


eGastroenterology Application of artificial intelligence in the diagnosis of hepatocellular carcinoma

Benjamin Koh ¹, Pojsakorn Danpanichkul,² Meng Wang,³ Darren Jun Hao Tan,¹ Cheng Han Ng³

To cite: Koh B, Danpanichkul P, Wang M, *et al.* Application of artificial intelligence in the diagnosis of hepatocellular carcinoma. *eGastroenterology* 2023;1:e100002. doi:10.1136/egastro-2023-100002

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/egastro-2023-100002>).

Received 30 June 2023

Accepted 20 September 2023

ABSTRACT

Hepatocellular carcinoma (HCC) is a major cause of cancer-related deaths worldwide. This review explores the recent progress in the application of artificial intelligence (AI) in radiological diagnosis of HCC. The Barcelona Classification of Liver Cancer criteria guides treatment decisions based on tumour characteristics and liver function indicators, but HCC often remains undetected until intermediate or advanced stages, limiting treatment options and patient outcomes. Timely and accurate diagnostic methods are crucial for enabling curative therapies and improving patient outcomes. AI, particularly deep learning and neural network models, has shown promise in the radiological detection of HCC. AI offers several advantages in HCC diagnosis, including reducing diagnostic variability, optimising data analysis and reallocating healthcare resources. By providing objective and consistent analysis of imaging data, AI can overcome the limitations of human interpretation and enhance the accuracy of HCC diagnosis. Furthermore, AI systems can assist healthcare professionals in managing the increasing workload by serving as a reliable diagnostic tool. Integration of AI with information systems enables comprehensive analysis of patient data, facilitating more informed and reliable diagnoses. The advancements in AI-based radiological diagnosis hold significant potential to improve early detection, treatment selection and patient outcomes in HCC. Further research and clinical implementation of AI models in routine practice are necessary to harness the full potential of this technology in HCC management.

INTRODUCTION

Hepatocellular Carcinoma (HCC) one of the leading cancers globally.^{1 2} A census conducted by the global cancer observatory in 2020 showed rising rates of HCC globally, particularly within North African and East Asian populations.¹ The aetiology of HCC is unique in its potential for curative therapy through surgical resection, radiofrequency ablation and liver transplants for patients with early stages of the disease.^{3–7} Currently, the most widely accepted classification of HCC is the Barcelona Classification of Liver Cancer (BCLC) criteria which characterise

HCCs based on the number and size of nodules, supported by secondary parameters such as liver function indicators.⁸ The BCLC guidelines also guide treatment options by physicians and curative treatments are generally reserved only for BCLC A and BCLC 0 patients.⁸ Unfortunately, symptomatic presentation of HCC generally only occurs in intermediate or advanced stage disease (BCLC B and C, respectively) where treatment options and efficacy are limited.^{9–11} As such, timely and accurate diagnostic methods are crucial as they are open to the possibility for more curative therapies and are likely to significantly improve patient outcomes.^{12–14} Current first-line surveillance methods include laboratory parameters such as alpha-feto protein (AFP) and imaging techniques such as liver ultrasounds, which are preferred due to their non-invasive nature.^{15–17} Recent advances in artificial intelligence (AI) have opened new applications in conjunction with existing diagnostic tools. Improvements in deep learning (DL) and neural network models have seen comparable or in some cases superior outcomes compared with physicians in the diagnosis of disease.^{18–20} Within HCC, AI models have been incorporated in the radiological and histological detection of the disease. This paper will discuss the recent advances of AI in the radiological diagnosis of HCC in the hope of providing a clearer landscape of the present and future of its application.

TERMINOLOGY

The field of AI has constantly evolved since its inception with new concepts and applications arising every few years. As such, adequate understanding of frequently used terminologies is indispensable in discussions on the topic. This section will provide a brief introduction to commonly used nomenclature



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Medicine, National University of Singapore Yong Loo Lin School of Medicine, Singapore

²Department of Microbiology, Chiang Mai University, Chiang Mai, Thailand

³Division of Gastroenterology and Hepatology, Department of Medicine, National University Hospital, Singapore

