Journal of Associated Medical Sciences 2023; 56 (3): 105-112

JAMS
Journal of Associated Medical Sciences

Scopus Indexed & Thai-Journal Citation Index Centre (TCI)

Journal of Associated Medical Sciences

Journal homepage: https://www.tci-thaijo.org/index.php/bulletinAMS/index

Two-stage method for hepatocellular carcinoma screening in B-mode ultrasound images

Sutthirak Tangruangkiat[1] Anchali Krisanachinda[1] Wongsakorn Ruangprapawut[2] Supatana Auethavekiat[3*]

[1]Department of Radiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.
[2]R&D engineer, Intronics company limited, Bangkok, Thailand.
[3]Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand.

ARTICLE INFO

Article history:
Received 15 June 2023
Accepted as revised 23 June 2023
Available online 17 July 2023

Keywords:
HCC screening, two-stage model, YOLOv4 detector, classifier, liver ultrasound

ABSTRACT

Background: Hepatocellular carcinoma (HCC) is a significant global health concern that requires early detection for effective treatment.

Objectives: The objective of this study was to develop a system for screening HCC in B-mode ultrasound images.

Materials and methods: The dataset consisted of 1665 hemangioma (HEM) images, including 961 typical HEM, 704 atypical HEM, and 543 HCC images. Four YOLOv4 models were trained: one for HCC detection, one for the conventional two-class detection of HEM and HCC, one to detect typical HEM and suspicious lesions, and the last one was our two-stage model consisting of a detector and classifier. In the first stage, a YOLOv4-based detector with ResNet-50 as the backbone was used to identify focal liver lesions. The second stage utilized ResNet-50 as a classifier to classify the lesions into HCC, atypical HEM, or typical HEM. Differentiating between HCC and atypical HEM is not necessary, as both require further investigation with CT or MR imaging.

Results: The evaluation of the developed HCC screening system using ten-fold cross-validation showed that grouping HCC and atypical HEM together significantly increased precision from 0.74 to 0.88 and improved HCC recall from 0.64 to 0.68. Furthermore, employing the two-stage method further improved HCC recall from 0.68 to 0.72.

Conclusion: The results indicate that combining HCC and atypical HEM into a single class and using a two-stage approach for detection led to substantial improvements in precision and HCC recall. These findings highlight the potential of the developed system for effective HCC screening in B-mode ultrasound images. The two-stage method provided better detection than the detector-only method. More accurate detection was achieved when lesions were classified based on appearance and clinical protocols.

Introduction

Ultrasound imaging is the common liver screening protocol and often the first tool to detect the early stage of hepatocellular carcinoma (HCC), the most common liver cancer. However, it must be followed by other imaging modals (contrast-enhanced ultrasound: CEUS, computed tomography: CT, or magnetic resonance imaging: MR) for definite diagnosis due to the shared sonographic appearance of HCC and hemangioma (HEM) (Figure 1).[1,2] Deep learning models have been applied for liver lesion classification as well as detection.[3-11] It is hypothesized that the model is capable of capturing the difference invisible

∗ **Corresponding contributor.**
Author's Address: *Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand.*
E-mail address: supatana.a@chula.ac.th
doi: 10.12982/JAMS.2023.060
E-ISSN: 2539-6056

