



OPEN

Detection of Hepatocellular Carcinoma in Contrast-Enhanced Magnetic Resonance Imaging Using Deep Learning Classifier: A Multi-Center Retrospective Study


Junmo Kim¹  , Ji Hye Min², Seon Kyoung Kim² , Soo-Yong Shin^{3,4,5} & Min Woo Lee^{1,2,5} 

Hepatocellular carcinoma (HCC) is one of the most common malignant tumors and a leading cause of cancer-related death worldwide. We propose a fully automated deep learning model to detect HCC using hepatobiliary phase magnetic resonance images from 549 patients who underwent surgical resection. Our model used a fine-tuned convolutional neural network and achieved 87% sensitivity and 93% specificity for the detection of HCCs with an external validation data set (54 patients). We also confirmed whether the lesion detected by our deep learning model is a true lesion using a class activation map.

Primary liver cancer is the fifth most common malignant tumor worldwide and the third most common cause of cancer-related mortality, with hepatocellular carcinoma (HCC) accounting for 85–90% of primary liver cancers^{1,2}. Many practice guidelines for HCC management state that tumor size is one of the important prognostic factors in patients with HCC, along with liver function and patient performance status^{3,4}. Therefore, earlier detection and diagnosis of HCC would be of paramount importance for better survival outcomes after treatment.

In terms of the diagnosis of HCC, magnetic resonance imaging (MRI) provides higher sensitivity than computed tomography (CT). Currently, gadoxetic acid-enhanced liver MRI is widely used for HCC diagnosis and has shown significantly higher sensitivity than MRI performed with other contrast agents⁵. This improved sensitivity of gadoxetic acid-enhanced MRI is mainly attributed to hepatobiliary phase images as most HCCs (80–90%) are hypointense in this phase^{6,7}. However, the per lesion sensitivity for HCC on gadoxetic acid-enhanced MRI was 87% (95% confidence interval: 83–92%) in a recent meta-analysis⁵. This implies that computational decision-support tools may play an important role in improving the diagnostic performance of gadoxetic acid-enhanced MRI⁸.

Deep learning has shown remarkable results in the field of computer vision⁹. Deep learning-based methods have also demonstrated that they are well suited for recognition and classification of medical images¹⁰ and they can be used as an effective screening tool in medical image analysis¹¹. Therefore, deep learning systems can be an auxiliary diagnostic system for the diagnosis of HCC, as well. To our knowledge, however, there are no deep learning-based HCC detection systems using liver MRI in the English literature. Therefore, the purpose of this study was to develop a fully automated deep learning model to detect HCC using hepatobiliary phase MR images in patients who underwent surgical resection for HCC and evaluate its performance in detecting HCC on liver MRI compared to human readers.

¹Department of Health and Sciences and Technology, SAIHST, Sungkyunkwan University, 81, Irwon-ro, Gangnam-gu, Seoul, 06351, Korea. ²Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81, Irwon-ro, Gangnam-gu, Seoul, 06351, Korea. ³Department of Digital Health, SAIHST, Sungkyunkwan University, 115, Irwon-ro, Gangnam-gu, Seoul, 06351, Korea. ⁴Big Data Research Center, Samsung Medical Center, 81, Irwon-ro, Gangnam-gu, Seoul, 06351, Korea. ⁵These authors jointly supervised this work: Soo-Yong Shin and Min Woo Lee.  e-mail: skyjunmo@gmail.com; leeminwoo0@gmail.com

Regularization Layer	Kernel size	Activation function	Optimizer	Number of Epochs	Accuracy
BN	2×2	ReLU ^a	Adam ^b	46	93.7%
BN and Dropout (0.1) ^c	2×2	ReLU	Adam	28	74.6%
BN and Dropout (0.2)	2×2	ReLU	Adam	165	92.7%
BN and Dropout (0.3)	2×2	ReLU	Adam	48	86.6%
BN and Dropout (0.4)	2×2	ReLU	Adam	46	85.0%
BN and Dropout (0.5)	2×2	ReLU	Adam	55	73.9%
Dropout (0.1)	2×2	ReLU	Adam	59	86.6%
Dropout (0.2)	2×2	ReLU	Adam	116	88.2%
Dropout (0.3)	2×2	ReLU	Adam	96	87.7%
Dropout (0.4)	2×2	ReLU	Adam	112	87.9%
Dropout (0.5)	2×2	ReLU	Adam	28	72.1%

Table 1. Comparison results of combination of batch normalization (BN) and dropout rate. BN only showed the best accuracy. ^aReLU: rectified linear unit. ^bAdam: Adam optimizer. ^cDropout (0.1) indicates a dropout rate of 0.1.

Activation function	Number of Epochs	Accuracy
LeakyReLU (0.1) ^a	37	84.2%
PReLU ^b	28	90.4%
ELU (0.1) ^c	73	91.6%
ReLU ^d	46	93.7%

Table 2. Comparison of diverse activation functions. ReLU showed the best accuracy. ^aLeakyReLU: leaky rectified linear function. ^bPReLU: parametric ReLU. ^cELU: exponential linear unit. ^dReLU: rectified linear unit.