Correspondence to

Cheng Han Ng;
chenhanng@gmail.com

Overview of Nomenclature in Artificial Intelligence

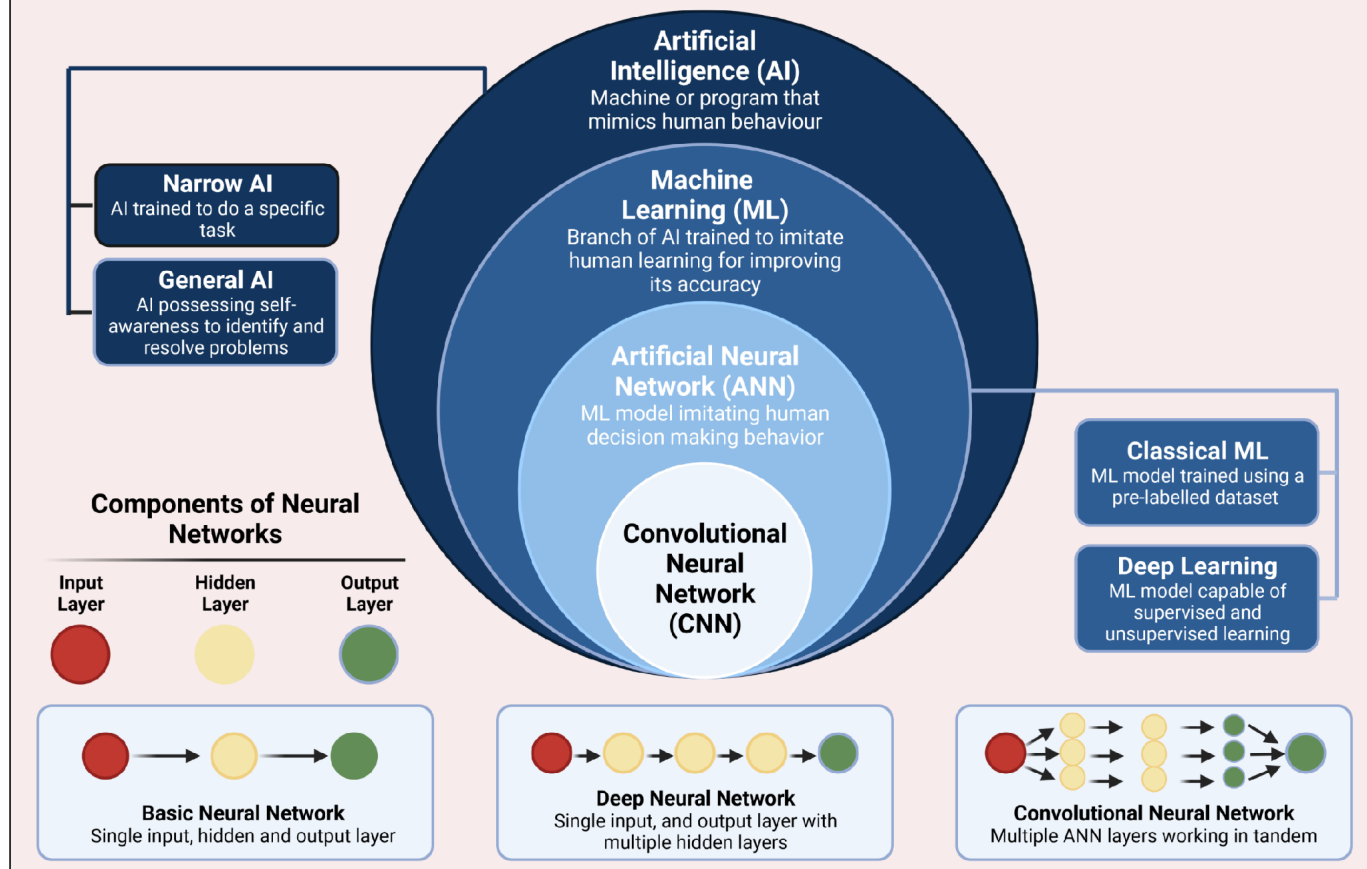


Figure 1 Overview of nomenclature associated with artificial intelligence. Definitions are based on the International Business Machines (IBM) corporation descriptions. The different components of a neural network have been colour-coded; red circles represent the input layer, yellow circles represent the hidden layers and green circles represent the output layer.

within the field of AI. First, AI refers to any science or engineering involved in making intelligent machines, especially intelligent computer programmes, for problem solving.²¹ AI covers a broad range of concepts that aim to allow computers to understand human intelligence but is not confined by what is biologically observable. AI models may be generally separated into two concepts; narrow AI which are trained to support a specific task and general AI which would possess the self-awareness to identify, learn and solve problems with minimal intervention²² (figure 1). Currently, general AI models are purely theoretical and no true general AI models have been created.²³ As such, this paper will focus on narrow AI which has seen implementation clinically. DL and machine learning (ML) are subfields of AI and refer to a system that seeks to identify meaningful relationships and patterns from observed data.^{24 25} Technically, DL models differ from ML models through its method of pattern recognition. ML systems refer to any system that extracts information from one or more supplied data sets and applies it to achieve a specified outcome. DL models are a subset of ML which uses multiple layers of non-linear processing units for feature extraction and classification,

creating a hierarchical system capable of incorporating large amounts of data for analysis.²⁶ Additionally, classical ML models rely on supervised learning, in which human operators manually delineate the pertinent features required to classify the image, while DL models can be trained using both labelled and unlabelled data sets.²⁷ Lastly, artificial neural networks (ANNs) are an application of ML which aim to imitate human pattern recognition behaviour by modelling the way biological neurons signal one another.^{28 29} This involves three general layers: a node layer, one or more hidden layers and an output layer. The node layer receives, and weighs data provided to produce a quantifiable signal. If the produced signal exceeds a specified threshold, the node is 'lit' and activates downstream layers, mimicking the way neurons are fired through sufficient stimuli. Through that, ANNs can perform complex analytics and adapt to unique tasks based on the activation of different layers. Convolutional neural networks (CNNs) are an advancement of ANNs, consisting of multiple layers of ANNs which surpass the classical method in performance, pattern recognition and object recognition.^{30 31} It is important to note that while these terminologies describe unique concepts

within the field of AI, significant overlaps exist between them. For example, while DL is considered a subset of ML, its ability for unsupervised learning can be applied to ANNs and thus could also be considered a subset of ANNs. As such, accurate nomenclature is crucial when discussing the topic.

