Opportunistic Detection of Hepatocellular Carcinoma Using Noncontrast CT and Deep Learning Artificial Intelligence

Chengzhi Peng, MBBS[a], Philip Leung Ho Yu, PhD[b,c], Jianliang Lu, MPhil[a], Ho Ming Cheng, PhD[a], Xin-Ping Shen, MD[d], Keith Wan-Hang Chiu, MD[d,e,f], Wai-Kay Seto, MD[a,g,h]

Abstract

Objective: Hepatocellular carcinoma (HCC) poses a heavy global disease burden; early diagnosis is critical to improve outcomes. Opportunistic screening—the use of imaging data acquired for other clinical indications for disease detection—as well as the role of noncontrast CT have been poorly investigated in the context of HCC. We aimed to develop an artificial intelligence algorithm for efficient and accurate HCC detection using solely noncontrast CTs.

Methods: A 3-D convolutional block attention module (CABM) model was developed and trained on noncontrast multiphasic CT scans. HCC was diagnosed following American Association for the Study of Liver Disease guidelines and confirmed via 12-month clinical composite reference standard. CT observations were reviewed by radiologists; observations in at-risk patients were annotated via the Liver Imaging Reporting and Data System. Internal validation, independent external testing, and sensitivity analyses were performed to evaluate model performance and generalizability.

Results: In all, 2,223 patients were included. The CBAM model achieved an area under the receiver operating curve (AUC) of 0.807 (95% confidence interval [CI] 0.772-0.841) on the internal validation cohort, comparable to radiological interpretation at 0.851 (95% CI 0.820-0.882). Among at-risk patients, cases with definite HCC outcomes, indeterminate scans, and scans with small lesions < 2 cm in size, the model attained AUCs of 0.769 (95% CI 0.721-0.817), 0.815 (95% CI 0.778-0.853), 0.769 (95% CI 0.704-0.834), and 0.773 (95% CI 0.692-0.854). On external testing cohort with 584 patients, the CBAM model achieved an AUC of 0.789 (95% CI 0.750-0.827).

[a]Department of Medicine, School of Clinical Medicine, The University of Hong Kong, Hong Kong.
[b]Professor of the Department of Mathematics and Information Technology and Associate Director of University Research Facility of Data Science and Artificial Intelligence, Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong.
[c]Honorary Professor, Department of Computer Science, The University of Hong Kong, Hong Kong.
[d]Department of Medical Imaging, The University of Hong Kong-Shenzhen Hospital, Shenzhen, China.
[e]Consultant in Radiology, Department of Diagnostic and Interventional Radiology, Queen Elizabeth Hospital, Hong Kong.
[f]Honorary Associate Professor of the Department of Diagnostic Radiology, Department of Diagnostic Radiology, School of Clinical Medicine, The University of Hong Kong, Hong Kong.
[g]Department of Medicine, The University of Hong Kong-Shenzhen Hospital, Shenzhen, China.
[h]Clinical Professor in Gastroenterology and Hepatology, State Key Laboratory of Liver Research, The University of Hong Kong, Hong Kong; Assistant Dean (Research), LKS Faculty of Medicine, The University of Hong Kong.
Corresponding author and reprint requests: Wai-Kay Seto, MD, Department of Medicine, The University of Hong Kong-Shenzhen Hospital, No. 1, Haiyuan 1st Road, Futian District, Shenzhen 518009, China; e-mail: wkseto@hku.hk and Keith Wan Hang Chiu, MD, Department of Medical Imaging, The University of Hong Kong-Shenzhen Hospital, No. 1, Haiyuan 1st Road, Futian District, Shenzhen 518009, China; e-mail: kwhchiu@hku.hk.

Wai-Kay Seto, MD, reports financial support was provided by Innovation and Technology Fund, Health and Medical Research Fund, and United Ally Research Limited; a relationship with AstraZeneca Pharmaceuticals LP that includes speaking and lecture fees; a relationship with Abbott that includes consulting or advisory and speaking and lecture fees; relationships with Pfizer, Inc, Alexion Pharmaceuticals, Inc, Ribo Life Sciences, and Boehringer Ingelheim, Ltd that include funding grants; a relationship with Gilead Sciences that includes consulting or advisory, funding grants, and speaking and lecture fees; and patent #63/382,198 pending to Versitech Limited, Hong Kong and the Education University of Hong Kong. The other authors state that they have no conflict of interest related to the material discussed in this article. All authors are non-partner/non-partnership track/employees.

1546-1440/25/$36.00 ■ https://doi.org/10.1016/j.jacr.2024.12.011

Discussion: The CBAM model achieved a diagnostic accuracy comparable to radiological interpretation during internal validation. Artificial intelligence analysis of noncontrast CTs has a potential role in HCC opportunistic screening.

Key Words: artificial intelligence, CT, deep learning, liver cancer, noncontrast

J Am Coll Radiol 2025;22:249-259.

INTRODUCTION

Liver cancer poses a heavy disease burden globally, being the seventh commonest cancer and the second most deadly [1]. Hepatocellular carcinoma (HCC) is the most common type of liver cancer, accounting for approximately 90% of liver cancer cases [2]. The highest incidence rates are seen in the Asia-Pacific and sub-Saharan regions, in which hepatitis B virus infection is the main risk factor, and traditionally low-incidence areas such as North America and Europe are seeing increased incidence due to rising rates of obesity, diabetes, and metabolic-associated steatotic liver disease [3]. Overall, the number of new liver cancer cases and deaths is predicted to rise by over 55% by 2044 [4].

The prognosis of HCC is generally poor, with a 5-year survival rate of around 6% to 7% [5,6]. Prognosis and survival are strongly influenced by the stage at presentation [7], with the 5-year survival in the United States ranging from 25.7% for localized disease to 3.5% for cases with distant metastases [8]. However, because early-stage liver cancer is often asymptomatic, many patients do not present until the disease is advanced and treatment options are limited [9]. Therefore, early detection and diagnosis are crucial for improving patient outcomes.

Currently, the recommended method of HCC diagnosis is by multiphase CT or MRI, and reported through the CT-MRI Liver Imaging Reporting and Data System (LI-RADS) [10]. The LI-RADS system classifies radiologically detected nodules into five main categories based on malignancy risk, from definitively benign (LR-1) to definitively HCC (LR-5) [11]. However, no definitive diagnosis can be made for the intermediate classes, and the recommended management is to return to surveillance, repeat diagnostic imaging, or perform liver biopsy [12], which can lead to significant diagnostic delays. Additionally, the LI-RADS system also faces other issues, such as limited sensitivity of LR-5 for HCC [13], unclear evidence on the application of ancillary features [14], presence of interreader variability [15], and as potential discrepancies between features on CT and MRI, which follow the same basic algorithm under LI-RADS [14,15].