Optimizer	Learning Rate	Number of Epochs	Validation loss	Accuracy
AdaGrad ^a	0.001	23	0.83	55.7%
Adam ^b	0.001	46	0.18	93.7%
RMSprop ^c	0.001	62	0.18	91.4%
SGD ^d	0.001	14	0.98	58.7%

Table 3. Comparison of diverse optimizers. Optimization with Adam had the best performance. ^aAdaGrad: adaptive gradient algorithm. ^bAdam: a method for stochastic optimization. ^cRMSprop: a mini-batch version of rprop. ^dSGD: stochastic gradient descent.

Results

Our CNN architecture. Tables 1–3 show the experimental results of combinations of heuristically chosen hyperparameters to optimize the CNN architecture for HCC detection in liver MRI. Table 1 shows the results of the comparison of batch normalization (BN)¹² and dropout¹³ to prevent overfitting. Since training was terminated when there was no improvement in the accuracy of validation datasets within 10 epochs, the number of epochs in each case were different. As a result, BN only showed the best performance.

To solve the vanishing gradient problem, various activation functions, including the leaky rectified linear function (LeakyReLU)¹⁴, a Parametric Rectified Linear Unit (PReLU)¹⁵, Exponential Linear Units (ELUs)¹⁶, and Rectified Linear Unit (ReLU)¹⁷ were compared. Among these activation functions, ReLU showed the best performance (Table 2).

To minimize information loss, we prefixed the stride as 1 and then changed the kernel size from 2×2 to 7×7 . As shown in Fig. 1., the 2×2 kernel achieved the minimum validation loss. To choose the right optimizer for the lowest possible error and steady learning speed, we compared optimization functions, including the adaptive gradient algorithm (AdaGrad)¹⁸, a method for stochastic optimization (Adam)¹⁹, a mini-batch version of rprop (RMSprop)²⁰, and stochastic gradient descent (SGD)²¹. The Adam optimizer was found to be the most accurate optimization function that affected learning speed and probability (Table 3).

The CNN components included widely used components such as the convolution filter, pooling, BN, dropout, padding, activation functions, and optimization functions. Each component also had diverse parameters, such as convolution filter size, pooling method, and stride size. Figure 2 shows the designed CNN architecture.

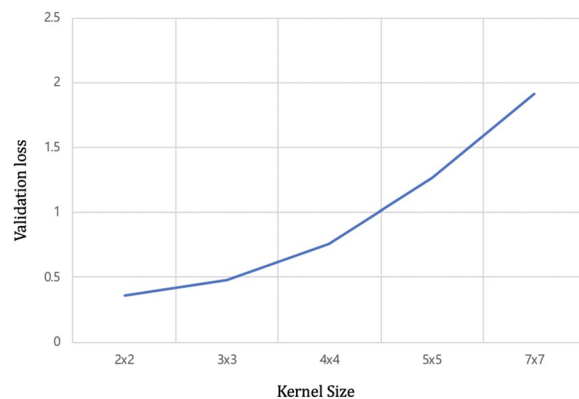


Figure 1. Comparison results of diverse kernel size. Kernel size with 2×2 showed the smallest loss.

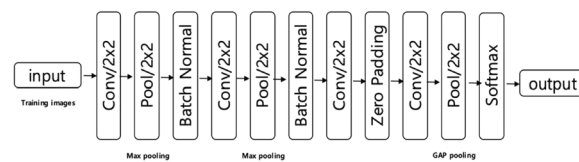


Figure 2. CNN structure of our model for HCC detection in MR images.

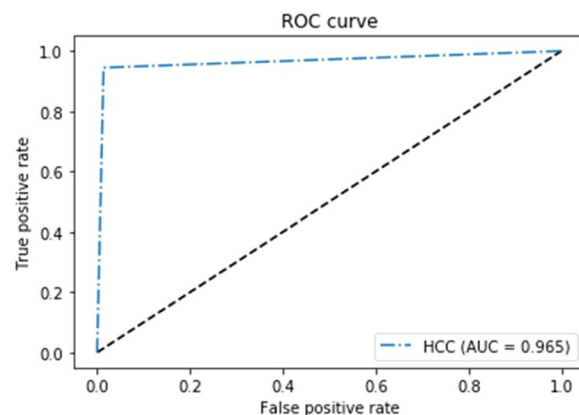


Figure 3. ROC analysis for the proposed model in test datasets.

Performance. The optimized CNN architecture achieved 94% sensitivity, 99% specificity, and 0.97 area under curve (AUC) for HCC cases in the test dataset (Fig. 3) and achieved 87% sensitivity and 93% specificity and an AUC of 0.90 for external validation datasets (Fig. 4). The mean size of HCCs that were missed by the less experienced radiologist but detected by our model was 1 ± 0.2 cm (Fig. 5). Therefore, our model seemed to have advantages over the performance of the less experienced radiologist in detecting very small HCCs (Fig. 5). However, our model showed false positive detections including the gallbladder, blood vessels, and heart (Fig. 6). The overall performance was not significantly different between our model and the less experienced radiologist (Table 4).