ROLE OF AI IN DIAGNOSIS OF HCC

The advantages of AI in diagnosing HCC are widespread but will be summarised here in three roles: the reduction of diagnostic variability, the reallocation of healthcare resources and the optimisation of data analysis. First, reduction of diagnostic variability. The diagnosis of HCC relies on an interplay of radiological, histological and cytological parameters.^{32–33} Of the three, radiological identification of HCC remains the preferred method for its non-invasive nature but is the most prone to variability due to factors such as experience of the radiologist and interpreting physician, patient factors and workflow variability.^{34–35} In a study by Covert *et al* evaluating the effect of interoperator variability on MRI-based manual segmentation of 140 HCC lesions, interoperator variability was found to be relatively insignificant (reliability coefficient for response evaluation criteria in solid tumours (RECIST) diameter=0.966) but was estimated to affect the perception of tumour-control probability by up to 25%.³⁶ Theoretically, this could indicate a difference in treatment regimen and outcome between clinicians. As such, the benefit of AI is twofold; its empirical analysis enables objective interpretation of an image while its non-biological nature helps ensure consistent analysis regardless of time or patient load. The second role of AI in imaging would be the reallocation of healthcare resources. The non-biological nature of AI ensures consistent, empirical analysis of patient data while consistent research has shown comparable, if not superior, diagnostic capabilities of AI to physicians in a variety of diseases.³⁷ This holds true in the diagnosis of HCC where, in a preliminary study evaluating the use of a CNN for the interpretation of 228 ultrasound videos, the overall detection rate of the AI (89.8%; 95% CI 84.5% to 95.0%) was significantly higher compared with non-radiologist physicians (29.1%; 95% CI 21.2% to 37.0%, $p<0.001$) and radiologists (70.9%; 95% CI 63.0% to 78.8%, $p<0.001$).³⁸ As such, the implementation of AI systems could help ease the workload of physicians by incorporating a constantly available tool to assist the diagnosis of HCC. Lastly, the optimisation of data analysis. The integration of information systems in healthcare has provided clinicians with an assortment of patient data at their disposal, enabling more reliable and informed diagnoses. However, this increase in volume also necessitates optimal integration to ensure maximum benefit. An example within HCC is a model developed by Kim *et al* which integrated patient history (such as hepatitis B/C and cirrhosis), patient characteristics (such as age and sex), antiviral treatments, imaging

data and laboratory parameters to generate a risk score for developing HCC in patients with chronic hepatitis B.³⁹

PRINCIPLES OF RADIOMICS

Radiomics is the emerging field involving the extraction of quantitative features and subsequent analysis of radiological images using AI models.⁴⁰ The process of training a radiomics model has been extensively described in literature^{41–45} and can be summarised into five steps: image requisition and preprocessing, segmentation, feature extraction, model training and model validation.⁴⁴ Image requisition involves selecting imaging data and identifying potential features causing variability when training the AI model. Crucially, developing a protocol for image requisition requires a balance of standardisation (to reduce noise and confounding) and variability (to ensure generalisability of the model in a clinical setting).⁴⁶ Imaging data are then segmented into regions of interests delineating the tumour and its surrounding areas before being fed into the selected AI model for feature extraction. The models are then refined for accuracy, complexity and efficiency before being validated through a validation data set (usually derived from a subset of the original data). Lastly, the model is tested on a 'blind' training data set to determine its performance. Current radiomics models in HCC have been developed for ultrasound, CTs and MRIs. The following sections will explore some of the notable developments of AI models for each imaging technique.

APPLICATION IN ULTRASOUND

Liver ultrasonography has been a standard radiological modality for HCC surveillance due to its non-invasiveness and absent risk of radiation or contrast exposure.⁴⁷ However, the utility of ultrasound in detecting HCC is being increasingly debated due to its suboptimal performance, particularly in early HCC, and its operator-dependent variability. Ultrasound detection of HCC remains relatively reliable with a meta-analysis of 32 studies (comprising 13 367 patients of varying HCC severity) concluding with an overall sensitivity of 84% (95% CI 76% to 92%). However, the study also evaluated the sensitivity of detecting early HCC at 47% (95% CI 33% to 61%).⁴⁸ This disparity is likely due to most patients with HCC presenting with confounding comorbidities such as cirrhosis which produce a coarse pattern on ultrasound, precluding identification of small HCC nodules.⁴⁹ Additional factors such as obesity, fatty-liver disease and fibrosis may similarly impair the quality of liver ultrasound, thereby obfuscating the presence of smaller HCC nodules in early HCC.⁵⁰ While newer applications, for example, contrast ultrasounds and diagnostic guidelines have helped limit the diagnostic variability of ultrasounds in HCC, recent interest has been on the implementation of AI models in interpreting ultrasound images. A CNN model developed by Tiyyarattanachai *et al*, highlighted the diagnostic