The potential value of noncontrast CT in HCC detection has been poorly investigated. Many soft tissue tumors have densities similar to that of the background liver, making them difficult to visualize without contrast [16]; therefore, previous studies on the value of noncontrast CT mostly considered only its possible role as an adjunct to portovenous phase or biphasic images in diagnosis [17,18]. However, artificial intelligence (AI), which has been increasingly applied to many fields in health care, may be capable of detecting subtle differences and patterns on noncontrast scans that are not apparent to the human eye. One recent study found a deep learning model was capable of detecting pancreatic cancers on noncontrast CT with a sensitivity of 92.9% and specificity of 99.9%, greatly surpassing radiologist performance [19].

Development of an AI algorithm for HCC detection on noncontrast CT scans would provide many potential benefits. In particular, an accurate and efficient AI algorithm would allow the application of noncontrast CT for opportunistic screening of HCC, and with the advent of technologies such as mobile CT [20], it also has the potential to provide a new accessible option for HCC screening in the general population. This study aims to investigate whether an AI algorithm is capable of detecting HCC on noncontrast CT with a performance matching that of radiological interpretation on multiphasic CT.

METHODS

Data Collection

This study received approval from the institutional review boards (IRBs) at each participating center: Pamela Youde Nethersole Eastern Hospital (IRB number HKECREC-2019-004), Kwong Wah Hospital/Queen Elizabeth Hospital (IRB number KC/KE-19-0056/ER-4), Queen Mary Hospital/The University of Hong Kong (IRB number UW-18-628), and the University of Hong Kong-Shenzhen Hospital (IRB number [2019]294). It was conducted in compliance with all applicable ethical guidelines and local privacy protection requirements.

Multiphasic CT liver scans acquired between March 2013 and August 2020 and relevant clinical information were collected retrospectively from five medical centers in our East Asian locality. All scans were performed on patients aged ≥18 years. All scans had the noncontrast phase available, with at least one untreated liver observation included. The scans were acquired from various commercial CT scanners using different image acquisition protocols (e-only Table S1). Scans performed after locoregional therapy, such as thermal ablation, transarterial chemo-embolization, radio-embolization, and external beam radiation therapy, were excluded as they are evaluated differently under the LI-RADS treatment response algorithm.

All liver observations were reviewed and annotated on multiphasic CT by two board-certified radiologists, each with more than 10 years of expertise in abdominal cross-sectional

imaging, and blinded to the observations' ground truths, patients' clinical information, and results of deep learning analysis. Liver observations in at-risk patients were further categorized using LI-RADS version 2018 into LR-1 to LR-5 or LR-M. For the remaining patients, radiologists graded each observation on the binary outcome of HCC versus non-HCC. Any discrepancies during external testing were resolved by consensus review.

Image Preprocessing

All scans were collected in DICOM format, and the scan range typically extended from the lung base to the iliac crest. Quality checks were conducted and scans with artifacts were excluded. Preprocessing was then performed to standardize and optimize the data for training purposes; all images were windowed using an abdominal soft tissue window of level 50 and width 400 and normalized to [0, 1] to facilitate model training, and then resized from an original size of 512 × 512 × N to a uniform resolution of 128 × 128 × 128 using spline interpolation in order to conserve graphical processing units memory [21,22].

Ground Truth Definitions

A composite clinical reference standard was employed for establishing the ground truth, to ensure representativeness and generalizability. Cases satisfying histological criteria (surgical resection, explant, or excisional biopsy) or standardized radiological criteria (equivalent to LR-5) were classified as definite HCC. For cases not initially satisfying these criteria, relevant clinical and radiological progress over the next 12 months was reviewed, and a diagnosis of probable HCC was made if the aforementioned criteria were satisfied within the 12-month time frame. Because HCC has a doubling time of around 6 months [23], a 12-month time frame ensures the accuracy of our reference standard. A negative diagnosis was made based on negative histology, image stability, or spontaneous reduction or disappearance without treatment. Cases with lesions clinically treated as HCC without definitive diagnosis were classified as real-world HCC to better reflect real-life clinical practices.

We marked cases at risk of HCC development, defined as male patients with chronic hepatitis B aged ≥40 years or female patients aged ≥50 years or patients with any disease etiology and underlying liver cirrhosis, suggested by a Fibrosis-4 score of >1.45 (Fibrosis-4 4 is a noninvasive risk score calculated from patient age and several serum biomarkers), platelet count, aspartate aminotransferase and alanine aminotransferase. Each ground truth diagnosis was reviewed and validated by two clinical investigators.

Model Development and Training

We developed a 3-D convolutional neural network-based deep learning model, the convolutional block attention

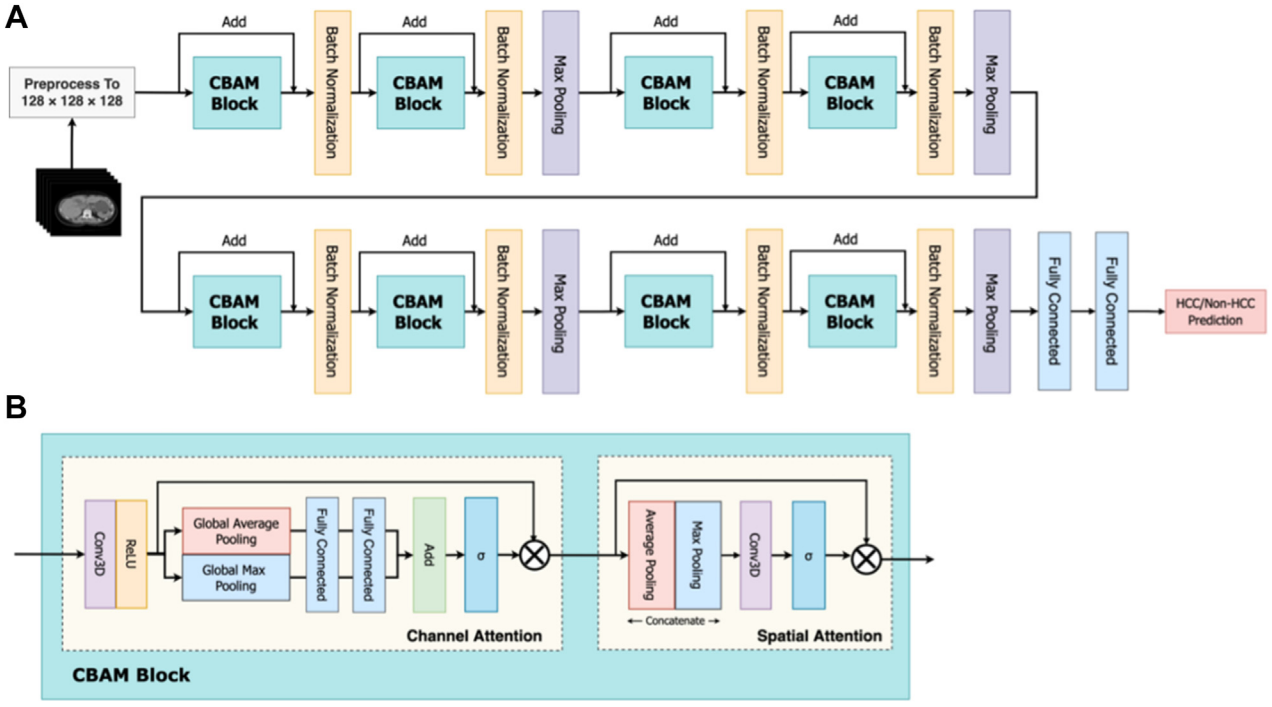

Fig. 1. Model architecture of the deep learning model, the convolutional block attention module (CBAM) model. (A) overall model architecture; (B) structure of the CBAM block. Conv3D = 3-dimensional convolution layer; HCC = hepatocellular carcinoma; ReLU = rectified linear unit.