In terms of image classification time, our model was much faster than human readers, regardless of the degree of experience (Table 4). Our model took 0.03 seconds to classify one image and the average image classification time of 100 images in a single patient was 3.4 seconds using a commercial PC (3.8 GHz Intel Core i5, 16 GB RAM, Radeon Pro 580 8 G). We used the CPU version considering the PC without an expensive GPU usually used in the doctor's office. In comparison, it took 0.18 seconds to classify one image and average image classification time of 100 images in a single patient was 18 seconds in both radiologists.

Discussion

Recently, deep learning has gained attention in the field of medical imaging, including radiologic imaging^{8,22–24}. In this study, we performed deep learning of the hepatobiliary phase images of 92,645 gadoxetic acid-enhanced MR images using a fine-tuned CNN. External validation using the training generation model for 4,537 images obtained by various MR scanners from multiple vendors showed an 87% sensitivity for HCC, 93% specificity, and

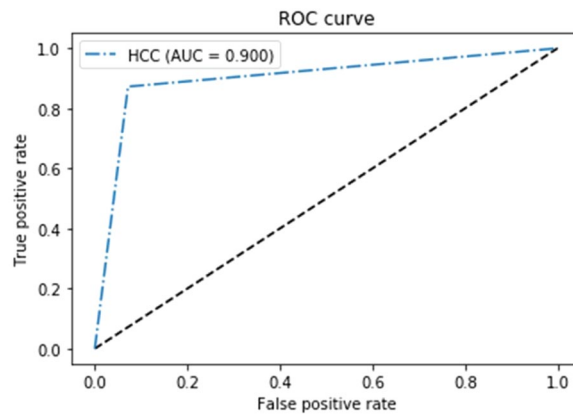


Figure 4. ROC analysis for the proposed model in external validation datasets.

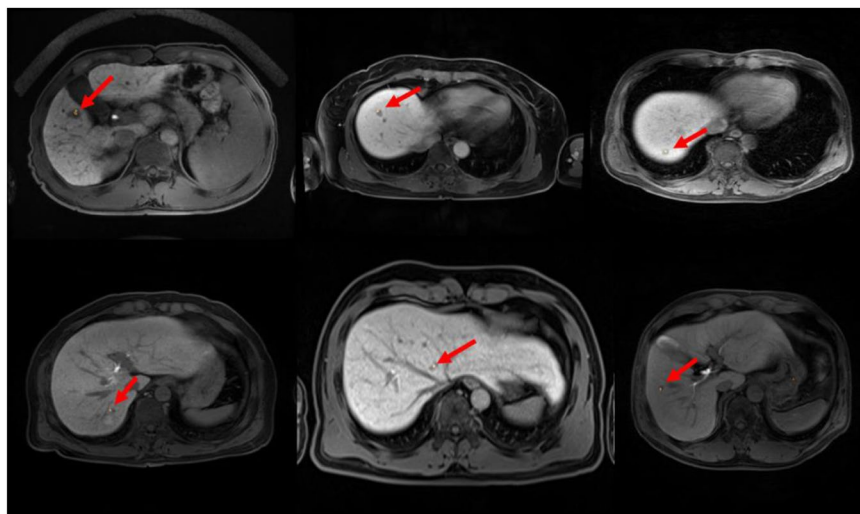


Figure 5. True positive detection (arrow) of HCC by our trained model showing the area of interest of the trained model using CAM method. Although the expert radiologist was able to detect these small HCCs, the less experienced radiologist did not find them. Also, it took longer time for both human readers to detect these small HCCs compared to our model.

an AUC of 0.90. Our model seems to be more sensitive than less experienced radiologists in detecting very small HCCs. Furthermore, the classification time of the HCC nodule was 30 milliseconds per image, approximately six times faster than human readers (180 milliseconds). The accuracy of HCC detection was as high as 90%. Based on these results, our deep learning system may be used as an effective decision-support tool for the detection of small HCCs (i.e., sub centimeter HCCs) particularly by less experienced radiologists (Fig. 7).

However, our model exhibited a not infrequent rate of false positive detection. Contrary to our expectation, hepatic cysts which show low signal intensity in the hepatobiliary phase was not a frequent false positive. Instead, intrahepatic vessels, gallbladder, and heart were common false positives in this study. Among the various false positive detections, extrahepatic structures may be explained by our algorithm in which cropping of images was not performed in our model. Instead, whole MR images were used as input data. However, given that extrahepatic structures such as the heart can be easily discriminated from true HCCs by human readers, this problem may not seriously affect our model's performance. Unlike our model, most studies dealing with deep learning regarding radiologic imaging require preprocessing of input data. This means that it is necessary to select images containing lesions, which is performed by human readers, and then regions of interest are cropped^{8,22,23}. Therefore, the cropped images include only the liver mass and surrounding liver parenchyma or the organ and cropped images are entered into the model as input images. This process may be marginally more complicated and time-consuming. In contrast, our method has advantages over other deep learning models as it does not require the process of selecting an image containing target lesions and cropping of images. In our model, the only thing required is to upload entire hepatobiliary phase MR images. Upon image uploading, our model automatically checks for potential candidate HCC nodules in a very short time. In addition, the location of the HCC as detected by our model was confirmed using a class activation map (CAM)²⁵ method. Another advantage of our model is that it works very fast. The average image classification time of 100 images in a single patient was 3.4 seconds,