potential of AI systems in liver ultrasounds.³⁸ There, a pretrained CNN model (RetinaNet⁵¹) was fed 25 557 images of various common ultrasound findings (eg, HCC, liver cysts and haemangiomas) followed by refinement using 228 ultrasound videos with difficult frames to create a model specialising in the differentiation of ultrasound findings. The model was then tested against 175 videos containing 127 lesions and its performance was compared with physicians (both non-radiologists and radiologists) to assess its potential. Overall, the AI system achieved an overall detection rate of 89.8% (114/127 lesions; 95% CI 84.5% to 95.0%) and a 100% detection rate for HCC (23/23 lesions; 95% CI 85.2% to 100%). Comparatively, overall detection rates for non-radiologist physicians and radiologists were 29.1% and 70.9%, respectively (both $p<0.001$) while detection rates for HCC were 39.1% ($p<0.001$) and 69.6% ($p=0.016$), respectively. This study demonstrated the feasibility of AI systems in detecting HCC through ultrasound with detection rates significantly higher than both non-radiologist and radiologist clinicians. Current clinical guidelines suggests early surveillance for HCC using ultrasound and serum AFP.⁵² However AFP remains negative in nearly two-thirds of patients at any stage of HCC⁵³ and screening outcomes using ultrasound alone remain suboptimal.⁴⁸ A DL model detailed by Zhang *et al* highlighted the potential of AI as a screening tool for AFP-negative HCCs.⁵⁴ The CNN model (Xception⁵⁵) was trained using a total of 305 images of HCC and focal nodular hyperplasia taken using B-mode ultrasound and model testing was done using 102 B-mode ultrasound images. HCC staging, lesion size, echogenicity and liver function were heterogeneous in both training and testing data sets. Diagnostic performance was then compared with four other available CNN models (MobileNet,⁵⁶ Resnet50,⁵⁷ Densenet121,⁵⁸ InceptionV3).⁵⁹ In total, the model displayed promising diagnostic capabilities with an overall area under the curve (AUC) of 93.68% (95% CI 88.6% to 98.8%), sensitivity of 96.08%, specificity of 76.92% and accuracy of 86.41%, outperforming other CNN models and providing an optimistic non-invasive tool for HCC surveillance. Table 1 summarises the key studies evaluating the use of AI in the interpretation of liver ultrasound.

APPLICATION IN CT

CT is a primary surveillance method for HCC, exhibiting utility in particular for detecting dysplastic lesions and early stage HCC.^{60 61} Current guidelines recommend a multiphasic CT with extracellular contrast agents as a first-line option for the diagnosis and staging of HCC.⁶² A meta-analysis by Lee *et al* evaluating the diagnostic capabilities of CT imaging for detecting HCC concluded with an overall sensitivity of 72% (95% CI 62% to 80%).⁶³ Additionally, the study reported a sensitivity of 74% and specificity of 81%

when CT imaging was used as the initial diagnostic tool for focal liver lesions detected during surveillance. These results highlight that while acceptable, there is significant room for optimisation in CTs as a diagnostic tool for HCC and the incorporation of AI has shown promise in that regard. A DL model outlined by Wang *et al* reported optimistic performance in identifying patients with HCC from CT data with an AUC of 0.887 (95% CI 0.855 to 0.919) for an internal data set and 0.883 (95% CI 0.855 to 0.911) for an external data set.⁶⁴ In the paper, 647 individuals with HCC of various stage of disease and 6865 non-HCC individuals were analysed using plain and contrast-enhanced CT. Obtained images were then fed into two CNN models (NoduleNet and HCCNet) for training, and performance testing was done on an internal and external data set. A subset of the test data set was also reviewed by three radiologists to obtain a comparison of performance. For the internal data set, the AI model achieved a sensitivity of 78.4% (95% CI 72.4% to 83.7%), specificity of 84.4% (95% CI 78.0% to 89.6%) and an overall accuracy of 81.0% (95% CI 76.8% to 84.8%). For the external data set, the model achieved a sensitivity of 89.4% (95% CI 85.0% to 92.8%), specificity of 74.0% (95% CI 68.5% to 78.9%) and an accuracy of 81.3% (95% CI 77.8% to 84.5%). Comparatively, the AI model performed similarly to the radiologists in terms of accuracy (0.853 vs 0.818, $p=0.107$), sensitivity (0.815 vs 0.753, $p=0.064$) and specificity (0.902 vs 0.903, $p=0.981$) for the internal data set, with similar findings of the external data set. However, the true utility of the model was displayed when it was used to assist the radiologists in improving their diagnostic performance as seen by a significant improvement in diagnostic accuracy for the internal data set (0.873 vs 0.818, $p=0.026$) and external data set (0.854 vs 0.793, $p=0.017$). This study indicated the potential for AI systems to standardise the diagnosis and stratification of HCC when using CT imaging. Table 2 summarises the key studies evaluating the use of AI in the interpretation of liver CTs.