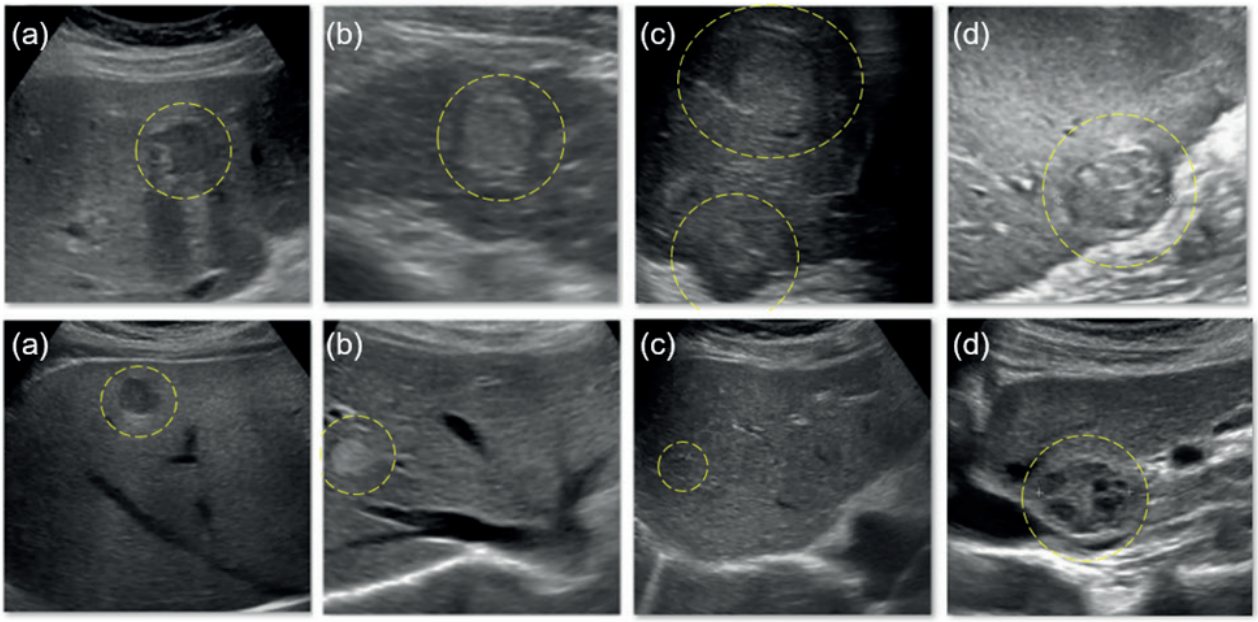

Figure 1 *Echogenic patterns of HCC and HEM lesions. Top row: HCC lesions. Bottom row: HEM lesions. a: hypoechoic, b: hyperechoic, c: isoechoic, d: mixed echoic appearances.*

to humans, if it has been trained by a sufficiently large dataset. The limited dataset is among the major problems of deep learning models. From our survey, the RetinaNet in Tiyarattanachai *et al* was trained by the largest dataset (20432 lesions which included 2414 HCC).[12] Nevertheless, the recall of HCC was the lowest among other liver lesions and had the highest deviation. The high deviation indicated that the number of data was too small.

In this paper, we hypothesized that the available dataset was not sufficient for differentiating HCC from HEM. Thus, we followed the clinical protocol and categorized HEM into two groups, typical and atypical. The typical HEM is uniform hyper-echogenicity and a well-defined margin.[13] It will be monitored for change during the follow-up. The atypical hemangioma exhibits various imaging features and shares similarities with HCC.[13] Both are grouped into suspicious lesions and will be sent for further investigation by CT or MR imaging.[2,14,15] We tackled the problem of a limited dataset by applying the two-stage method which had the detection of HCC and HEM lesions followed by the classification to typical HEM, atypical HEM, and HCC.

Materials and methods

Dataset

The abdominal ultrasound images used in this retrospective study encompass both upper abdominal ultrasound images and whole abdominal ultrasound images. The study received approval from Chulabhorn Research Institute, Thailand (CRI No. 098/2563), as well as the Institution Review Board of the Faculty of Medicine, Chulalongkorn University, Thailand (IRB No. 485/2563). The images utilized were obtained from the period 2015 to 2019 at these two institutions. The inclusion criteria for selecting the images are outlined below.

- 2D ultrasound images from a curvilinear transducer. Due to the diverse range of ultrasound machine brands employed across the two hospitals, the dimensions and resolution of the images varied.
- HCC and HEM were confirmed by CT or MR reports.

The dataset consists of 961 typical HEM, 704 atypical HEM, and 543 HCC images. A skilled sonographer drew the lesion boundary. Compared to previous studies, Our dataset is larger than previous studies, however, it is much smaller than Tiyarattanachai *et al*.[12,16,17]

For data preparation, all images were converted to grayscale and cropped to focus on the liver by removing extraneous black areas. They were resized to 224x224 pixels to fit the input requirement of the ResNet-50 model. Patient information was removed. Some images contained markers, but their presence did not significantly impact the detector due to the mixed presentations (Figure 2). Therefore, the markers were not removed in this study.

Methods

The proposed two-stage method is depicted in Figure 3. The detector in the first stage was trained to detect focal liver lesions. Both HCC and HEM are focal liver lesions, so the training data becomes a combination of HCC and HEM images. The size of the training dataset was larger than the model where HEM and HCC were separately considered. Furthermore, the shared sonographic appearance of HEM and HCC can be exploited for better detection.

The result of the first stage was resized to 224x224 images and inputted to the classifier in the second stage. The classifier categorized lesions into three classes: typical HEM, atypical HEM, and HCC. Differentiating HCC from atypical HEM was not crucial as both required further investigations.