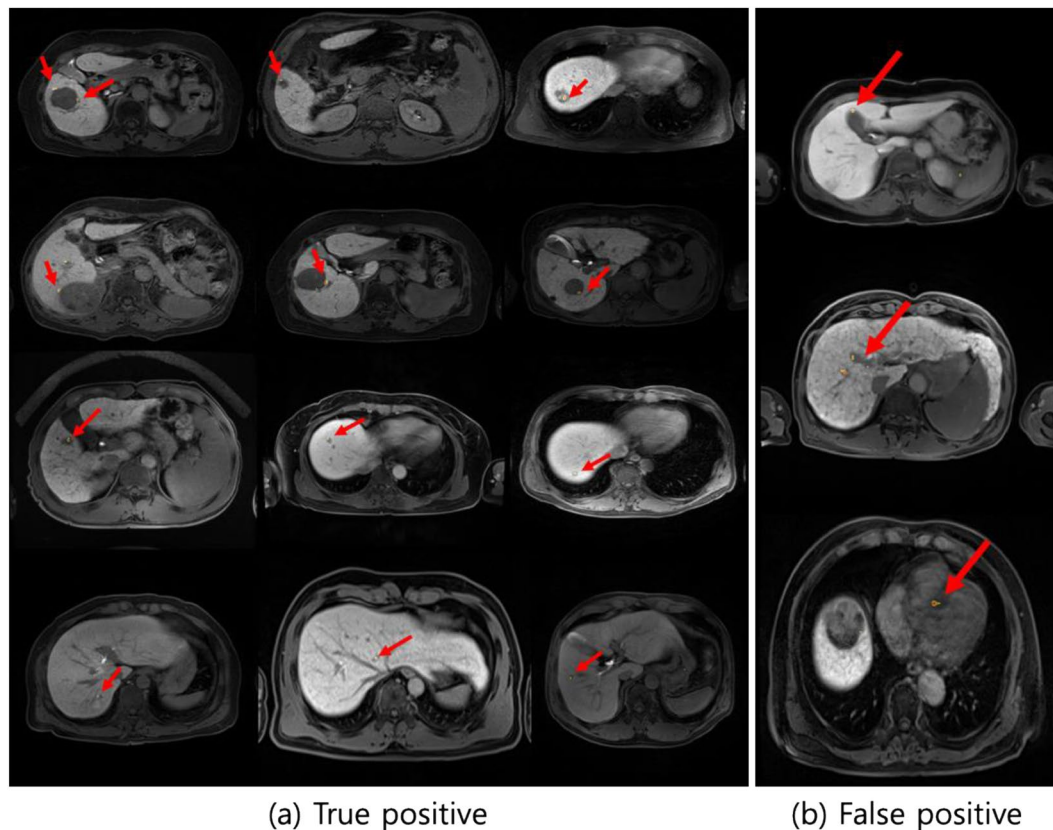


Figure 6. Examples of true positive and false positive detections (arrow) of HCC by our trained model using CAM method. A physician can intuitively discriminate true HCC from pseudo lesions based on HCC candidate indicated by color map.

Class	Sensitivity	Specificity	Accuracy	Interpretation time per 100 images (seconds)
Model	0.87	0.93	0.90	3.4
Human ^a	0.86	0.92	0.91	18
Human ^b	0.98	0.93	0.94	18

Table 4. Comparison between proposed model and human performance in external validation dataset. ^aless experienced radiologist. ^bexpert radiologist.

which is faster than a previous study in which 10 seconds was required for the computation of 100 images²³. However, any direct comparison is limited because the previous study used CT images and the PC performance may be different from ours.

There are several limitations of our study. First, as a preliminary study, only the hepatobiliary phase of gadoxetic acid-enhanced liver MRI was used for the detection of HCCs as the image quality of the arterial phase was frequently affected by transient severe motion artifacts in the arterial phase^{26,27}. However, arterial enhancement on cross-sectional imaging is one of the key imaging findings in the diagnosis of HCC. Moreover, given that deep learning with CNN using multiphase CT images yields higher accuracy than those using single phase CT images²⁶, a deep learning model using multiphase MR images may provide higher accuracy. Further study is warranted using other MRI sequences, including arterial phase images. Nevertheless, our preliminary study found that deep learning can be applied in the detection of small HCCs in the hepatobiliary phase of gadoxetic acid-enhanced liver MRI, which showed similar accuracy to that of less-experienced radiologists with a faster interpretation time. Second, our study population for training, validation, and test sets had relatively good liver function as the patient had undergone surgical resection for HCC. In addition, it may be difficult to detect atypical HCCs with our deep learning model as whole tumors included in this study showed low signal intensity in hepatobiliary phase MRI. That implies that our model may work only for patients with good hepatic function and typical HCCs with low signal intensity on hepatobiliary phase MRI. Third, our training data set used MR images obtained from a single vendor MR scanner (Philips Healthcare, Best, The Netherlands), which may have