module (CBAM) model (Fig. 1). Apart from the 3-D nature of the model, which allows it to learn features and patterns in all orientations and preserves cross-slice information, the CBAM model enhances the model's learning ability by incorporating two mechanisms: channel attention, which allows the model to place more emphasis on more essential features, and spatial attention, which enables the model to focus on regions in the image more relevant for diagnosis.

The channel attention module is composed of two parallel pathways containing a global average pooling layer and global max pooling layer, respectively, and then passed separately through two shared, fully connected layers. The outputs of the two pathways are added together and passed through a sigmoid activation function. This trains the module to learn a relative weighting of different channels within the feature map, which is multiplied with the original input and transmitted to the spatial attention module.

The spatial attention module receives the output from the channel attention module, which is passed through an average pooling layer and a max pooling layer, with the outputs concatenated. The concatenated outputs are fed into a 3-D convolutional layer with a single filter and a sigmoid activation function. This yields a weighting map highlighting important spatial regions in the feature map, which is again multiplied with the input to obtain the final output. The detailed structure of the channel and spatial attention modules are shown in Figure 1B.

Tensorflow 2.4.0 (Google Brain, Mountain View, California) was used for development of the deep learning model; training and validation were performed on four NVIDIA (Nvidia Corporation, Santa Clara, California) V100 80G graphical processing units. Several methods of real-time data augmentation were adopted, including horizontal flipping, vertical flipping, 90° rotations, and combinations of these techniques. Stochastic gradient descent optimizer was used with a learning rate of 0.001, and binary cross-entropy was adopted as the loss function. The model was trained for 100 epochs and saved at each epoch. The model, which minimized the loss function on the internal validation dataset, was chosen as the final model.

The dataset was divided randomly by a ratio of 7:3 into training and internal validation sets. We further tested the robustness of our model by evaluating it on an independent external testing dataset from two centers, which had the same ground truth definitions as the internal validation cohort.

Statistical Analyses

Statistical analyses were performed using Python 3.10.9 (Python Software Foundation, Wilmington, Delaware) and R version 4.3.0 (R Core Development Team, Vienna, Austria). The overall performance of our binary classifier model was assessed by computing the area under the receiver operating curve (AUC) using the pROC package in R. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated with the cutoff threshold chosen, which maximizes the Youden index [24]. Additionally, sensitivity analyses were performed on the internal validation dataset, in which the model was tested on several subgroups: (1) patients at risk for HCC, defined as patients with cirrhosis of any etiology, or male patients aged $\geq$40 years or female patients aged $\geq$50 years with chronic hepatitis B; (2) exclusion of real-world HCC cases; (3) cases with indeterminate nodules as determined by radiological interpretation and LI-RADS (LR-2, LR-3, LR-4, and LR-M); (4) cases with lesions $<$2 cm in size and cases with lesions 2 to 5 cm in size.

Heat Map Visualization

To improve the explainability and transparency of our model, we developed heat maps for our CT scans to better visualize which regions of the liver were most crucial for the model in making a diagnosis of HCC versus non-HCC. The gradient-weighted class activation mapping [25] method was adopted, which calculates the gradients of a model's prediction and uses it to generate a localization map. The final convolutional layer (ie, the convolutional layer in the spatial attention module of the last CBAM) was used for heat map generation.

RESULTS

The basic clinical characteristics of patients in the training and internal validation cohorts are given in Table 1. In all, 2,574 cases from five centers were screened, with 2,223 (86.4%) patients eventually included. The patient disposition of the study is shown in Figure 2. Overall, included patients had a mean age of 58.3 (±14.3) years, with 1,317 (59.2%) having underlying chronic liver disease and 1,178 (53.0%) individuals at risk for HCC. There were 675 (30.4%) patients diagnosed to have HCC, among whom 497 (73.6%) were diagnosed by radiological or histological criteria, and the other 178 (26.4%) were classified as real-world HCC.

Diagnostic Performance

The diagnostic performance of the CBAM deep learning model on noncontrast CT and radiological interpretation of multiphasic CT on different datasets is shown in Table 2, and the AUC curves are shown in Figure 3. The CBAM model achieved an AUC of 0.807 (95% confidence interval [CI] 0.772-0.841) on the internal validation dataset, which was comparable to that of radiological interpretation (0.851, 95% CI 0.820-0.882). The sensitivity of the CBAM model was 0.834 (95% CI 0.778-0.878), specificity was 0.633

Table 1. Baseline characteristics of the 2,223 patients used for training and internal validation, and the 584 patients used for external testing