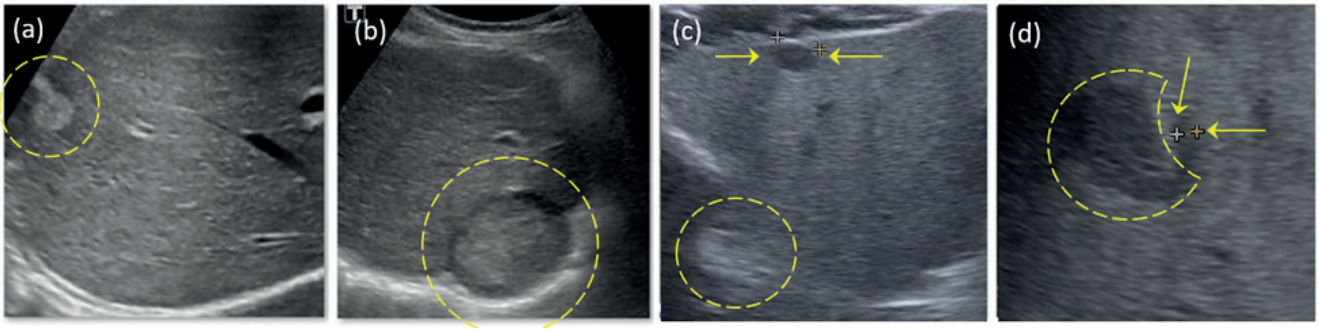

Figure 2 *Ultrasound images with and without markers. HEMs and HCCs show inside a dashed circle. a: HEM without a marker, b: HCC without a marker, c: a marker for hepatic cyst in HEM image, d: a marker for vessels near HCC.*

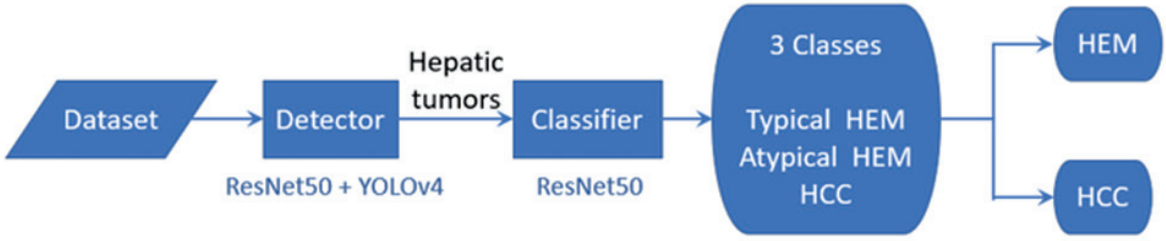

Figure 3 *Proposed two-stage method.*

The models for the detector and the classifier were selected from the available models in the Deep Learning Toolbox of MATLAB 2022a (license number 40662904). In the preliminary experiment, we compared the performance of four detectors: regions with convolutional neural networks (R-CNN), single shot detector (SSD), You Only Look Once (YOLO) v2, and YOLOv4. SSD failed to provide accurate detection. R-CNN and YOLO had comparable accuracy, but R-CNN required much longer training times. YOLOv2 and YOLOv4 had comparable accuracy but YOLOv4 offered a more precise lesion location. Thus, YOLOv4 was chosen as the detector. The architecture of YOLO consists of a backbone, neck, and head. The backbone acts as the feature extractor, while the neck is used to connect the features to the head, which provides the detection output. Pretrained convolutional networks are used as the backbone. In our preliminary experiment, ResNet-50 provided a performance better than CSPDarkNet53. Therefore, we used YOLOv4 with ResNet-50 as the backbone of this study. For the classifier, GoogLeNet, VGG-16, ResNet-18, and ResNet-50 were tested. ResNet-50 offered the highest accuracy, so it was selected as the classifier. All CNN networks were pre-trained using the ImageNet database.

Setting

All models were implemented in MATLAB 2022a on a personal computer (CPU: Intel Xeon, RAM: 128 GB, Video Card: NVIDIA 16 GB). Ten-fold cross-validation was used for performance comparison. We compared the proposed two-stage method with the following three detectors.

1. Model 1: HCC detector trained by HCC images only.
2. Model 2: HCC and HEM detector where HCC and HEM were considered as separate classes.
3. Model 3: typical HEM and suspicious lesion detector. HEM was divided into a typical HEM and an atypical HEM. Atypical HEM and HCC were grouped into suspicious lesion classes.

In most previous works detectors were trained to find HCC as a distinct lesion from HEM.[3,7,12,16,17] So, the first two models were used as the baseline models. The third model follows the clinical protocol and divided the lesions into typical HEMs for future monitoring and suspicious lesions for additional investigation. These three models were compared with Model 4, which is the initial stage of our two-stage method. Model 4 was trained to detect focal liver lesions which combine HCC and HEM in the same class.

To address the variability in the direction of the ultrasound beam, which can depend on the user (radiologist/sonographer), image data augmentation techniques were employed. Specifically, a rotation of ±5 degrees and vertical/horizontal flipping were applied. The dataset for the classifier consisted of manually drawn lesion areas extracted from the image dataset used to train the detector. These lesion areas were resized to 224x224 pixels.

The experiment was divided into three parts to assess the performance of the proposed two-stage method. The first part focused on investigating the detector's accuracy and error. The second part examined the classification accuracy of the ResNet-50 model. Finally, the overall accuracy of the two-stage method was evaluated against the detector-only method.