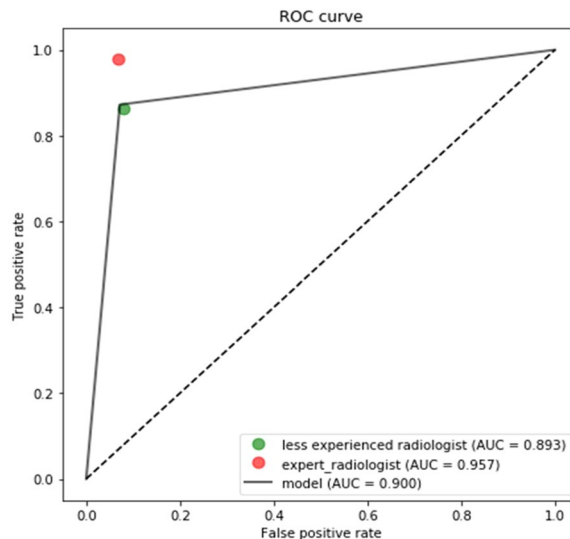


Figure 7. ROC analysis for the proposed model and human readers in external validation dataset.

Manufacturer	Model Name (Tesla)	Number of patients
PHILIPS	Achieva (1.5)	388
PHILIPS	Achieva (3.0)	117
PHILIPS	Ingenia (3.0)	25

Table 5. List of data collection equipment.

resulted in an overfitting issue and thereby slightly lower accuracy in the validation data set in which a variety of MR scanners from multiple vendors were used. We compared our model with various CNNs networks, and our model seemed to perform well. However, in order to get a clearer conclusion, it is necessary to confirm more cases through additional experiments.

Methods

Data collection. The study protocol was in line with ethical guidelines of the 1975 Declaration of Helsinki. This study was approved by the institutional review board (IRB) of Samsung Medical Center (2019-03-101-002), and the IRB waived the requirement to obtain written informed consent from the patients. We reviewed the hepatobiliary phase images of the pre-operative gadoteric acid-enhanced liver MRI of 549 patients from 2010 to 2014 who were confirmed to have HCC after surgical resection. The equipment used for MR acquisition is listed in Table 5.

Data categorization. Among 549 patients (442 male patients and 107 female patients with a mean age of 56 years \pm SD 9.7), 94 patients were excluded due to severe motion artifacts ($n = 31$), missing images ($n = 44$), low image quality ($n = 18$), or absence of preoperative MR images ($n = 1$) (Fig. 8). The remaining 455 patients who all had Child-Pugh score A were included in this study. There was a total of 92,645 hepatobiliary phase MR images. They were categorized into no HCC (41,485 images) and HCC (51,160 images) according to whether HCC was present in the image. Among the 92,645 images, 70%, 15%, and 15% were chosen as the training dataset, validation dataset, and test dataset, respectively.

Data preprocessing and augmentation. Due to the various structures and image sizes included in the MR images, it was difficult to accurately and efficiently learn the characteristics of HCC. Thus, pre-processing that standardizes image size and eliminates unwanted noise was important for improving model learning results and accuracy.

Therefore, all images were scaled to 320×320 pixels using bicubic interpolation and area interpolation since MR images have diverse pixel sizes (from 256×256 to approximately 400×400)^{28,29}. In addition, among the approximately 100 MR images of each patient, only 3–10 images usually contained an HCC nodule. This relative data shortage problem can lead to excessive over-fitting of the model into classes with large amounts of data in learning. Therefore, we augmented the data in various ways to prevent this. First, the HCC area in the chosen image was extracted using a mask. The mask was generated using a human-annotated label map which distinguished the HCC area. To increase the number of data, the HCC images were augmented using rotation, shift, and zooming as shown in Table 6. We tried not to distort the images since image distortion can reduce performance. Therefore, image rotation was only permitted within 90° . Image shift was performed within 10 pixels for all directions. The image was zoomed from 0.8 to 1.2 times. In addition, shift and zoom were combined. Consequently, we had 44,765 HCC images following the image augmentation process (Fig. 9).

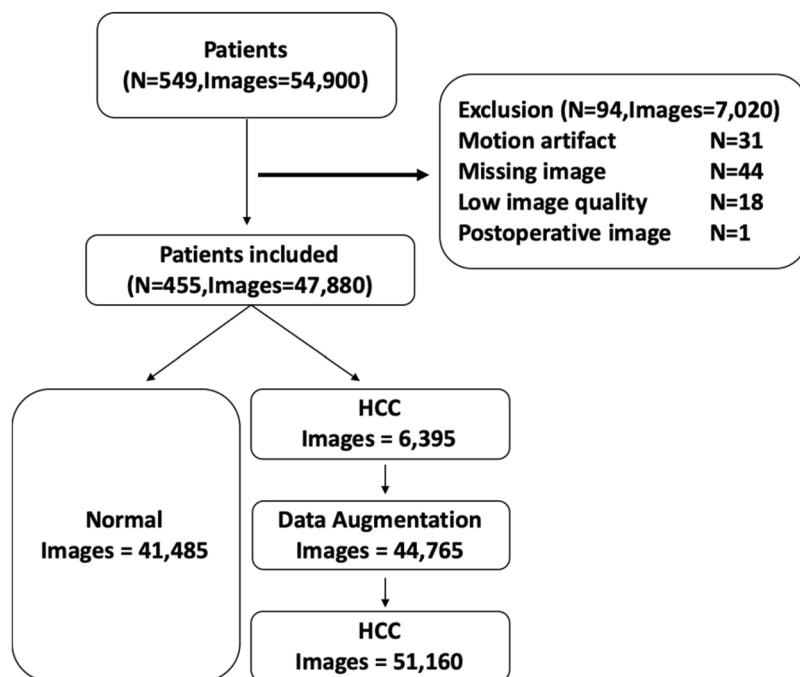


Figure 8. Patient inclusion and exclusion criteria in dataset.

