

Computer aided diagnosis system developed for ultrasound diagnosis of liver lesions using deep learning

Makoto Yamakawa
Graduate School of Medicine
Kyoto University
Kyoto, Japan
yamakawa.makoto.6x@kyoto-
u.ac.jp

Tsuyoshi Shiina
Graduate School of Medicine
Kyoto University
Kyoto, Japan
shiina.tsuyoshi.6w@kyoto-
u.ac.jp

Naoshi Nishida
Faculty of Medicine
Kindai University
Osaka, Japan
naoshi@med.kindai.ac.jp

Masatoshi Kudo
Faculty of Medicine,
Kindai University
Osaka, Japan
m-kudo@med.kindai.ac.jp

Abstract—The Japan Society of Ultrasonics in Medicine (JSUM) is currently constructing an ultrasound image database. This database collects B-mode images of liver tumors and breast tumors, and B-mode videos of heart disease. In the past year, 31,000 liver tumor images have been collected from 11 institutions and 14,000 breast tumor images have been collected from 5 institutions. We are developing computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems for liver and breast tumors based on deep learning using this database. In this paper, we report on CADx to estimate liver tumor types as a first trial. The data used in this study are 159 cyst cases (338 images), 68 hemangioma cases (279 images), 73 hepatocellular carcinoma (HCC) cases (241 images), and 24 metastatic liver cancer cases (122 images), collected at one facility. We developed the CADx system that estimates four types of liver tumor using a convolutional neural network based on VGGNet. The accuracy of the developed 4-class classification CADx was 88.0%. The accuracy by tumor type was 98.1% for cysts, 86.8% for hemangiomas, 86.3% for HCC, and 29.2% for metastatic liver cancer, with increasing accuracy observed for larger data sets. We also developed CADx to estimate whether a liver tumor is benign or malignant. The accuracy of this 2-class classification CADx was 94.8%, the sensitivity was 93.8%, and the specificity was 95.2%. Both 4-class classification and 2-class classification CADx had relatively high accuracy. However, in this study, we used only a small amount data collected from a single facility. In the future, we plan to verify our results using a larger amount of data collected from multiple facilities. In addition, we prototyped CAD software and are currently developing it with feedback from doctors.

Keywords — artificial intelligence, deep learning, convolutional neural network, computer-aided diagnosis, liver tumor, ultrasound image database

I. INTRODUCTION

In recent years, development of computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems for medical images using deep learning has been widely investigated. Using deep learning, high diagnostic accuracy has been reported for CADx for diabetic retinopathy, CADx for skin cancer, CADe and CADx for endoscopic images [1-4]. On the other hand, there are still few examples of applying deep

learning to ultrasound images, such as diagnosis and detection of breast cancer and detection of the fetal heart [5-7]. This is probably because there is no large-scale database of ultrasonic data. Therefore, the Japan Society of Ultrasonics in Medicine (JSUM) has been constructing a large-scale database of ultrasound images with the support of the Japan Agency for Medical Research and Development (AMED) since 2018. Currently collected data are B-mode images and screening videos of liver and breast tumors, and B-mode videos of heart disease. This data is collected from 11 centers for liver tumors, 5 centers for breast tumors, and 5 centers for heart disease. In the past year, 31,000 liver tumor images and 14,000 breast tumor images have been collected. This database currently stores the tumor position in the image and tumor size in pixel units together with the B-mode image. Screening videos of liver and breast tumors and B-mode videos of heart disease will be collected later. The JSUM project also aims to develop CADe and CADx systems using deep learning as well as database construction. In this paper, we report on the development of CADx that estimates the liver tumor type from the ultrasound live tumor image, which was the first trial in this project.

II. METHOD

A. Data and preprocessing

The data analyzed here are ultrasound B-mode images of 324 liver tumors (980 images) collected at Kinki University with the approval of the Ethics Committee in 2018. This data includes 159 cases of cysts (338 images), 68 cases of hemangioma (279 images), 73 cases of hepatocellular carcinoma (HCC) (241 images), and 24 cases of metastatic liver cancer (122 image).

During preprocessing, we first cropped the region of interest (ROI) so that the maximum diameter of the tumor is 60% of the ROI size (Fig. 1). This ratio was determined based on the preliminary verification results [8]. However, because some tumors were very small, we set a lower limit (64×64 pixels) on the ROI size. In addition, for data augmentation, we added left-right inverted images, $\pm 5^\circ$, $\pm 10^\circ$ rotated images and their left-right inverted images to the training data. The B-mode images of the liver tumor used were all measured with a convex probe, and the maximum deflection angle of the ultrasonic beam was

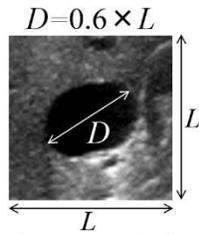


Fig. 1. Relationship between ROI image and maximum tumor diameter.