Performance evaluation

Intersection over Union (IoU) is the ratio of the

area of overlap and the area of union. It is often used to evaluate the result of a detector. In this experiment, the result of the detector was considered correct if the IoU was at least 50%. The classification was then evaluated by the following metrics.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{(i)}$$

$$\text{recall} = \frac{TP}{TP + FN} \quad \text{(ii)}$$

$$\text{precision} = \frac{TP}{TP + FP} \quad \text{(iii)}$$

$$\text{negative predictive value} = \frac{TN}{TN + FN} \quad \text{(iv)}$$

$$\text{F1 - score} = \frac{2 \times \text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad \text{(v)}$$

where *TP, TN, FP* and *FN* are the number of true positive, true negative, false positive, and false negative, respectively.

In these evaluation metrics, a value close to 1 indicates good performance, while lower values indicate poorer performance. In addition, the average precision was also used. The average precision is the precision averaged over all the detection results. The higher the average precision indicates the better detector. Since ten-fold cross-validation was used, all metrics were averaged from the 10 experiments. Note that recall will be mostly focused since it is the most important metric for screening tools.

Result and discussion

Performance evaluation: detector

The detection result is presented in Table 1. The target lesion was considered positive. All models were tested with both HCC and HEM images. The results indicated that the detector trained to specifically detect HCC (Model 1), achieved a higher recall rate compared to the two-class model used in Model 2, but it came at the cost of the inability to differentiate HEM from HCC (low precision). The recall of Model 2 varied from 0.53 to 0.97 which indicated low repeatability. Model 3 had a high recall rate for detecting suspicious lesions, but when only HCC was considered, the recall rate dropped to 0.68.

The finding is consistent with other studies on the detection of malignant tumors, such as Cao *et al.* who used SSD to detect breast tumors in ultrasound images, and Tanaka *et al*. who developed a computer-aided diagnosis (CAD) system for classifying breast cancer but achieved a detection rate of less than 50% of breast tumors in ultrasound images.[18,19] A recent study in 2021 by Tiyarattanachai *et al*.[12] reported a high recall of 0.74 for HCC detection using RetinaNet, but this was achieved by lowering the IoU threshold to 0.2.

The best detection result was achieved by Model 4. The combination of HEM and HCC in the focal liver lesion group provided a larger dataset that could be used to train the detector to identify the distinct characteristics of both types. Notably, the detector successfully detected HCC lesions missed by the first three models, as shown in Figure 4.

To ensure that the higher recall of Model 4 led to better HCC detection. The detection result was categorized into 3 classes: HCC, HEM, and others (incorrect detection) and shown in Table 2. It is worth noting that certain images contained multiple HEM/HCC lesions, and YOLOv4 did not detect all of them. Some lesions were detected multiple times, as shown in the last row of Table 2 and Figure 5. The result indicated that Model 4 outperformed the other three models, with recall rates of 0.78 for HCC and 0.86 for HEM.

Model 4 exhibited two types of detection errors. The first type involved the failure to detect focal liver lesions, impacting the recall of the two-stage method. The second type was the misdetection of other areas/lesions as focal liver lesions, affecting precision. The second type of error constituted less than 5% of the total test data and could be easily dismissed by radiologists during follow-up.

Among 135 undetected HCC, 118 lesions (87%) did not have the sonographic appearance of HCC. Most of these lesions displayed features such as faint opacity, isoechoic tumor, or incomplete border. These lesions were

Table 1 Detection results of four YOLOv4 models. The values in the parenthesis were the range of the matrices.

	Model 1	Model 2	Model 3	Model 4
	1 class 543 HCC	2 classes 1) 1665 HEM 2) 543 HCC	2 classes 1) 961 typical HEM 2) 1247 suspicions	1 class 2208 HEM and HCC like lesions
Accuracy	0.52 (0.39-0.61)	0.85 (0.73-0.89)	0.72 (0.70–0.77)	0.86 (0.82-0.88)#
Precision	0.54 (0.43-0.61)	0.74 (0.50-0.88)	0.88 (0.84-0.93)#	0.88 (0.82-0.91)
Recall	HCC 0.67 (0.54-0.86)	HCC 0.64 (0.53-0.97)	Suspicion 0.70 (0.63-0.75)	HEM and HCC 0.84 (0.79-0.89)#
F1-score	0.71 (0.64-0.87)	0.68 (0.51-0.93)	0.78 (0.75-0.82)	0.86 (0.83-0.89)#
NPV*	0.60 (0.48-0.70)	0.87 (0.82-0.92)#	0.72 (0.65-0.76)	0.86 (0.81-0.88)
mAP**	0.50 (0.32-0.68)	0.49 (0.37-0.61)	0.60 (0.51-0.65)	0.76 (0.73-0.84)#