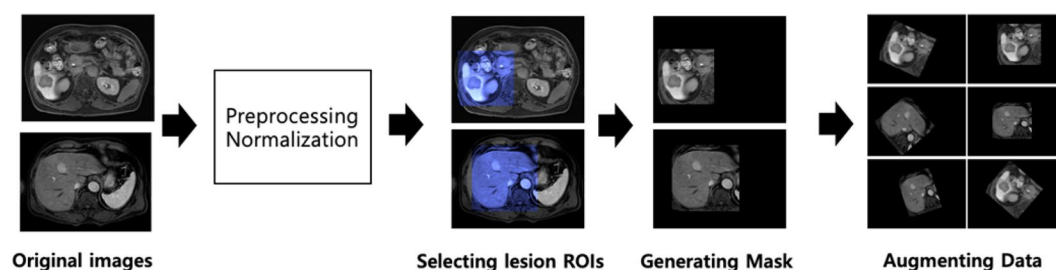


Figure 9. Examples of image augmentation.

Method	Range
Rotate	$-90^\circ \leq \text{rotation angle} \leq 90^\circ$
Shift	1–10 pixels
Zoom	0.8–1.2×
Shift and Zoom	Shift (1–10 pixels) and zoom (0.8–1.2×)

Table 6. Data augmentation methods.

Overall procedure. The overall process of the proposed deep learning system to detect HCC is explained in Fig. 10. The detailed explanation will be described in the following subsections.

Convolutional neural network (CNN) architecture. Since there is no solid theory for hyperparameter optimization, we experimented to identify the best combination of the chosen hyperparameters, including batch normalization, dropout, activation function, kernel size, and optimizer. We randomly selected 11,117 images (4,902 no HCC images and 6,215 HCC images) from the training dataset to optimize the CNN architecture. Then we selected 9,449 images (4,167 no HCC images and 5,282 HCC images) as the learning dataset and 1,668 images (736 no HCC images and 932 HCC images) as the validation data set.

CNN training details. First, all training images were shuffled. Training was terminated when there was no improvement of accuracy in the validation datasets within 20 epochs. The batch size was 128 to balance training quality and convergence speed. The parameters were initialized using the He initializer¹⁷ and the learning rate was 0.001. ReLU was used as the activation function and the Adam Optimizer was applied. Cross entropy was used for

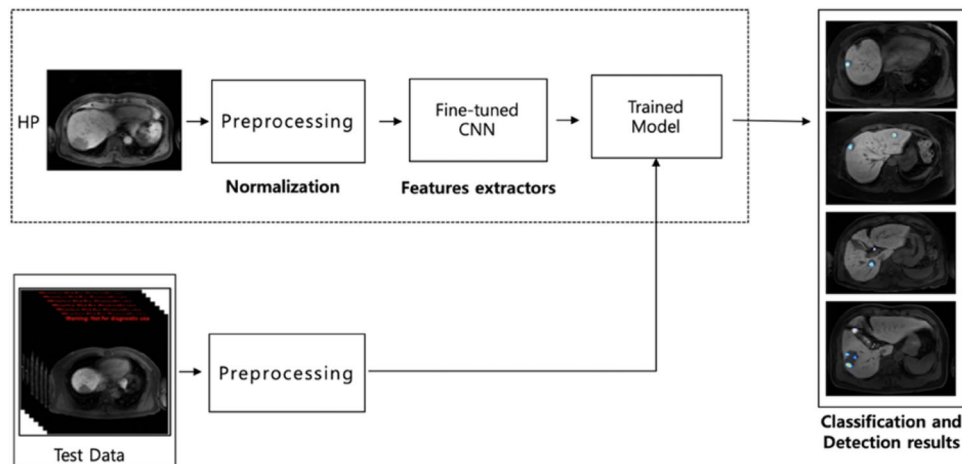


Figure 10. Our proposed Deep Learning system for HCC detection.