APPLICATION IN MRI

Liver MRI is another first-line imaging tool for the diagnosis and classification of HCC.⁶² Advantages of MRI include its high-quality imaging of the entire liver, contrasting enabling detection and classification of HCC and lack of ionising radiation allowing safer routine use.^{65–67} MRIs have also shown utility in detecting small HCC nodules more consistently than CTs. In a meta-analysis of 34 articles, MRIs had an overall sensitivity of 88% (95% CI 83% to 92%) and specificity of 94% (95% CI 85% to 98%). However, MRIs also had a significantly lower sensitivity in detecting lesions less than 2 cm with a sensitivity of 62% for lesions <2 cm and 95% for lesions >2 cm ($p<0.02$). These results highlight that while MRIs generally have higher detection rates, more optimisations

Table 1 Overview of studies involving AI in HCC ultrasound as of January 2023

Author	Model type	Data set	Model description	Results
Shiraishi <i>et al</i> ⁸⁶	ANN	This study used 103 FLLs obtained from contrast ultrasound of 97 patients, consisting of 26 metastases, 16 haemangiomas and 61 HCCs.	This computerised scheme consists of six major steps; (1) Extraction of a series of microflow images from the original clip; (2) Automated segmentation of a kidney, vessel-like patterns, the adjacent liver parenchyma region and a central/peripheral region of an FLL; (3) Estimation of time-intensity curves for an FLL by use of a series of temporally averaged microflow images for determination of temporal features; (4) Extraction of morphological and grey-level image features for an FLL; (5) Stepwise feature selections for distinguishing two groups in six different decisions; and (6) Application of a cascade of 6 independent ANNs	The classification accuracies for the 103 FLLs were 88.5% for metastasis, 93.8% for haemangioma and 86.9% for all HCCs.
Streba <i>et al</i> ⁸⁷	CNN	112 patients with HCC (n=41), hypervascular (n=20) and hypovascular (n=12) liver metastases, hepatic haemangiomas (n=16) or focal fatty changes (n=23) who underwent contrast-enhanced ultrasonography were prospectively included. Full-length movies of all contrast uptake phases were used and postprocessed offline by selecting two areas of interest (one for the tumour and one for the healthy surrounding parenchyma) and consecutive analysis.	The difference in maximum intensities, the time to reaching them and the aspect of the late/portal phase, as quantified by the neural network and a ratio between median intensities of the central and peripheral areas were analysed by a feed forward back propagation multilayer neural network which was trained to classify data into five distinct classes, corresponding to each type of liver lesion.	The neural network had 94.45% training accuracy (95% CI 89.31% to 97.21%) and 87.12% testing accuracy (95% CI 86.83% to 93.17%). The automatic classification process registered 93.2% sensitivity, 89.7% specificity, 94.42% positive predictive value and 87.57% negative predictive value.
Huang <i>et al</i> ⁸⁸	ML	Two data sets were collected. Data set 1 contains 155 cases with FNH and 49 cases with atypical HCC. Data set 2 was an imbalanced data set including 102 cases with FNH while only 36 cases with atypical HCC. Regions of interests were identified by a radiologist and randomly divided into a training and testing data set. Data set 1 includes the training data (169 FNH nodules and 167 atypical HCC nodules) and the testing data (15 FNH nodules and 15 atypical HCC nodules). Data set 2 includes the training data (186 FNH nodules and 201 atypical HCC nodules) and the testing data (15 FNH nodules and 15 atypical HCC nodules).	The model automatically calculates three types of semantic features with equally down-sampled frames based on contrast-enhanced ultrasound. A support vector machine classifier is trained to make the final diagnosis.	The average accuracy was 94.40%, recall rate 94.76%, F1-score value 94.62%, specificity 93.62% and sensitivity 94.76%, indicating good merit for automatically diagnosing atypical HCC cases.
Tiyarattanachai <i>et al</i> ⁸⁹	CNN	Images from upper abdominal ultrasounds were retrospectively retrieved from three different hospitals. A total of 40 397 images with 20 432 FLLs were included in the training set, while 6191 images with 845 FLLs and 18 922 images with 1195 FLLs were included in the internal test set and external validation set, respectively.	A CNN architecture called RetinaNet was used. The training was done in two main steps. First, the backbone ResNet50 was trained on a publicly available image data set. Subsequently, the whole CNN was fine-tuned on ultrasound images in the training set. The CNN was trained for 500 000 iterations on ultrasound images. To enable the CNN to recognise diverse configurations of images and to maximise the number of training images, image augmentation was performed by horizontal translation, vertical translation, rotation, scaling, horizontal flipping, motion blur, contrast, brightness, hue and saturation adjustment at each iteration.	For diagnosis of HCC, the CNN yielded sensitivity, specificity and negative predictive value of 73.6% (95% CI 64.3% to 82.8%), 97.8% (95% CI 96.7% to 98.9%) and 96.5% (95% CI 95.0% to 97.9%) on the internal test set, and 81.5% (95% CI 74.2% to 88.8%), 94.4% (95% CI 92.8% to 96.0%) and 97.4% (95% CI 96.2% to 98.5%) on the external validation set, respectively.