Parameter	Training and Internal Validation			External Testing		
	All Patients (n = 2,223)	HCC (n = 675)	Non-HCC (n = 1,548)	All Patients (n = 584)	HCC (n = 223)	Non-HCC (n = 361)
Age, y	58.3 (±14.3)	60.7 (±13.5)	57.3 (±14.5)	66.5 (±13.0)	63.8 (±9.0)	68.1 (±14.7)
Male patients (%)	1,362 (61.3)	541 (80.1)	821 (53.0)	386 (66.1)	189 (84.8)	197 (54.6)
Chronic liver disease (%)						
Present	1,317 (59.2)	608 (90.1)	709 (45.8)	256 (43.8)	213 (95.5)	43 (11.9)
HBV	1,020 (77.4)	521 (85.7)	499 (70.4)	213 (83.2)	176 (82.6)	37 (86.0)
HCV	95 (7.2)	41 (6.7)	54 (7.6)			
MASLD	137 (10.4)	10 (1.6)	127 (17.9)	40 (15.6)	39 (18.3)	1 (2.3)
Alcoholic liver disease	83 (6.3)	49 (8.1)	34 (4.8)	26 (10.2)	25 (11.7)	1 (2.3)
≥2 liver diseases	75 (5.7)	40 (6.6)	35 (4.9)	41 (16.0)	41 (19.2)	0 (0.0)
Others	63 (4.8)	31 (5.1)	32 (4.5)	18 (7.0)	14 (6.6)	4 (9.3)
No known liver disease	906 (40.8)	67 (9.9)	839 (54.2)	328 (56.2)	10 (4.5)	318 (88.1)
HCC risk (%)						
At risk	1,178 (53.0)	607 (89.9)	571 (36.9)	222 (38.0)	186 (83.4)	36 (9.97)
Not at risk	711 (31.4)	52 (7.7)	659 (42.6)	293 (50.2)	36 (16.1)	257 (71.2)
Indeterminate	334 (15.0)	16 (2.4)	318 (20.5)	69 (11.8)	1 (0.4)	68 (18.8)
Albumin	41 (37-44)	39 (34-43)	42 (39-45)	38 (34-42)	41 (37-44)	37 (31-40)
Bilirubin	13 (9-19)	15 (10-24)	12 (8-17)	10 (7-14)	12 (9-16)	9 (7-13)
ALT, U/L	27 (18-45)	37 (24-64)	24 (16-36)	25 (18-39)	32 (21-47)	22 (16-36)
AST, U/L	33 (23-60)	51 (31-93)	26 (19-38)	33 (25-52)	35 (26-57)	28 (21-39)
ALP, U/L	80 (64-112)	102 (74-167)	75 (61-95)	81 (63-103)	85 (69-115)	75 (60-96)
Platelet count, $\times 10^9$/L	200 (145-255)	168 (116-237)	213 (166-264)	205 (156-271)	169 (126-215)	243 (189-302)
INR	1.1 (1.0-1.2)	1.1 (1.1-1.2)	1.1 (1.0-1.2)	1.1 (1.0-1.2)	1.1 (1.0-1.1)	1.1 (1.1-1.2)
AFP, ng/mL	4 (3-21)	30 (5-506)	3 (2-5)	8 (3-99)	15 (4-145)	3 (2-4)

Continuous data presented as median (interquartile range). HCV data were unavailable for the external testing cohort. AFP = alpha-fetoprotein; ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; HBV = hepatitis B virus; HCV = hepatitis C virus; HCC = hepatocellular carcinoma; INR = international normalized ratio; MASLD = metabolic associated steatotic liver disease.

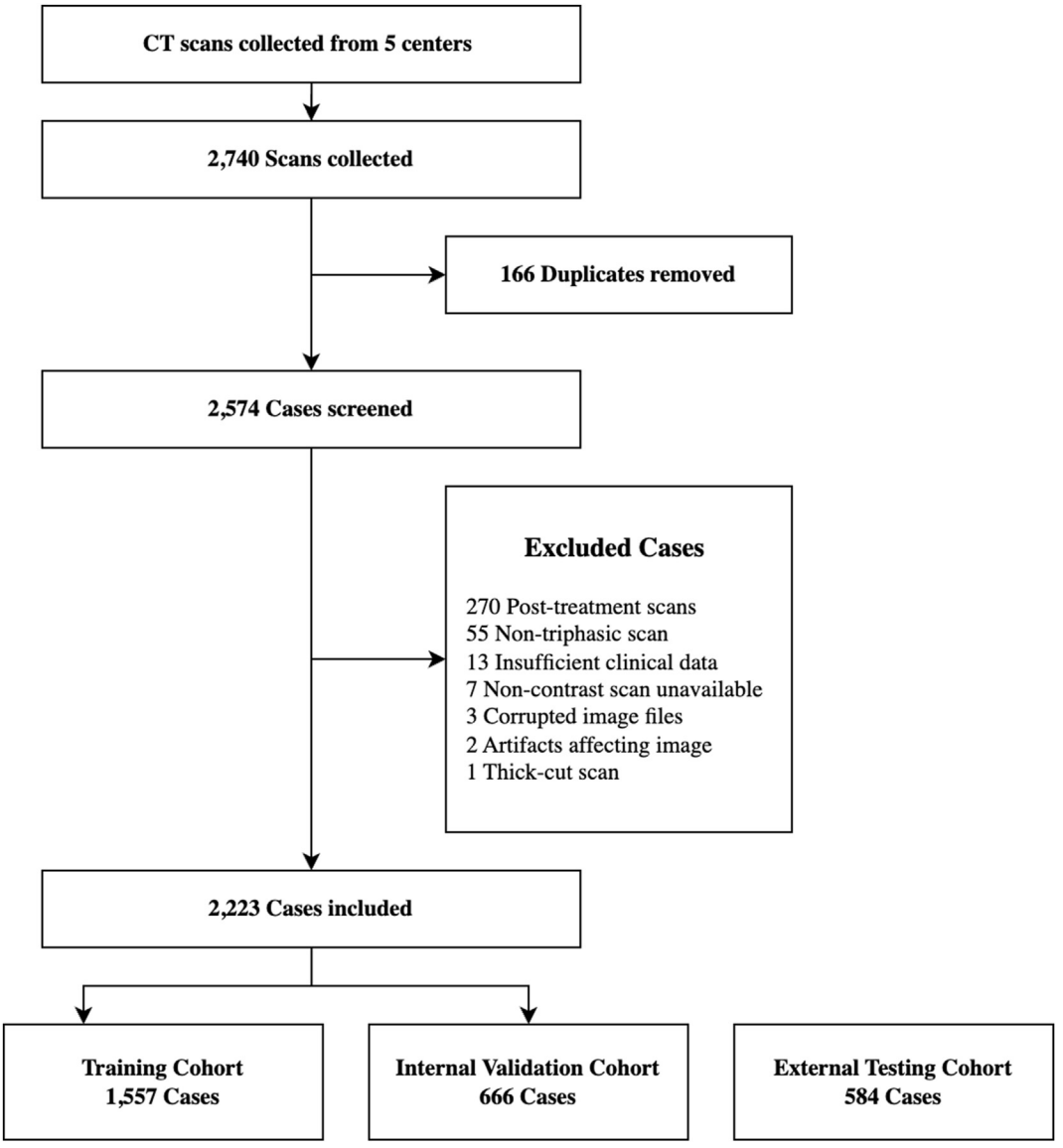

Fig. 2. Patient disposition of the current study. Posttreatment scans after percutaneous, transcatheter, or radiation therapy were excluded as they are assessed differently under Liver Imaging Reporting and Data System. Two scans had artifacts due to previously inserted intra-abdominal vascular coils and were excluded.

(95% CI 0.588-0.676), PPV was 0.513 (95% CI 0.460-0.566), and NPV was 0.892 (95% CI 0.853-0.921).

Sensitivity Analyses

Sensitivity analyses of different scenarios demonstrated CBAM to have comparable diagnostic performance (Table 2). Among at-risk patients, CBAM had a slightly lower AUC of 0.769 (95% CI 0.721-0.817), as compared with LI-RADS (0.849, 95% CI 0.816-0.883). Among patients with definite HCC outcomes, CBAM attained an AUC of 0.815 (95% CI 0.778-0.853), as compared with AUC 0.867 (95% CI 0.832-0.901) for radiological interpretation. The AUCs of CBAM for indeterminate lesions, lesions 2 to 5 cm, and lesions <2 cm were 0.769 (95% CI 0.704-0.834), 0.756 (95% CI 0.697-0.814), and 0.773 (95% CI 0.692-0.854), respectively, comparable to that of radiological interpretation (AUCs 0.636-0.867). NPVs of CBAM were high, ranging from 0.891 to 0.966.