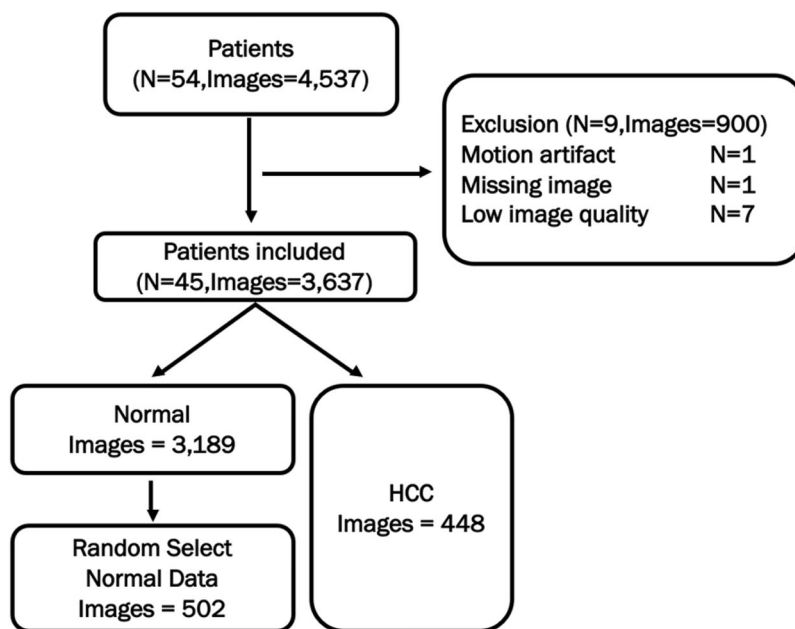


Figure 11. Patient inclusion and exclusion criteria in external validation set.

the loss function. A global average pooling layer was applied to the last layer instead of fully-connected layer, since the fully-connected layer loses location information from the image. By using the global average pooling layer, we were able to reduce the size of the parameters and apply the CAM method to generate the heat map. After this layer, softmax was adopted to predict each class. We used a commodity PC (3.7 GHz \times 12 Intel Core i7, 64 GB RAM, GeForce GTX 1080Ti 8 G \times 2) and TensorFlow V1.8.0.

Performance evaluation of our model. *Data collection.* To verify the performance of our model, we also collected the hepatobiliary phase images from the pre-operation gadoteric acid-enhanced MRI of 54 patients (42 male and 12 female patients with a mean age of 57 years \pm SD 9.6), who had undergone MR imaging at one of four external hospitals from 2015 to 2017. Their histopathologic results were available as they had undergone hepatic surgery in our institution. Among them, nine patients were excluded due to motion artifact ($n = 1$), missing image ($n = 1$), or low image quality ($n = 7$). We randomly selected 502 hepatobiliary phase MR images from 3,189 images with no HCC (Fig. 11). We validated the model using these 502 images and 448 images in which the HCC nodules were included. The equipment used for MR acquisition is listed in Table 7. As in Table 7, the external dataset consisted of MR images obtained with a variety of MR scanners compared to the validation and test datasets.

Manufacturer	Model Name (Tesla)	Number of patients
PHILIPS	Achieva (1.5)	9
PHILIPS	Ingenia (3.0)	8
PHILIPS	Ingenia CX (3.0)	1
SIEMENS	Avanto (1.5)	6
SIEMENS	Skyra (3.0)	18
SIEMENS	SMS Avanto (1.5)	1
SIEMENS	Verio (3.0)	3
GE	DISCOVERY MR750w (3.0)	2
GE	SIGNA EXCITE (1.5)	2
GE	Signa HDxt (3.0)	4

Table 7. List of data collection equipment for external dataset.

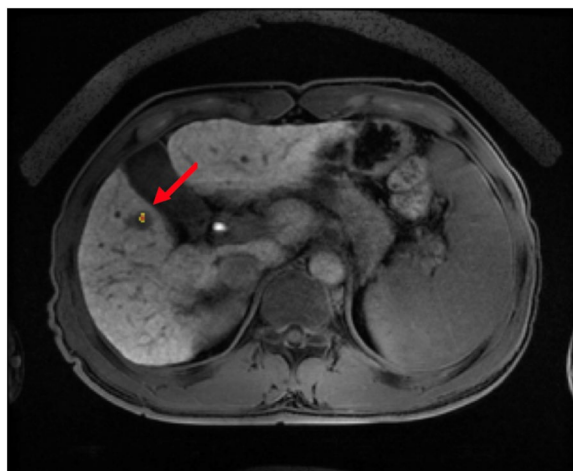


Figure 12. An example of true positive detection (arrow) of HCC by our trained model using CAM method showing the area of interest of the trained model.

Comparison of performance between our model and human readers. To validate the performance of our model, the sensitivity, specificity, and accuracy of HCC detection were compared between our model and radiologists. Two radiologists (a board-certified abdominal radiologist with 10 years of experience with abdominal imaging and a trainee with 4 years of experience in the department of radiology) participated in this validation study. The two radiologists were blinded to the development of the model and the results of the MR reports and histopathologic results of the external validation datasets. They were only informed that the patients might have risk factors for HCC. Therefore, the radiologists were not aware of the presence, number, or location of the HCCs. They were instructed to record the image number containing HCC nodules in the datasheet when reviewing MR images using a picture archiving and communication system (PACS; Centricity Radiology RA 1000; GE Healthcare, Chicago, IL, USA). They were also requested to record the interpretation time using a stopwatch. The interpretation time was defined as the time between image opening and finishing filling out the datasheet.

Image validation. To validate the model, we applied a CAM that points to the correct location and provides clues to the physician. Figure 12 shows where the model automatically predicted the HCC.

Comparison with other CNNs. We compared our own CNN with three other popular deep learning models including ResNet50³⁰, AlexNet³¹, VGG-16³² and Inception-ResNetV2³³. We found that our own CNN architecture outperformed ResNet50, AlexNet, VGG-16 and Inception-ResNetV2. The summary of the result is shown in Table 8 and Supplementary Fig. S1.