Continued

Table 1 Continued

Author	Model type	Data set	Model description	Results
Li <i>et al</i> ²⁰	ML	A total of 226 focal liver lesions, including 107 atypical HCC and 119 FNH lesions, examined by contrast-enhanced ultrasound were reviewed retrospectively. 3132 features were extracted from the images of the baseline, arterial and portal phases. The entire data set was randomly divided into a training data set (comprising 80% of subjects) and a validation data set (comprising the remaining 20% of subjects).	The model was built using the training data set and then evaluated using the validation data set repeatedly. Subsequently, the model with the best classification performance was selected as the best model. The predictive model was constructed using the support vector machine method trained with the following groups: ultrasonics features, radiologist's score, and combination of ultrasonics features and radiologist's score.	The model presented good performance in differentiating FNH and atypical HCC with an AUC of 0.86 (95% CI 0.80 to 0.89), a sensitivity of 76.6% (95% CI 67.5% to 84.3%) and a specificity of 80.5% (95% CI 70.6% to 85.9%). The model trained with a combination of ultrasonics features and the radiologist's score achieved a significantly higher AUC (0.93, 95% CI 0.89 to 0.96) than that trained with the radiologist's score (AUC: 0.84, 95% CI 0.79 to 0.89, $p < 0.001$).
Virmani <i>et al</i> ⁶⁰	ANN	A total of 108 B-mode liver ultrasound images were collected (18 normal, 12 cyst, 15 haemangiomas, 28 HCCs and 28 metastatic lesions). Regions of interest identified by both radiologist and model were classified as both internal regions of interest (IROIs) and surrounding regions of interest (SROIs). Regions of interests were then bifurcated into a training data set (SROIs $n=59$; IROIs $n=200$) and testing data set (SROIs $n=52$; IROIs $n=180$).	The first step of the classification module consists of a five class principal component analysis neural network (PCA-NN) based primary classifier which yields probability outputs for five liver image classes. The second step of classification module consists of 10 binary PCA-NN based secondary classifiers. The probability outputs of five class PCA-NN based primary classifier is used to determine the first two most probable classes for a test instance, based on which it is directed to the corresponding binary PCA-NN based secondary classifier for crisp classification between two classes.	The first step classification model alone reported an overall accuracy of 87.7% while incorporating both models increased the overall accuracy to 95%.
Mitrea <i>et al</i> ⁶¹	CNN	The contrast-enhanced ultrasound images corresponding to 48 patients affected by HCC were considered. The pixel data from the B-mode ultrasound images and contrast-enhanced ultrasound was fused, combining the two types of information provided by each imaging modality, the first being tissue reflections by ultrasound and the second being the vessel structure under contrast. Data were split into 60%, 15% and 25% for the training, validation and testing data set, respectively. Concerning the traditional classification techniques, 75% of the data were included in the training set and 25% of the data constituted the test set.	The performance of various classifiers was evaluated against a novel method of classification using both B-mode ultrasound and contrast-enhanced ultrasound through feature-level fusion. Multiple feature-level fusion classifiers were evaluated to identify the best model.	After evaluating six different feature-level fusion models, the best performances were obtained: accuracy of 93.2%, AUC of 93.2%, sensitivity of 93.6%, specificity of 95%.
Zhang <i>et al</i> ⁶⁴	DL	A total of 407 cases were enrolled in the study, comprising 209 HCC and 198 FNH cases. B-mode ultrasound images of liver lesions were obtained. The model cohort and test cohort were set at a ratio of 3:1, in which the test cohort was composed of AFP-negative HBV-infected cases.	Five deep learning models (Xception, MobileNet, Resnet50, DenseNet121 and InceptionV3) were constructed as comparative baselines.	The AUC of the model, Xception, achieved 93.68% in the test cohort, superior to other baselines (89.06%, 85.67%, 83.94% and 78.13%, respectively, for MobileNet, Resnet50, DenseNet121 and InceptionV3).

Search terms included 'Hepatocellular Carcinoma' AND 'Artificial Intelligence' OR 'AI' OR 'Machine Learning' OR 'Ultrasound' on the Pubmed library. Search results were manually filtered to only include studies discussing the diagnosis of HCC.
AFP, alpha-feto protein; AI, artificial intelligence; ANN, artificial neural network; AUC, area under the curve; CNN, convolutional neural network; DL, deep learning; FLL, focal nodular hyperplasia; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; ML, machine learning.