External Testing

We further validated our model on an independent external testing cohort, comprised of 584 patients, among which 223 (38.2%) were categorized as having HCC. The basic clinical characteristics of patients in the external testing cohort are shown in Table 1. The CBAM model achieved an AUC of 0.789 (95% CI 0.750-0.827) on the external dataset, which was lower compared with radiological interpretation at 0.927 (95% CI 0.904-0.950). The CBAM model had a sensitivity of 0.700 (95% CI 0.636-0.756), specificity of

Table 2. Diagnostic performance of the deep learning model (CBAM) on noncontrast CT and radiological interpretation of multiphasic CT for patient-level diagnosis of HCC in the internal validation and external testing dataset and various subgroups

Dataset	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
CBAM					
Internal validation	0.807 (0.772-0.841)	0.834 (0.778-0.878)	0.633 (0.588-0.676)	0.513 (0.460-0.566)	0.892 (0.853-0.921)
At-risk	0.769 (0.721-0.817)	0.842 (0.784-0.887)	0.544 (0.469-0.617)	0.672 (0.610-0.729)	0.756 (0.673-0.823)
Exclusion of real-world HCC	0.815 (0.778-0.853)	0.843 (0.778-0.891)	0.633 (0.587-0.676)	0.451 (0.396-0.508)	0.918 (0.882-0.944)
Indeterminate nodules by radiological interpretation	0.769 (0.704-0.834)	0.780 (0.659-0.866)	0.618 (0.544-0.688)	0.411 (0.324-0.503)	0.892 (0.823-0.936)
Lesions 2-5 cm	0.756 (0.697-0.814)	0.815 (0.717-0.884)	0.610 (0.541-0.675)	0.458 (0.379-0.540)	0.891 (0.827-0.933)
Lesions <2 cm	0.773 (0.692-0.854)	0.682 (0.473-0.836)	0.679 (0.624-0.730)	0.136 (0.084-0.213)	0.966 (0.932-0.984)
External testing	0.789 (0.750-0.827)	0.700 (0.636-0.756)	0.734 (0.686-0.777)	0.619 (0.558-0.677)	0.798 (0.752-0.838)
Radiological interpretation					
Internal validation	0.851 (0.820-0.882)	0.711 (0.646-0.768)	0.991 (0.978-0.997)	0.974 (0.935-0.990)	0.881 (0.850-0.906)
At-risk LI-RADS	0.849 (0.816-0.883)	0.711 (0.642-0.770)	0.988 (0.958-0.997)	0.985 (0.948-0.996)	0.754 (0.694-0.806)
Exclusion of real-world HCC	0.867 (0.832-0.901)	0.742 (0.669-0.804)	0.991 (0.977-0.996)	0.967 (0.919-0.987)	0.915 (0.886-0.937)
Lesions 2-5 cm	0.732 (0.677-0.787)	0.469 (0.364-0.577)	0.995 (0.972-0.999)	0.974 (0.868-0.995)	0.822 (0.769-0.865)
Lesions <2 cm	0.636 (0.541-0.732)	0.273 (0.132-0.482)	1.000 (0.987-1.000)	1.000 (0.610-1.000)	0.949 (0.918-0.968)
External testing	0.927 (0.904-0.950)	0.857 (0.804-0.896)	0.997 (0.984-1.000)	0.995 (0.971-0.999)	0.918 (0.887-0.942)

AUC = area under curve; CBAM = convolutional block attention module; CI = confidence interval; HCC = hepatocellular carcinoma; LI-RADS = Liver Imaging Reporting and Data System; NPV = negative predictive value; PPV = positive predictive value.

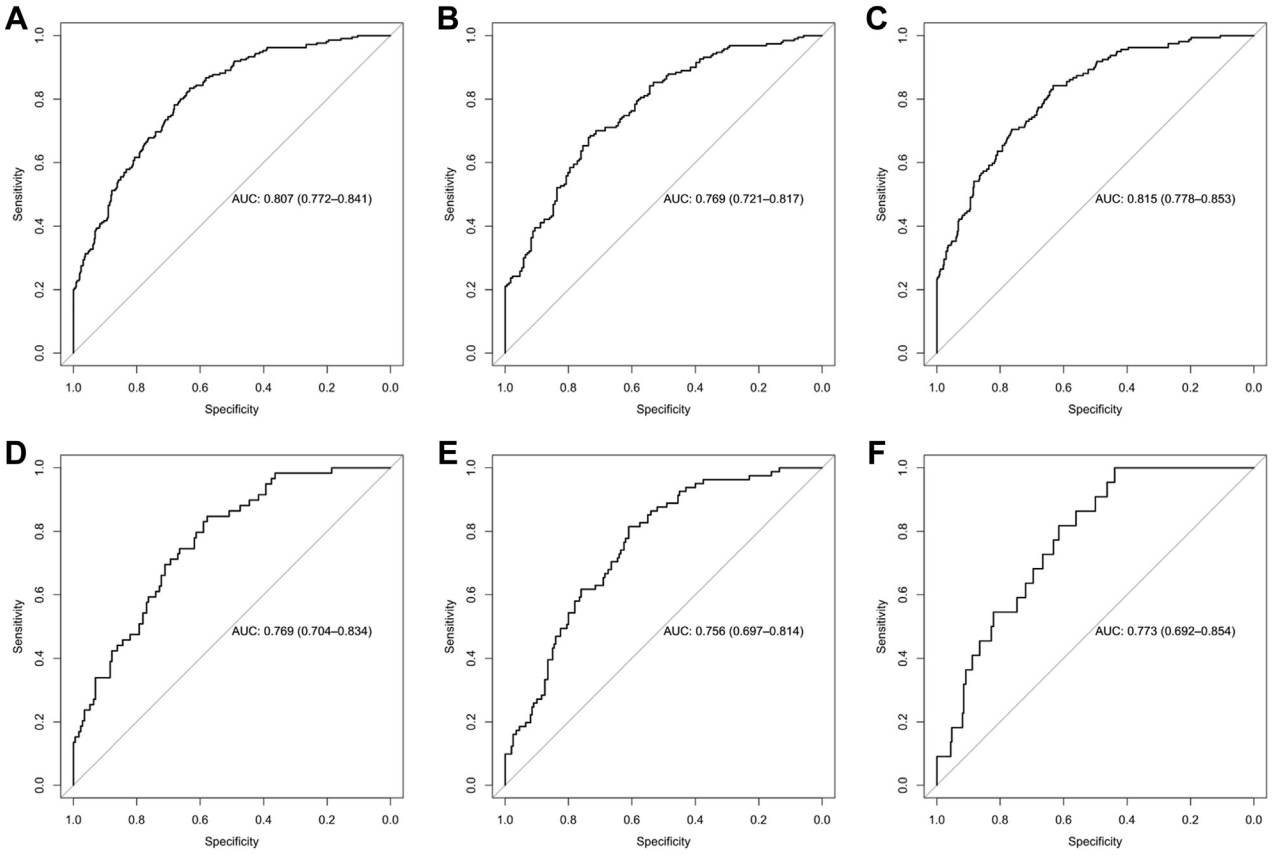

Fig. 3. Area under the receiver operating curve (AUC) of the convolutional block attention module model on (A) the internal validation cohort, (B) at-risk patients, (C) cases after exclusion of real-world hepatocellular carcinoma, (D) cases with indeterminate lesions, (E) cases with lesions 2 to 5 cm in size, (F) cases with lesions <2 cm in size.