Note: *#the best result for the given evaluation matrix, *NPV: negative predictive value, **mAP: mean average precision*

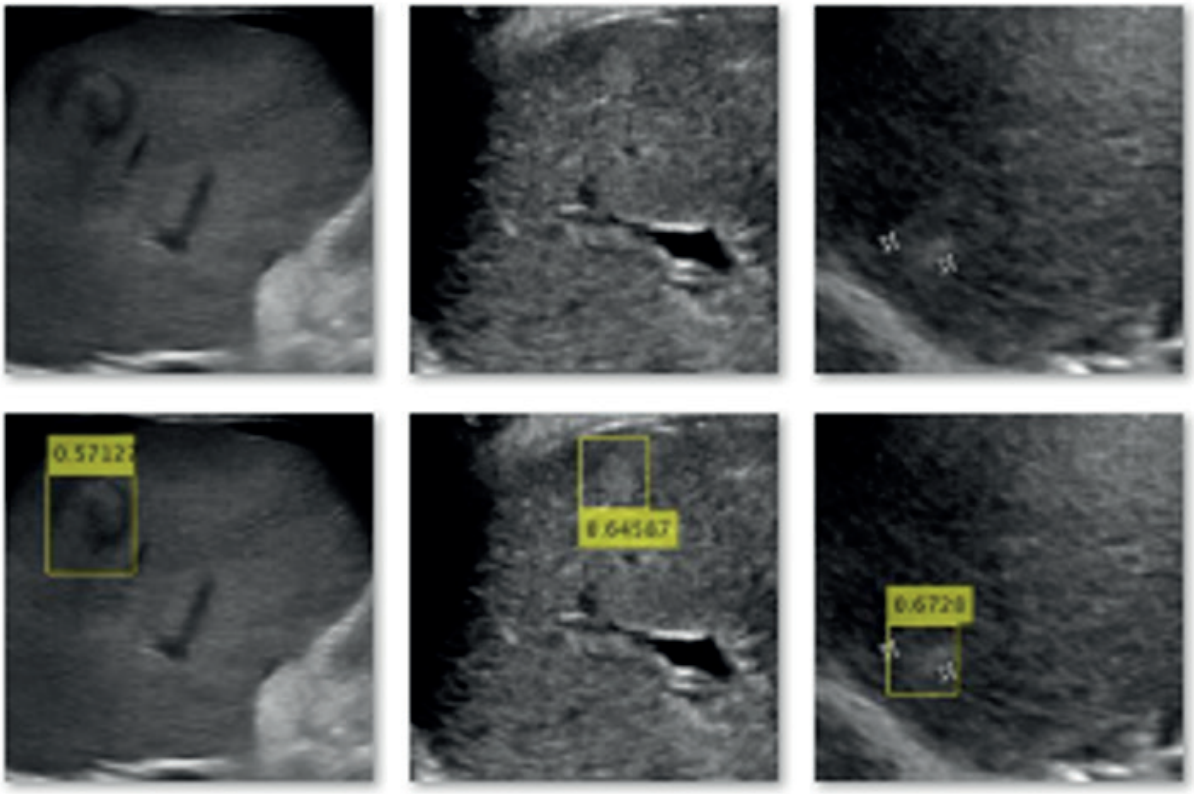

Figure 4 *Training YOLOv4 to detect HEM and HCC as one class improved HCC detection (bottom row) compared to training them as separate classes (top row) which failed to detect the lesions.*

Table 2 Detection results of Model 4 as grouped by lesion type.

Detector model YOLOv4	The number of detected lesions (actual value)			
	HCC	HEM	Others	Total*
Images	472 (543) 86.92%	1455 (1665) 87.39%	68 (0)	1927 (2208) 87.73%
Lesions	480 (615) 78.05%	1479 (1721) 85.94%	68 (0)	1954 (2336) 83.64%
Lesion + Redundancy	489 (624) 78.37%	1494 (1734) 86.16%	68 (0)	1983 (2358) 84.10%