Table 2 Overview of studies on the application of AI in HCC CTs as of January 2023

Author	Model type	Data set	Model description	Results
Weston <i>et al</i> ⁹²	CNN	Data from 2715 examinations were collected from 1429 patients who had been imaged for treatment of pancreatic cancer, renal cell carcinoma, transitional cell carcinoma or gastrointestinal cancer. A secondary data set of 2369 CT examinations from 1083 patients who were being treated for HCC was also added to assess the generalisability of the model. The data set was randomly partitioned into training (90%; 2430 of 2700) and test (10%; 270 of 2700) sets. The training set was further partitioned into training (90%; 2187 of 2430) and validation (10%; 243 of 2430) sets.	A convolutional deep neural network referred to as a U-Net was adapted to perform segmentation. The U-Net architecture consists of cascading layers of learnable convolutional filters.	The model achieved Dice Scores (mean±SD) of 0.98±0.03 in the test set, and 0.94±0.05 in the secondary HCC data set.
Nayak <i>et al</i> ⁹³	DL	A clinical data set of 40 patients identified using CE-CT was retrospectively analysed in this study. The data set had three groups of subjects: healthy (n=14/40), cirrhosis (n=12/40) and cirrhosis with HCC (n=14/40). Five phases of CT images, no-contrast, arterial phase, portal-venous phase, hepatic-venous phase and delayed phase, were acquired.	A novel method for the automatic 3D segmentation of liver using modified region-growing segmentation technique was developed and compared with the state-of-the-art deep learning based technique. Using support vector machine, multiphase analysis of CT images was performed to extract 24 temporal features for detecting cirrhosis and HCC liver.	The proposed method produced improved 3D segmentation with a Dice coefficient of 90% for healthy liver, 86% for cirrhosis and 81% for HCC subjects compared with the deep learning algorithm (healthy: 82%; cirrhosis: 78%; HCC: 70%).
Yamada <i>et al</i> ⁹⁴	CNN	215 consecutive patients with histologically proven primary liver cancers, including 6 early, 58 well-differentiated, 109 moderately differentiated, 29 poorly differentiated HCCs and 13 non-HCC malignant lesions containing cholangiocellular components were retrospectively evaluated. The input images with specific degrees of misalignment were divided into training (70%) and test (30%) sets.	Transfer learning was done using various pretrained CNNs and preoperative three-phasic dynamic contrast-enhanced CT (CE-CT). Three-phasic dynamic CE-CT images were manually registered to correct respiratory motion. The registered dynamic CE-CT images were then assigned to the three colour channels of an input image for transfer learning: precontrast, early phase and delayed phase images for the blue, red and green channels, respectively.	The highest mean diagnostic accuracy for CNNs after transfer learning with pixel shifts, rotations and skew misalignments were 44.1%, 44.2% and 43.7%, respectively.
Shi <i>et al</i> ⁹⁵	CNN	342 patients scanned with a four-phase CT protocol (precontrast, arterial, portal-venous and delayed phases) were retrospectively enrolled. 449 FLLs were categorised into HCC and non-HCC groups based on the best available reference standard. Models were trained on 80% of lesions and evaluated in the other 20%.	Three convolutional dense networks with the input of four-phase CT images (model A), three-phase images without portal-venous phase (model B) and three-phase images without precontrast phase (model C) were trained.	The diagnostic accuracy in differentiating HCC from other FLLs on test sets was 83.3% for model A, 81.1% for model B and 85.6% for model C, and the AUCs were 0.925, 0.862 and 0.920, respectively. The AUCs of models A and C did not differ significantly (p=0.765), but the AUCs of models A and B (p=0.038) and of models B and C (p=0.028) did.
Mokrane <i>et al</i> ⁹⁶	ML	178 patients from 27 different institutions were included in this study. Patients were randomly assigned to a discovery cohort (142 patients) and a validation cohort (36 patients). Each liver nodule was segmented on each phase of triphasic CT scans, and 13920 quantitative imaging features were extracted.	Using machine learning techniques, the signature was trained and calibrated (discovery cohort), and validated (validation cohort) to classify liver nodules as HCC versus non-HCC. Effects of segmentation and contrast enhancement quality were also evaluated.	The signature reached an AUC of 0.70 (95% CI 0.61 to 0.80) and 0.66 (95% CI 0.64 to 0.84) in discovery and validation cohorts, respectively.