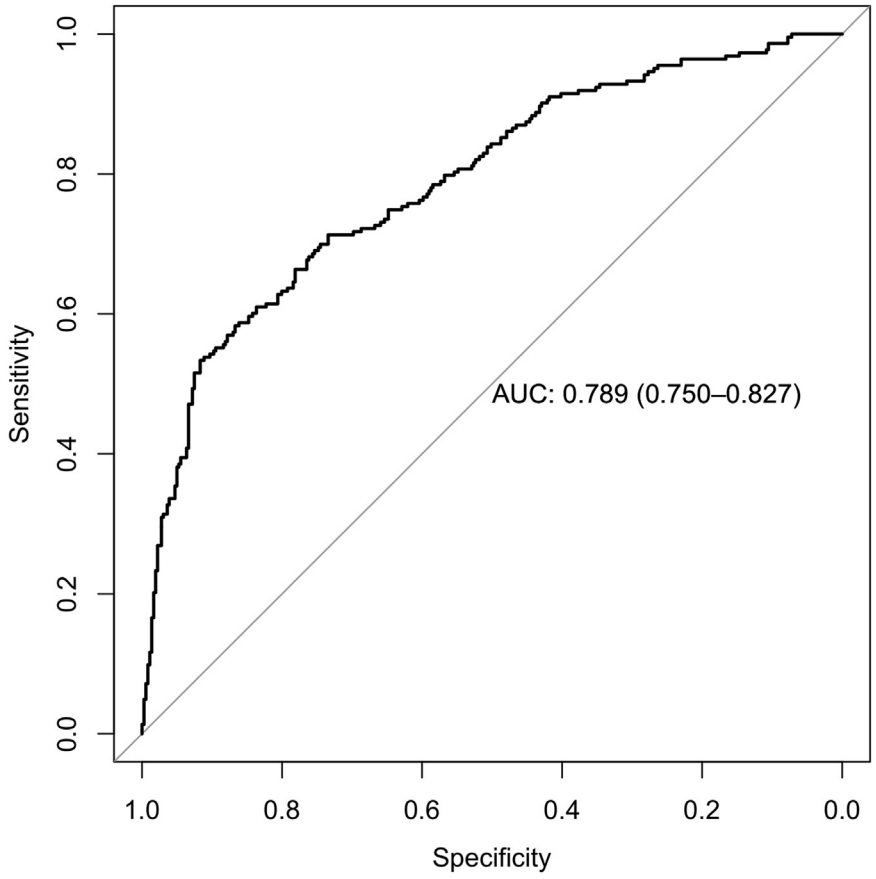

Fig. 4. Area under the receiver operating curve (AUC) of the convolutional block attention module model on the external testing dataset.

0.734 (95% CI 0.686-0.777), PPV of 0.619 (95% CI 0.558-0.677), and NPV of 0.798 (95% CI 0.752-0.838). The AUC curve of the CBAM model on the external testing cohort is shown in Figure 4.

Heat Map Visualization

Heat maps of the CBAM model's predictions on internal validation cases were generated; the results are shown in Figure 5A and B. By calculating the importance of each image position in outputting the prediction for each class, the gradient-weighted class activation mapping heat maps visually illustrate the areas most crucial in leading the model to make its final decision, which in the figures are highlighted in red and blue for HCC and non-HCC, respectively. As shown in the figures, the regions highlighted in the heat maps generally correspond to the location of the lesion as determined based on multiphasic scans.

DISCUSSION

Our study shows that deep learning models are able to diagnose HCC on noncontrast CT with an accuracy comparable to that of radiological interpretation on multiphasic CT; the model was robust in different subgroups with similar diagnostic performance. This demonstrates the great potential of deep learning models in detecting HCC using noncontrast CT scans. Although noncontrast phase has generally been overlooked for the purpose of HCC diagnosis due to poor tissue differentiation, our study shows that AI can overcome these difficulties to a great extent and has the capacity to differentiate HCC and non-HCC, using solely noncontrast CT, at a performance level comparable to that of expert radiological interpretation on multiphasic scans.

The strength of our approach lies in several factors. First is the quantity, quality, and representativeness of our data. Compared with traditional machine learning methods, deep learning models have the capacity to achieve higher accuracy but typically require much larger datasets to reach an optimal performance [26]. Our dataset with more than 2,000 CT scans from multiple medical centers is well suited to meet these high demands. Moreover, we did not adopt a purely histology-based ground truth [27] but instead chose a clinical composite reference standard, taking into consideration various clinical