Conclusions

We have created a fully automated, deep learning system that detects and classifies HCCs in gadoteric acid-enhanced MRI using a new fine-tuned CNN structure. The system classified HCCs six times faster than human readers and achieved 87% sensitivity and 93% specificity in an external validation data set. This result seems to be comparable to the performance of less experienced radiologists. However, our deep learning model

Models	NO. of trainable parameters	validation accuracy
ResNet50	23,532,418	93%
AlexNet	43,744,890	57%
VGG-16	241,222,338	56%
Inception-ResNetV2	54,278,690	92%
Our own CNN	1,077,186	94%

Table 8. Comparison of CNNs architecture. Our own CNN architecture had the best performance.

has the advantage of detecting very small HCCs better than less experienced radiologists. Finally, for use as a decision support system, we have created a program that categorizes HCCs with a single click and shows the location of candidate HCCs in hepatobiliary phase MR images.

Received: 17 November 2019; Accepted: 9 May 2020;

Published online: 11 June 2020

References

- Parkin, D. M. Global cancer statistics in the year 2000. *The lancet oncology* **2**, 533–543 (2001).
- El-Serag, H. B. & Rudolph, K. L. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* **132**, 2557–2576 (2007).
- European Association for the Study of the Liver. Electronic address, e.e.e. & European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* **69**, 182–236 (2018).
- Marrero, J. A. *et al.* Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology* **68**, 723–750 (2018).
- Lee, Y. J. *et al.* Hepatocellular carcinoma: diagnostic performance of multidetector CT and MR imaging—a systematic review and meta-analysis. *Radiology* **275**, 97–109 (2015).
- Narita, M. *et al.* Expression of OATP1B3 determines uptake of Gd-EOB-DTPA in hepatocellular carcinoma. *Journal of gastroenterology* **44**, 793–798 (2009).
- Kitao, A. *et al.* Hepatocellular carcinoma: signal intensity at gadoxetic acid-enhanced MR imaging—correlation with molecular transporters and histopathologic features. *Radiology* **256**, 817–826 (2010).
- Hamm, C. A., *et al.* Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *European radiology*, 1–10 (2019).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436 (2015).
- Yan, Z. *et al.* Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition. *IEEE transactions on medical imaging* **35**, 1332–1343 (2016).
- Nayak, A., *et al.* Computer-aided diagnosis of cirrhosis and hepatocellular carcinoma using multi-phase abdomen CT. *International journal of computer assisted radiology and surgery*, 1–12 (2019).
- Ioffe, S. & Szegedy, C. J. Batch normalization: Accelerating deep network training by reducing internal covariate shift. (2015).
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. *Dropout: a simple way to prevent neural networks from overfitting*. **15**, 1929–1958 (2014).
- Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. In Proc. icml, Vol. 30 3 (2013).
- He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision 1026–1034 (2015).
- Clevert, D.-A., Unterthiner, T. & Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289 (2015).
- Agarap, A. F. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375 (2018).
- Duchi, J., Hazan, E. & Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* **12**, 2121–2159 (2011).
- Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
- Tieleman, T. & Hinton, G. Divide the gradient by a running average of its recent magnitude. coursea: Neural networks for machine learning. Technical Report. (2017).
- Bottou, L. Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010 177–186 (Springer, 2010).
- Yasaka, K., Akai, H., Kunimatsu, A., Abe, O. & Kiryu, S. Liver Fibrosis: Deep Convolutional Neural Network for Staging by Using Gadoteric Acid-enhanced Hepatobiliary Phase MR Images. *Radiology* **287**, 146–155 (2018).
- 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* **286**, 887–896 (2018).
- Lakhani, P. & Sundaram, B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **284**, 574–582 (2017).
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning Deep Features for Discriminative Localization. in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2921–2929 (2016).
- Motosugi, U., Bannas, P., Bookwalter, C. A., Sano, K. & Reeder, S. B. An Investigation of Transient Severe Motion Related to Gadoteric Acid-enhanced MR Imaging. *Radiology* **279**, 93–102 (2016).
- McClellan, T. R. *et al.* Intravenous Gadoteric Acid Disodium Administration Reduces Breath-holding Capacity in the Hepatic Arterial Phase: A Multi-Center Randomized Placebo-controlled Trial. *Radiology* **282**, 361–368 (2017).
- Keys, R. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing* **29**, 1153–1160 (1981).
- Tanbakuchi, A. Comparison of OpenCV Interpolation Algorithms. (2016).
- He, K. & Zhang, X. Deep Residual Learning for Image Recognition arXiv preprint arXiv:1512.03385 (2015).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 1097–1105, (2012).

32. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition arXiv preprint arXiv:1409.1556 (2014).
33. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, Inception-ResNet and impact of Residual Connections on Learning arXiv preprint arXiv:1602.07261 (2016).

Author contributions

Study concept and design: Shin S.Y. and Lee M.W. Collection of clinical data: Lee M.W. Statistical analysis and interpretation: Kim J.M., Min J.H., Kim S.K. Drafting of the manuscript: Kim J.M., Lee M.W. Critical revision of the manuscript for important intellectual content: Kim J.M., Lee M.W., and Shin S.Y. Final approval of manuscript: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-65875-4>.

Correspondence and requests for materials should be addressed to J.K. or M.W.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020