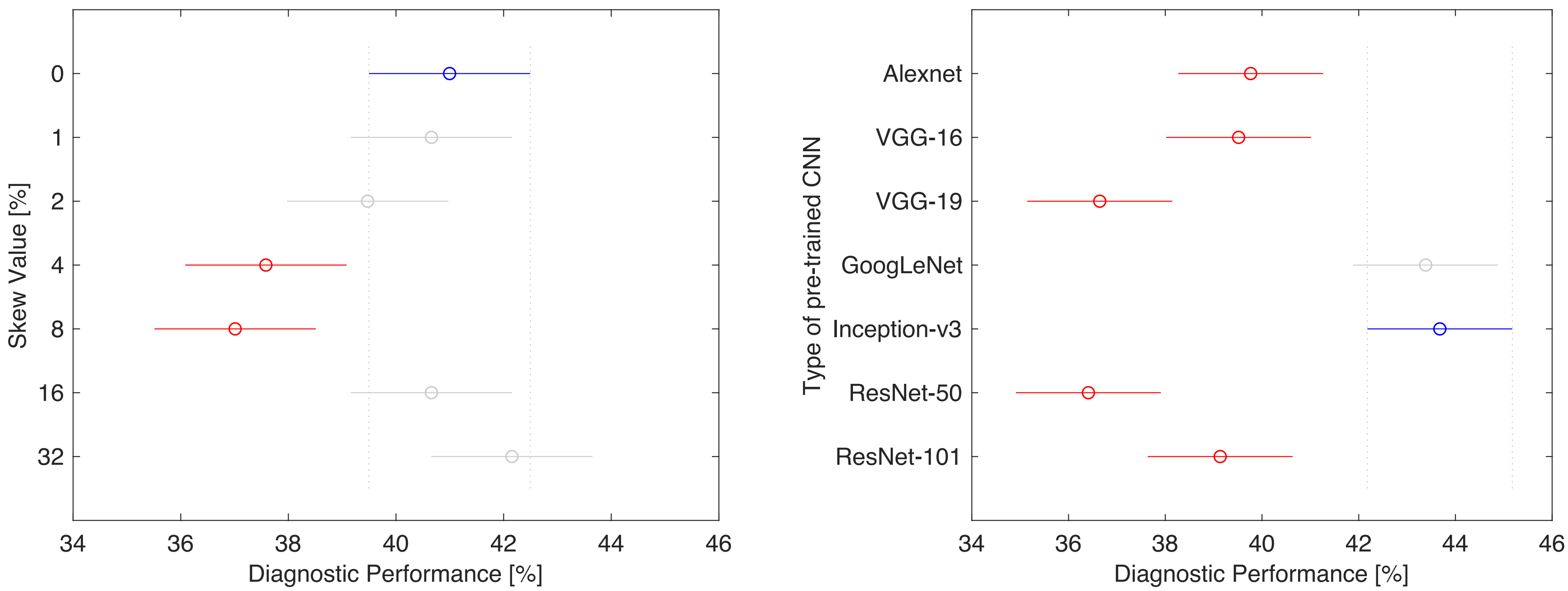

Skew Value [%]
0
1
2
4
8
16
32
34
36
38
40
42
44
46
Diagnostic Performance [%]
Type of pre-trained CNN
Alexnet
VGG-16
VGG-19
GoogLeNet
Inception-v3
ResNet-50
ResNet-101
34
36
38
40
42
44
46
Diagnostic Performance [%]

Table 1. Mean diagnostic performances of pre-trained CNNs after transfer learning in classification of primary liver malignant tumors using three-phasic DCE-CT misaligned by pixel shifts