**Total is the sum of the HCC and HEM only. Other lesions were not considered.*

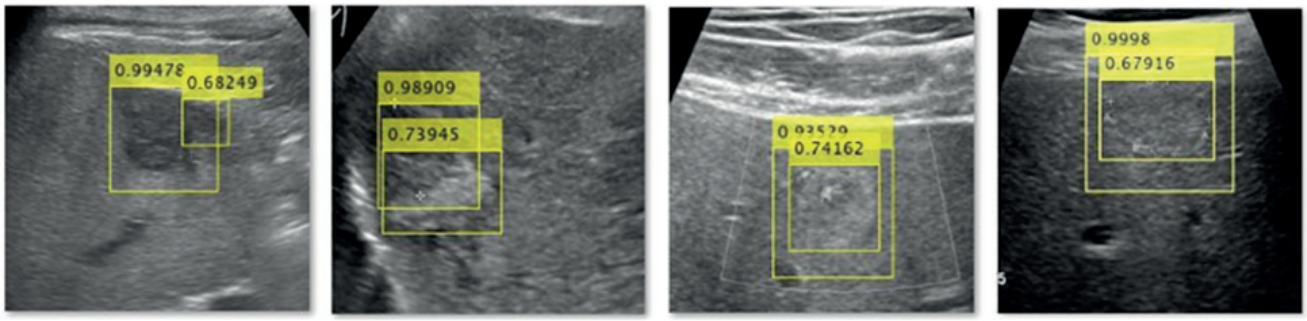

Figure 5 *Examples of multiple detection of the same lesion by YOLOv4.*

detected in further CT or MR scanning. In clinical protocol, if a new lesion appears where nothing was shown in the previous scanning, irrespective of the appearance, CT or MR scan is requested. Without the previous records, it is impossible to detect these HCC. Furthermore, if one HCC is detected, the entire liver will be scanned by CT or MR imaging. Thus, the detections of every HCC or at least one in an image have the same outcome. In this sense, it is possible to conclude that YOLOv4 could detect 86.92% of HCC patients. However, it is not guaranteed that at least one lesion would be detected, so all undetected HCC was considered false negative in this study. We concluded that the focal liver lesion detector had a 0.78 recall rate for HCC.

Performance evaluation: classifier

ResNet-50 was applied to classify focal liver lesions into three classes: typical HEM, atypical HEM, and HCC. Table 3 presents the confusion metric of the classification, where only the correct results of the first stage were considered. The 472 HCC images detected by the first-stage detector had 480 lesions (from Table 2), and the 1455 HEM images had 1479 lesions (570 atypical and 909 typical HEMs). Note that the number of detected HCC was 489 due to the multiple detection of some HCC lesions.

With the limited dataset, it is impossible to prove whether the deep learning model can differentiate the difference between atypical HEM and HCC in a B-mode ultrasound image. However, the further treatment plan for both lesions is the same, i.e., scheduled for CT or MRI examination. Therefore, the detection of HCC as atypical HEM did not pose a health risk. Table 3 was modified to Table 4 where the HCC incorrectly detected as atypical HEM is accepted as the correct classification. According to Table 4, the HCC recall rate of 0.92 (448/489) was achieved. The accuracy and the negative predictive value (HEM = negative) were 0.90 and 0.97, respectively. When HCC was considered positive, the precision (0.74) was much lower than the other values. This is because ultrasound imaging is not a tool to differentiate HCC from atypical HEM. Among 157 errors, 59 images were atypical HEMs. If atypical HEM was considered the same class as HCC (instead of HEM), the precision would jump to 0.82.

A more serious problem was an HCC incorrectly classified as a typical HEM. Out of the 41 HCC incorrectly classified as typical HEM, 22 lesions closely resembled typical HEM. These 22 lesions were well-defined and hyperechoic (Figure 6). Some of these lesions were detected in further CT or MR examinations because they were either presented 1) in a liver with multiple HCCs or 2) new lesions that appeared in the area without any lesions during the previous ultrasound screenings. Furthermore, some misclassification occurred, because the detector did not extract enough area of the HCC lesion as shown in the leftmost image of Figure 5.

The classifier was trained by the ground truth lesions. The classification result would be better if the classifier was also trained using the detection result. However, we would like to evaluate the performance independent of the detector, so the ground truth was used.

Performance comparison: two-stage method

In this experiment, we compared the proposed two-stage method (Model 4) with the detector-only model (Model 3). Since atypical HEM and HCC have the same appearance and require further CT or MR examination, distinguishing between them is unnecessary. We compared the results of Model 3 with the proposed two-stage method. HCC was considered positive, while HEM was considered negative. The incorrect detection of Model 4 was not classified but would be considered as getting a negative (HEM) classification. The accuracy and the recall rate were calculated based on the number of actual HCC (not the number of detected areas). If an HCC lesion was

Table 3 Results of YOLOv4 Detector and ResNet50 Classifier on 3x3 confusion matrix.

Predicted Class	Actual Class			
Class	**HCC**	**Atypical HEM**	**Typical HEM**	**Total**
HCC	337	59	98	494
Atypical HEM	111	486	135	732
Typical HEM	41	29	687	757
Total	489	574	920	1983

Table 4 Results of modified 3x3 confusion matrix as 2x2 confusion matrix.