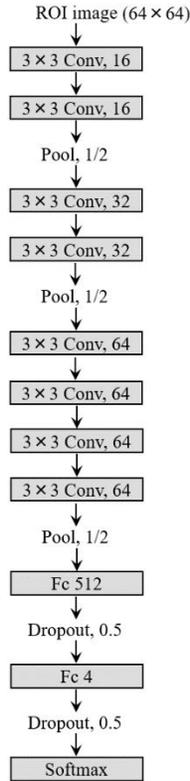


Fig. 2. CNN for liver tumor 4-class classification.

$\pm 30^\circ$. For this reason, we added rotated images within $\pm 10^\circ$ to the training data. In other words, we increased the training data to 10 times the original data. Finally, we resized the ROI image to the input image size (64×64 pixels) of our convolutional neural network (CNN).

B. Convolutional neural network

We used a CNN as shown in Fig. 2 based on VGGNet [9]. In liver B-mode images, many liver tumors are depicted as small in the image. Therefore, we set the input image size to 64×64 pixels. Because the size of input image is relatively small, many max-pooling processing cannot be performed. Therefore, we set the depth of CNN to 10 layers. In addition, we added a dropout process to prevent over-learning. Finally, we calculated the probability of 4 classes (cyst, hemangioma, HCC, metastatic liver cancer) using the softmax function, and we determined that the tumor type was the one with the highest probability. We used ReLU function as the activation function, Adam (adaptive moment estimation) as the optimizer, 0.001 as the learning rate, half of the training data as the batch size, and 20-30 as the

Table I Liver tumor 4-class classification result.

		True condition			
		Cyst	Hemangioma	HCC	Metastatic liver cancer
Predicted condition	Cyst	156	1	0	0
	Hemangioma	2	59	8	11
	HCC	1	7	63	6
	Metastatic liver cancer	0	1	2	7

number of epochs.

C. Validation

Owing to the small amount of data, we randomly divided all the data into 10 groups and verified them using the cross-validation method. When dividing into groups, we used cases as one unit such that images of the same case were not mixed with test data and training data. Additionally, we made the same percentage of tumor types in each group and evaluated the accuracy of cases as a unit. This is because the images of the same case are relatively similar, and the number of images significantly varies depending on the case. In the validation data, we used the accuracy in case units as the primary evaluation criterion. In the evaluation in case units, we calculated the probability of each type of tumor for all images included in the same case and calculated the average value of the probabilities. The tumor type with the highest probability was used as the estimated result for that case. We used only the original ROI image as test data and did not use the left-right reversed image or the rotated image.

III. RESULTS

A. 4-class classification result

Table I lists the results of classifying the four types of liver tumors by the CNN in Fig. 2. The overall accuracy was 88.0% (285/324 cases). Accuracy by liver tumor type is 98.1% for cysts (156/159 cases), 86.8% for hemangiomas (59/68 cases), 86.3% for HCC (63/73 cases), and 29.2% for metastatic liver cancer (7/24 cases). From this result, it was confirmed that the greater the amount of data, the higher the accuracy. Because cysts are relatively easier for humans to diagnose, it can be estimated with high accuracy in CAD. On the other hand, it was found that the accuracy of metastatic liver cancer is low, and it is often mistakenly estimated for hemangioma and HCC. One reason for this is the lack of data on metastatic liver cancer.

B. Benign / malignant classification result from 4-class classification result

Based on the above 4-class classification results, the classification accuracy of benign tumors (cysts, hemangiomas) and malignant tumors (HCC, metastatic liver cancer) is

Table II Benign / malignant classification result from 4-class classification result.

		True condition	
		Benign	Malignant
Predicted condition	Benign	218	19
	Malignant	9	78

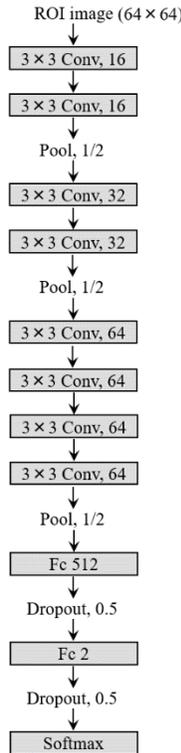


Fig. 3. CNN for liver tumor 2-class classification.

calculated as listed in Table II. From this result, the accuracy of benign and malignant classification using the CNN in Fig. 2 is 91.4% (296/324 cases) for accuracy, 80.4% (78/97 cases) for sensitivity, and 96.0% (218/227 cases) for specificity. Classification accuracy of benign tumors including cysts which were relatively easier to determine were high, but classification accuracy of malignant tumors including metastatic liver cancer which had low accuracy owing to the lack of data was slightly low.