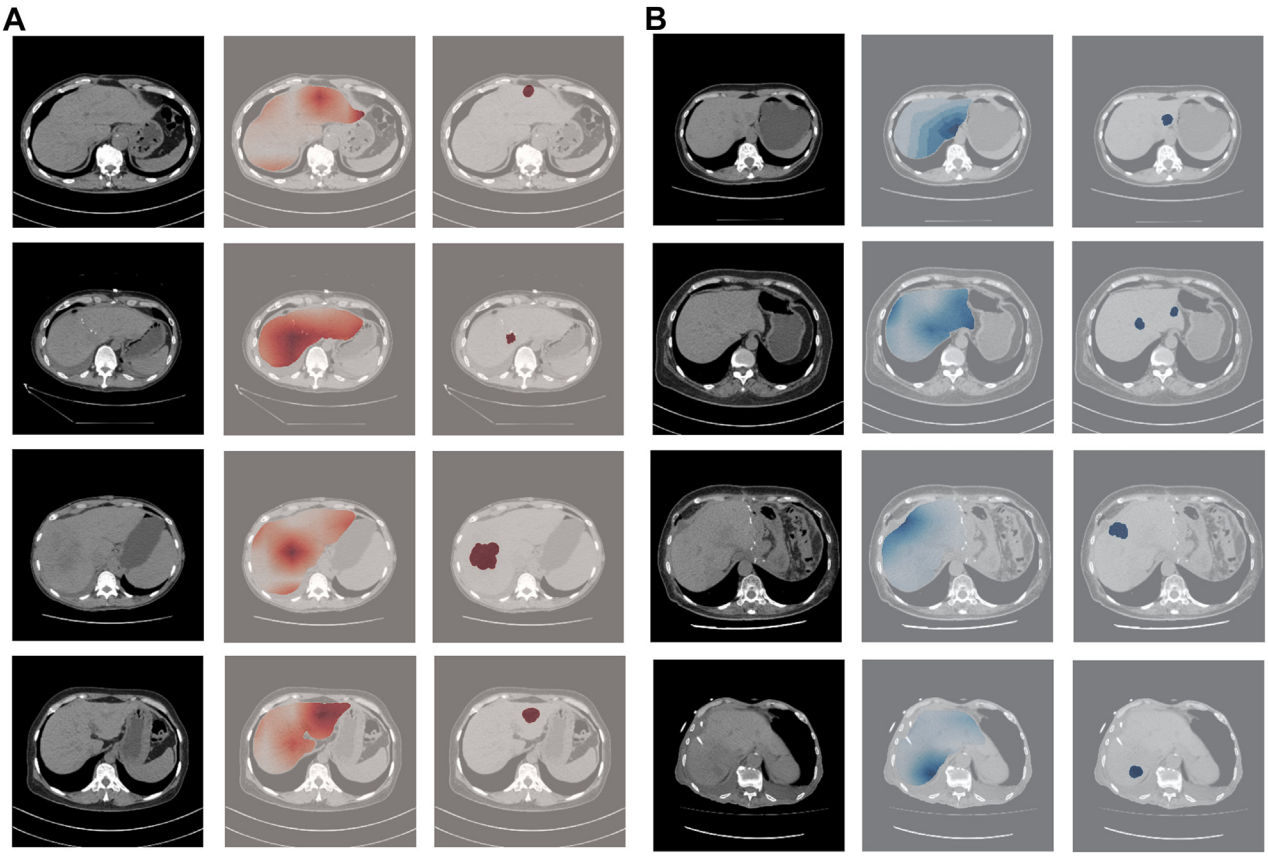

Fig. 5. Heat map visualization of the convolutional block attention module model's classification results for (A) four hepatocellular carcinoma (HCC) cases and (B) four non-HCC cases. The first column shows the CT scan in the noncontrast phase, the second column shows the heat map, and the third column shows the lesion contour, which was manually delineated based on portovenous scans. The red and blue colors show the most crucial regions that led the model to make a prediction of HCC or non-HCC, respectively.

and radiological parameters; this greatly expands our patient pool and also makes it more reflective of real-world settings. An adequately large and representative dataset is essential for training a model with high performance and good generalizability.

The power and capacity of our model architecture is also a crucial factor. For our model, a deep learning network incorporating attention mechanisms was adopted, allowing various features and spatial regions in an image to be highlighted during learning; our 3-D approach enabled the extraction of more abundant cross-slice radiological information, which would otherwise have been lost in 2-D models. We believe that all these factors were essential in enhancing the learning capacity of our model, allowing it to achieve radiologist-comparable performance on noncontrast scans with poor tissue differentiation.

Apart from model performance, we also aimed to improve model interpretability by the incorporation of heat maps. Heat maps generated by the model, which highlight the areas deemed most crucial by the model in discriminating HCC or non-HCC, can help users better understand the model's decision-making process by providing a direct visual representation. Additionally, they may potentially facilitate further assessment by alerting radiologists to possible regions of interest, enhancing the efficiency and accuracy of the diagnostic workflow.

These results have significant clinical implications, in particular for possible application as a method of opportunistic screening, which in radiology describes the practice of using information from imaging scans obtained for an unrelated indication [28]. CT scans are particularly well suited for the purpose of opportunistic screening: it is commonly used in clinical practice for a variety of indications, with more than 70 million scans performed annually in the United States [29] and approximately 300 million globally [30]; moreover, it provides detailed and comprehensive information of the imaged region, regardless of the original clinical indication [31]. Opportunistic screening is also reasonable in populations with high HCC incidence and prevalence, such as East Asia, Southeast Asia, and Africa [32].

An AI algorithm can serve as an efficient means of harnessing and interpreting this unused incidental information, with no additional radiation exposure and minor additional cost. Moreover, with its shorter procedure time, lower risk of adverse effects, and lower dosage of radiation exposure, population-based screening for HCC using noncontrast CT could be possible with a validated AI model, similar to low-dose CT for screening of lung cancer, which has been widely validated in many studies [33]. Patients who screen positive may be referred for further investigations such as contrast-enhanced CT or MRI; in the future, if lesion detection models are developed, additional characteristics such as lesion size may also be taken into consideration. Because noncontrast MRI has already been found to be a potentially more cost-effective alternative to the currently established modality of screening by ultrasound [34], an algorithm based on noncontrast CT and AI may serve as a more efficient and cost-effective screening method.

Limitations

Although the AUC of our model was comparable to that of radiological interpretation, there is still room for improvement in terms of performance; in particular the NPV, which is crucial for a model intended for screening purposes. Our model achieved an NPV of 89.2% on the internal validation cohort but only 79.8% on the external testing cohort; in comparison, in the context of HCC surveillance, the NPV of ultrasound has been reported to be around 80% [35], and the NPV of alpha-fetoprotein has been reported to be as high as 97% [36]. The use of a weighted Youden index in determining the prediction threshold may help better fine-tune the balance between sensitive and specificity, to improve the NPV while maintaining an acceptable PPV. Further improvements in the deep learning model may be attained by various modifications in the training process, such as applying a class weighting in the loss function to penalize false-negatives more, adjusting the data augmentation scheme, or adopting other oversampling techniques to help the model learn better representations of the positive cases, thereby reducing false-negatives. Also, despite the use of various techniques to reduce overfitting, the performance of our model on the external testing dataset was still lower than that of the internal validation set. Although this was expected due to the significantly different distribution of HCC and non-HCC cases, it does indicate some level of overfitting and shows room for improvement in terms of the model's generalizability. In the future, expanding the dataset to include more diverse patients and settings may allow the development of a more robust model and provide further evidence for a noncontrast CT-based opportunistic screening program for HCC.

TAKE-HOME POINTS

- The CBAM deep learning model achieved a patient-level AUC of 0.807 for HCC detection on the internal validation cohort, comparable with that of radiological interpretation using contrast-enhanced scans, but was slightly less efficacious in the external cohort with an AUC of 0.789.
- The CBAM deep learning model was robust and demonstrated comparable diagnostic performance for different subgroups, including for indeterminate lesions and small lesions.
- AI models combined with noncontrast CT may have potential future applications as a method of opportunistic screening for HCC.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Lung Yi Mak, Xianhua Mao, Sarah Yu, Juan Wu, Hung Tat Chan, King Ming Kwok, and Wai Kuen Kan for the acquisition of image data from the different institutions that participated in this present study, as well as Man-Fung Yuen for his intellectual input and critical review of the study. This study was supported by the Innovation and Technology Fund, The Government of the Hong Kong SAR (reference no.: ITS/122/18FP); Health and Medical Research Fund (reference no.: 07182346); and United Ally Research Limited, a subsidiary of Hong Kong Sanatorium and Hospital Limited.