Predicted Class	Actual Class		
Class	**HCC**	**HEM**	**Total**
HCC	448	157	605
HEM	41	1337	1378
Total	489	1494	1983

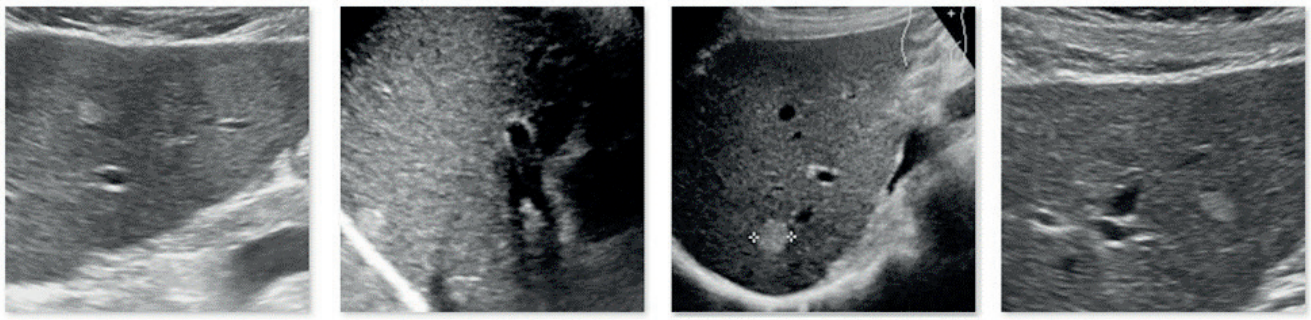

Figure 6 *Small oval-shaped hyperechoic HCC lesions were misclassified as typical HEM by the Classifier.*

Table 5 Results of HCC detection by the detector only and the two-stage methods.

	Model 3	Two-stage method		
		Model 4	Classifier	Overall
Accuracy	0.72	0.86	0.90	0.77*
Precision	0.88*	0.88*		
Recall	Suspicious 0.70	HCC+HEM 0.84	0.90	HCC+HEM 0.76*
	HCC 0.68	HCC 0.78	HCC 0.92	HCC 0.72*
F1-score	0.78	0.86*		
Negative predictive value	0.72	0.86		
Mean average precision	0.60	0.76*		

*Note : * the best result for the given evaluation matrix.*

detected more than once, only one instance classified as HCC was enough for further examination and would be considered as correct. The result is presented in Table 5. Except for precision, the proposed two-stage method provided better performance. Both detectors-only and two-stage methods provided the same precision.

The two-stage method outperformed the detector-only method in the experiment, improving the HCC recall from 0.68 to 0.72. This enhancement signifies a meaningful improvement in the ability to correctly identify and detect HCC cases. Despite using out-of-the-box models not specifically designed for medical imaging, the achieved recall rate of 0.72 was comparable to previous findings.[12] The dataset in our work is smaller so the number of the training image was much lower (615 HCC vs 2414 HCC). Furthermore, Tiyarattanachai *et al.*[12] reported a recall rate of 0.74 by setting the accepted IoU threshold to 0.2, which was considered incorrect detection in our work. There were other works that demonstrated high accuracy.[16,17] However, the database was too small to make a solid conclusion.

Our two-stage method allows for easy improvement as the detector and classifier can be trained separately. YOLOv4, the detector used in our study, has been surpassed by the more recent YOLOv8 (available at https://ultralytics.com/yolov8). Replacing YOLOv4 with YOLOv8 would lead to quick improvements in our method. Additionally, while ResNet-50 provided good classification, optimal results could be achieved by pre-training the network with medical images instead of the ImageNet database. We are currently developing a shallow network specifically for lesion classification in liver ultrasound images due to the limitations of training ResNet-50 with a small database.

Limitation

Two limitations of this experiment are dataset limitations and lack of external validation in real clinical settings that could limit the reliability and real-world applicability of the developed model. Additionally, the use of YOLOv4, as a deep learning model, may present challenges in understanding the decision-making process.

Conclusion

The proposed method for HCC detection from ultrasound images is a two-stage approach. In the first stage, a detector was trained to capture all focal liver lesions. In the second stage, the classifier was trained to distinguish HCC, atypical HEM, and typical HEM. The classification of HCC was not strict in the sense that HCC is allowed to be detected as atypical HEM since the future plan for HCC and atypical HEM is the same. The experiment showed that the two-stage method outperformed the detector-only method in HCC detection. The findings suggest that training separate models: detection and

classification models, led to higher efficiency and accuracy in detecting and classifying hepatic lesions.

Conflict of Interests

The authors declare that they have no conflicts of interest.

References

[1] Park HJ, Jang HY, Kim SY, Lee SJ, Won HJ, Byun JH, *et al*. Non-enhanced magnetic resonance imaging as a surveillance tool for hepatocellular carcinoma: comparison with ultrasound. J Hepatol. 2020; 72(4): 718-24. doi: 10.1016/j.jhep.2019.12.001.

[2] Chou R, Cuevas C, Fu R, Devine B, Wasson N, Ginsburg A, *et al*. Imaging techniques for the diagnosis of hepatocellular carcinoma: a systematic review and meta-analysis. Ann Intern Med. 2015; 162(10): 697-711. doi: 10.7326/M14-2509.