Pixel shift	Type of pre-trained CNN						
[pixel]	Alexnet	VGG-16	VGG-19	GoogLeNet	Inception-v3	ResNet-50	ResNet-101
0	42.7%	42.8%	37.7%	42.5%	40.8%	38.6%	41.9%
1	39.4%	37.8%	37.8%	44.0%	44.9%	33.8%	40.6%
2	44.0%	40.9%	38.5%	44.3%	40.3%	38.8%	39.7%
4	36.3%	39.1%	38.2%	44.3%	38.5%	41.5%	35.7%
8	41.8%	41.8%	47.1%	45.5%	43.4%	36.3%	38.2%
16	43.7%	40.9%	37.8%	47.1%	41.2%	39.7%	33.8%
32	39.1%	41.2%	40.9%	40.9%	43.4%	36.6%	39.1%

The image with 0 pixel shift represents the original registered image. Only images with the same pixel shift were used for the training and validation. Diagnostic performances (DP = [number of correctly classified cases] / [total number of cases] × 100) of various pre-trained CNNs after transfer learning in the test set are shown. Two-way ANOVA revealed that the type of pre-trained CNN ($P < 0.0001$) has a significant effect on the DP, but pixel-shift values in the input images for TL do not ($P = 0.17$). CNN = convolutional neural network, ANOVA = analysis of variance

Table 2. Mean diagnostic performances of pre-trained CNNs after transfer learning in classification of primary liver malignant tumors using three-phasic DCE-CT misaligned by rotation

Rotation	Type of pre-trained CNN						
[degree]	Alexnet	VGG-16	VGG-19	GoogLeNet	Inception-v3	ResNet-50	ResNet-101
0	42.7%	42.8%	37.7%	42.5%	40.8%	38.6%	41.9%
1	38.5%	38.8%	39.7%	41.2%	41.8%	37.8%	40.0%
2	39.7%	42.5%	41.2%	43.1%	43.7%	36.6%	38.2%
4	39.7%	37.8%	43.7%	47.4%	45.5%	32.3%	39.1%
8	40.6%	35.1%	34.2%	42.8%	42.5%	32.3%	41.5%
16	40.6%	39.1%	36.0%	48.3%	47.7%	46.2%	42.8%
32	40.0%	39.7%	40.3%	44.3%	41.5%	44.3%	44.9%

The image with 0 degree rotation represents the original registered image. Only images with the same rotation were used for the training and validation. Diagnostic performances (DP = [number of correctly classified cases] / [total number of cases] × 100) of various pre-trained CNNs after transfer learning in the test set are shown. Two-way ANOVA revealed that the type of pre-trained CNN ($P < 0.0001$) and rotation value in the input image for TL ($P = 0.001$) have significant effects on the DP. CNN = convolutional neural network, ANOVA = analysis of variance

Table 3. Mean diagnostic performances of pre-trained CNNs after transfer learning in classification of primary liver malignant tumors using three-phasic DCE-CT by misaligned skewing

Skew	Type of pre-trained CNN						
[%]	Alexnet	VGG-16	VGG-19	GoogLeNet	Inception-v3	ResNet-50	ResNet-101
0	42.7%	42.8%	37.7%	42.5%	40.8%	38.6%	41.9%
1	37.8%	41.2%	36.9%	43.7%	47.4%	36.9%	40.6%
2	35.4%	39.7%	38.8%	43.4%	45.5%	35.1%	38.5%
4	35.7%	38.2%	30.5%	44.3%	40.3%	35.4%	38.8%
8	39.1%	37.2%	35.7%	40.6%	36.6%	30.5%	39.4%
16	42.8%	40.0%	39.4%	45.2%	48.6%	36.6%	32.0%
32	44.9%	37.5%	37.5%	44.0%	46.5%	41.8%	42.8%

The image with 0% skew represents the original registered image. Only images with the same skewing were used for the training and validation. Diagnostic performances (DP = [number of correctly classified cases] / [total number of cases] × 100) of various pre-trained CNNs after transfer learning in the test set are shown. Two-way ANOVA revealed that the type of pre-trained CNN ($P < 0.0001$) and the skew value in an input image for TL ($P < 0.0001$) have significant effects on the DP. CNN = convolutional neural network, ANOVA = analysis of variance