ADDITIONAL RESOURCES

Additional resources can be found online at: https://doi.org/10.1016/j.jacr.2024.12.011.

REFERENCES

1. McGlynn KA, Petrick JL, El-Serag HB. Epidemiology of hepatocellular carcinoma. Hepatology 2021;73(Suppl 1):4-13.
2. Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. Nat Rev Dis Primers 2021;7:6.
3. El-Serag HB. Epidemiology of hepatocellular carcinoma. In: The Liver. Hoboken, NJ: John Wiley & Sons Ltd; 2020:758-72.
4. Rumgay H, Arnold M, Ferlay J, et al. Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol 2022;77:1598-606.
5. Chen JG, Zhu J, Zhang YH, et al. Liver cancer survival: a real world observation of 45 years with 32,556 cases. J Hepatocell Carcinoma 2021;8:1023-34.
6. Borie F, Bouvier AM, Herrero A, et al. Treatment and prognosis of hepatocellular carcinoma: a population based study in France. J Surg Oncol 2008;98:505-9.
7. Johnson P, Berhane S, Kagebayashi C, et al. Impact of disease stage and aetiology on survival in hepatocellular carcinoma: implications for surveillance. Br J Cancer 2017;116:441-7.
8. Momin BR, Pinheiro PS, Carreira H, Li C, Weir HK. Liver cancer survival in the United States by race and stage (2001-2009): findings from the CONCORD-2 study. Cancer 2017;123(Suppl 24):5059-78.

9. Nordenstedt H, White DL, El-Serag HB. The changing pattern of epidemiology in hepatocellular carcinoma. Dig Liver Dis 2010;42(Suppl 3):S206-14.
10. Marrero JA, Kulik LM, Sirlin CB, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 2018;68:723-50.
11. Mitchell DG, Bruix J, Sherman M, Sirlin CB. LI-RADS (Liver Imaging Reporting and Data System): summary, discussion, and consensus of the LI-RADS Management Working Group and future directions. Hepatology 2015;61:1056-65.
12. Chernyak V, Fowler KJ, Kamaya A, et al. Liver Imaging Reporting and Data System (LI-RADS) version 2018: imaging of hepatocellular carcinoma in at-risk patients. Radiology 2018;289:816-30.
13. Chernyak V, Fowler KJ, Do RKG, et al. LI-RADS: looking back, looking forward. Radiology 2023;307:e222801.
14. Kim YY, Choi JY, Sirlin CB, An C, Kim MJ. Pitfalls and problems to be solved in the diagnostic CT/MRI Liver Imaging Reporting and Data System (LI-RADS). Eur Radiol 2019;29:1124-32.
15. Elsayes KM, Fowler KJ, Chernyak V, et al. User and system pitfalls in liver imaging with LI-RADS. J Magn Reson Imaging 2019;50:1673-86.
16. Baron RL. Understanding and optimizing use of contrast material for CT of the liver. AJR Am J Roentgenol 1994;163:323-31.
17. Oliver JH 3rd, Baron RL, Federle MP, Rockette HE Jr. Detecting hepatocellular carcinoma: value of unenhanced or arterial phase CT imaging or both used in conjunction with conventional portal venous phase contrast-enhanced CT imaging. AJR Am J Roentgenol 1996;167:71-7.
18. Doyle DJ, O'Malley ME, Jang HJ, Jhaveri K. Value of the unenhanced phase for detection of hepatocellular carcinomas 3 cm or less when performing multiphase computed tomography in patients with cirrhosis. J Comput Assist Tomogr 2007;31:86-92.
19. Cao K, Xia Y, Yao J, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. Nat Med 2023;29:3033-43.
20. Goertz L, Al-Sewaidi Y, Habib M, et al. State-of-the-art mobile head CT scanner delivers nearly the same image quality as a conventional stationary CT scanner. Sci Rep 2024;14:6393.
21. Fu F, Shan Y, Yang G, et al. Deep learning for head and neck CT angiography: stenosis and plaque classification. Radiology 2023;307:e220996.
22. Zhang R, He A, Xia W, et al. Deep learning-based fully automated segmentation of regional muscle volume and spatial intermuscular fat using CT. Acad Radiol 2023;30:2280-9.
23. Ronot M, Fouque O, Esvan M, Lebigot J, Aubé C, Vilgrain V. Comparison of the accuracy of AASLD and LI-RADS criteria for the non-invasive diagnosis of HCC smaller than 3 cm. J Hepatol 2018;68:715-23.
24. Schisterman EF, Faraggi D, Reiser B, Hu J. Youden index and the optimal threshold for markers with mass at zero. Stat Med 2008;27:297-315.
25. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization 2017 IEEE International Conference on Computer Vision (ICCV) 2017:618-26.
26. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep learning in radiology: does one size fit all? J Am Coll Radiol 2018;15(3 Pt B):521-6.
27. Elmore JG, Barnhill RL, Elder DE, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. BMJ 2017;357:j2813.
28. Engelke K, Chaudry O, Bartenschlager S. Opportunistic screening techniques for analysis of CT scans. Curr Osteoporos Rep 2023;21:65-76.
29. Brenner DJ. Slowing the increase in the population dose resulting from CT scans. Radiat Res 2010;174:809-15.
30. Schöckel L, Jost G, Seidensticker P, Lengsfeld P, Palkowitsch P, Pietsch H. Developments in x-ray contrast media and the potential impact on computed tomography. Invest Radiol 2020;55:592-7.
31. Pickhardt PJ. Value-added opportunistic CT screening: state of the art. Radiology 2022;303:241-224.
32. Singal AG, Lampertico P, Nahon P. Epidemiology and surveillance for hepatocellular carcinoma: new trends. J Hepatol 2020;72:250-61.
33. Hoffman RM, Atallah RP, Struble RD, Badgett RG. Lung cancer screening with low-dose CT: a meta-analysis. J Gen Intern Med 2020;35:3015-25.
34. Kim DH, Choi SH, Shim JH, et al. Magnetic resonance imaging for surveillance of hepatocellular carcinoma: a systematic review and meta-analysis. Diagnostics (Basel) 2021;11:1665: https://doi.org/10.3390/diagnostics11091665.
35. Ravindran N, Agarwal A, Li F, Thuluvath PJ. Liver ultrasound in cirrhosis is inadequate for hepatocellular carcinoma surveillance compared to magnetic resonance imaging. Liver Cancer Int 2022;3:113-8.
36. Tayob N, Corley DA, Christie I, et al. Validation of the updated hepatocellular carcinoma early detection screening algorithm in a community-based cohort of patients with cirrhosis of multiple etiologies. Clin Gastroenterol Hepatol 2020;19:1443-50.e6.