C. 2-class classification result

Next, we created a 2-class classification CNN that performs benign and malignant classification as shown in Fig. 3, and trained this CNN. The result of 2-class classification CNN is listed in Table III. From this result, the accuracy was 94.8% (307/324 cases), the sensitivity was 93.8% (91/97 cases), and the specificity was 95.2% (216/227 cases). Because the 2-class classification CNN has high accuracy, it is considered that the accuracy of the 4-class classification is improved by using the 2-class classification CNN.

Table III Liver tumor 2-class classification result.

		True condition	
		Benign	Malignant
Predicted condition	Benign	216	6
	Malignant	11	91

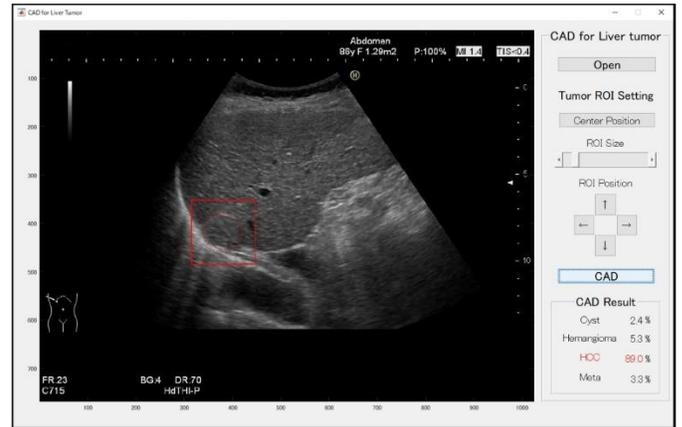


Fig. 4. CADx software developed for ultrasound diagnosis of liver tumor.

D. CADx software

We developed CADx software by adding a GUI to the trained CNN. The screenshot of the developed CADx software is shown in Fig. 4. The basic operation is to first read an image file (e.g., DICOM and JPEG), then set the ROI, and finally classify the liver tumor type. When setting the ROI, set the tumor so that it touches the red circle as shown in Fig. 4. In addition, the result is to display the probability of each tumor type. This software is currently being improved with doctor's feedback.

IV. CONCLUSION

In this paper, we tried to develop a CADx system that estimates the type of liver tumor using part of the ultrasound image database collected by the JSUM. Although the amount of data used was a little less than 1,000, relatively good accuracy was obtained. However, because the number of metastatic liver cancer cases was extremely small, the accuracy of this particular classification was low. Therefore, we are currently verifying our approach using a data set more than 10 times used in this study. We are also currently conducting verifications using multi-facility data. In addition, we developed CADx software and provided this software to doctors. We are improving the CADx software with doctor's feedback, and we plan to update the CNN by increasing the data input and training it again.

ACKNOWLEDGMENT

This work was supported by AMED under grant number JP191k1010035.

REFERENCES

- [1] D.S.W. Ting, C.Y. Cheung, G. Lim, G.S.W. Tan, N.D. Quang, A. Gan, et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes", *JAMA*, vol.318, no.22, pp.2211-2223, 2017.
- [2] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks", *Nature*, vol.542, pp.115-118, 2017.
- [3] M. Misawa S. Kudo, Y. Mori, T. Cho, S. Kataoka, A. Yamauchi, et al., "Artificial intelligence-assisted polyp detection for colonoscopy: initial experience", *Gastroenterology*, vol.154, pp.2027-2029, 2018.
- [4] T. Hirasawa, K. Aoyama, T. Tanimoto, S. Ishihara, S. Shichijo, T. Ozawa, et al., "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images", *Gastric Cancer*, vol.21, pp.653-660, 2018.
- [5] J. Z. Cheng, D. Ni, Y. H. Chou, J. Qin, C. M. Tiu, Y. C. Chang, et al., "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US image and pulmonary nodules in CT scans", *Scientific reports*, 6:24454, pp.1-13, 2016
- [6] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggielaar, et al., "Automated breast ultrasound lesions detection using convolutional neural networks", *IEEE J. Biomed. Health Infomatics*, vol.22, no.4, pp.1218-1226, 2018
- [7] V. Sundaresan, C. P. Bridge, C. Ioannou, J. A. Noble, "Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks", in *Proc. IEEE Int. Symp. Biomed. Img.*, pp.671-674, 2017
- [8] M. Yamakawa, T. Shiina, N. Nishida, M. Kudo, "Basic study on optimal cropping setting in convolution neural network for ultrasonic liver tumor diagnosis", in *Proc. Symp. Ultrasonic electronics 2019*, in press
- [9] K. Shimonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", in *Proc. ICLR 2015*, arXiv 1409.1556, 2015.