[3] Yamakawa M, Shiina T, Nishida N, Kudo M, editors. Computer aided diagnosis system developed for ultrasound diagnosis of liver lesions using deep learning. 2019 IEEE Int. Ultrason Symp; 2019: IEEE. doi: 10.1109/ULTSYM.2019.8925698.

[4] Yang Q, Wei J, Hao X, Kong D, Yu X, Jiang T, *et al*. Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: A multicentre study. EBioMedicine. 2020; 56: 102777. doi: 10.1016/j.ebiom.2020.102777.

[5] Bharti P, Mittal D, Ananthasivan R. Preliminary study of chronic liver classification on ultrasound images using an ensemble model. Ultrason Imaging. 2018; 40(6): 357-79. doi: 10.1177/0161734618787447.

[6] Brehar R, Mitrea D-A, Vancea F, Marita T, Nedevschi S, Lupsor-Platon M, et al. Comparison of deep-learning and conventional machine-learning methods for the automatic recognition of the hepatocellular carcinoma areas from ultrasound images. Sensors. 2020; 20(11): 3085. doi: 10.3390/s20113085.

[7] Ryu H, Shin SY, Lee JY, Lee KM, Kang H-j, Yi J. Joint segmentation and classification of hepatic lesions in ultrasound images using deep learning. Eur Radiol. 2021; 31: 8733-42. doi: 10.1007/s00330-021-07850-9.

[8] Karako K, Mihara Y, Arita J, Ichida A, Bae SK, Kawaguchi Y, *et al*. Automated liver tumor detection in abdominal ultrasonography with a modified faster region-based convolutional neural networks (Faster R-CNN) architecture. Hepatobiliary Surg Nutr. 2022; 11(5): 675-83. doi: 10.21037/hbsn-21-43.

[9] Tosaki T, Yamakawa M, Shiina T. A study on the optimal condition of ground truth area for liver tumor detection in ultrasound images using deep learning. J Med Ultrasound. 2023: 1-10. doi: 10.1007/s10396-023-01301-2.

[10] Zeng X, Wen L, Liu B, Qi X. Deep learning for ultrasound image caption generation based on object detection. Neurocomputing. 2020; 392: 132-41. doi: 10.1016/j.neucom.2018.11.114.

[11] Zhang Y, Dai X, Tian Z, Lei Y, Chen Y, Patel P, *et al*., editors. Liver motion tracking in ultrasound images using attention guided mask R-CNN with long-short-term-memory network. Medical Imaging 2022: Ultrason Imaging; 2022: SPIE. doi: 10.1117/12.2613013.

[12] Tiyarattanachai T, Apiparakoon T, Marukatat S, Sukcharoen S, Geratikornsupuk N, Anukulkarnkusol N, *et al*. Development and validation of artificial intelligence to detect and diagnose liver lesions from ultrasound images. PloS One. 2021; 16(6): e0252882. doi: 10.1371/journal.pone.0252882.

[13] Caturelli E, Pompili M, Bartolucci F, Siena DA, Sperandeo M, Andriulli A, et al. Hemangioma-like lesions in chronic liver disease: diagnostic evaluation in patients. Radiology. 2001; 220(2): 337-42. doi: 10.1148/radiology.220.2.r01au14337.

[14] Song DS, Bae SH. Changes of guidelines diagnosing hepatocellular carcinoma during the last ten-year period. Clin Mol Hepatol. 2012; 18(3): 258-67. doi: 10.3350/cmh.2012.18.3.258.

[15] Zheng S-G, Xu H-X, Liu L-N. Management of hepatocellular carcinoma: the role of contrast-enhanced ultrasound. World J Radiol. 2014; 6(1): 7. doi: 10.4329/wjr.v6.i1.7.

[16] Schmauch B, Herent P, Jehanno P, Dehaene O, Saillard C, Aube C, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. Diagn Interv Imag. 2019; 100(4): 227-33. doi: 10.1016/j.diii.2019.02.009.

[17] Hassan TM, Elmogy M, Sallam E-S. Diagnosis of focal liver diseases based on deep learning technique for ultrasound images. Arab J Sci Eng 2017; 42(8): 3127-40. doi: 10.1007/s13369-016-2387-9.

[18] Cao Z, Duan L, Yang G, Yue T, Chen Q. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. BMC Med. 2019; 19(1): 1-9. doi: 10.1186/s12880-019-0349-x.

[19] Tanaka H, Chiu S-W, Watanabe T, Kaoku S, Yamaguchi T. Computer-aided diagnosis system for breast ultrasound images using deep learning. Phys Med Biol. 2019; 64(23): 235013. doi: 10.1088/1361-6560/ab5093.