Continued

Table 2 Continued

Author	Model type	Data set	Model description	Results
Wang <i>et al</i> ⁶⁴	DL	The liver CT imaging data of 9185 patients were retrospectively collected. The deep learning AI system was trained on CT images from 7512 patients. Its performance was validated on one internal test set (n=385) and one external test set (n=556).	Two deep learning AI models NoduleNet and HCCNet were developed in this study. NoduleNet acted as an assistant to HCCNet and the results of its analysis were directly integrated into HCCNet. Both NoduleNet and HCCNet are two deep residual convolutional networks of 34 layers.	The system achieved an AUC in identifying HCC of 0.89 (95% CI 0.86 to 0.92) on the internal test set and 0.88 (95% CI 0.86 to 0.91) on the external test set. For the internal test set, accuracy was 81.0%, sensitivity was 78.4%, specificity was 84.4% and F1 was 0.824. For the external test set, accuracy was 81.3%, sensitivity was 89.4%, specificity was 74.0% and F1 metric was 0.819.
Liu <i>et al</i> ¹⁹	ML	This retrospective study included 85 patients aged 32–86 years with 86 histopathology-proven liver cancers: 24 combined hepatocholangiocarcinoma (cHCC-CC), 24 CC (cholangiocarcinoma) and 38 HCC who had MRI and CT. Preoperative CTs were available in 35/37 HCC, 15/24 cHCC-CC and 23/24 CC.	Support vector machine classifier was utilised to evaluate CT radiomics features for the prediction of cHCC-CC versus non-cHCC-CC and HCC versus non-HCC.	Precontrast and portal-phase CT showed excellent performance for the differentiation of HCC from non-HCC with an AUC of 0.81 (SD 0.06) and 0.71 (SD 0.15) for CT phases, respectively. The misdiagnosis of cHCC-CC as HCC or CC using radiologists read was 69% for CT.
Kim <i>et al</i> ⁹⁷	DL	A total of 1350 multiphase CT scans of 1280 hepatic malignancies (1202 HCCs and 78 non-HCCs) in 1320 patients at high risk for HCC were retrospectively analysed. The scans were categorised randomly into the training (568 scans), tuning (193 scans), and test (589 scans) sets.	Multiphase CT information was subjected to multichannel integration, and livers were automatically segmented before model development. A deep learning based model capable of detecting malignancies was developed using a mask region-based CNN.	This model exhibited a sensitivity of 84.8% with 4.80 false-positives per CT scan on the test set. The most frequent potential causes of false-negatives and false-positives were determined to be atypical enhancement patterns for HCC (71.7%) and registration/segmentation errors (42.7%), respectively.
Khan <i>et al</i> ⁹⁸	ML	A data set of 179 clinical cases which consisted of 98 benign and 81 malignant liver tumours was used.	A novel Fuzzy Linguistic Constant is designed for image enhancement. To classify the enhanced liver image as cancerous or non-cancerous, fuzzy membership function is applied. The extracted features are assessed for malignancy and benignancy using the structural similarity index.	Classification accuracy of 98.3% is achieved by Support Vector Machine. The proposed method automatically segmented the tumour with an improved detection rate of 78% and a precision value of 0.6.
Gao <i>et al</i> ¹⁸	DL	723 patients from two centres were pathologically diagnosed with HCC, ICC or metastatic liver cancer. The training and test sets consisted of 499 and 113 patients from centre 1, respectively. The external test set consisted of 111 patients from centre 2.	Deep learning model which takes advantage of deep CNN and gated recurrent neural network to effectively extract and integrate the diagnosis-related radiological and clinical features of patients.	The model achieved an accuracy of 86.2% and an AUC of 0.893 for classifying HCC and ICC on the test set. The model achieved an accuracy of 72.6% on the test set, comparable with the diagnostic level of doctors' consensus (70.8%).
Xu <i>et al</i> ⁹⁹	ML	Patients with pathological results of liver cancer underwent non-contrast CT between August 2018 and November 2019. The whole cohort was partitioned into two parts, 122 patients for training and 89 patients for validation.	Radiomic features were extracted from CT scans. Feature selection was operated with the least absolute shrinkage and operator logistic method. The support vector machine was selected to build a model.	In the training set, the AUC of the evaluation of the radiomics was 0.855 higher than for radiologists at 0.689. In the valuation cohorts, the AUC of the evaluation was 0.847 and the validation was 0.659.

Continued

Table 2 Continued

Author	Model type	Data set	Model description	Results
Rocha <i>et al</i> ¹⁰⁰	CNN	The study population is composed of 396 CT scans from healthy liver donors (20%) and patients with cirrhosis (80%) which are composed of four volumes each, summing in 178.633 slices.	An annotation platform and CNN was trained to automatically identify the CT scan phases. The algorithm was improved with hyperparameter tuning and evaluated with cross-validation methods.	Comparing its predictions with the radiologist's annotation, it achieved an accuracy of 94.6%, 98% and 100% in the testing data set for the slice, volume and exam evaluation, respectively.
Duc <i>et al</i> ¹⁰¹	CNN	The medical records and CT imaging studies of 110 patients with 115 HCC lesions were included in this retrospective study. The data set was randomly divided into a training set (92 patients), and a test set (18 patients).	The proposed CNN segmentation method was based on the U-Net architecture and trained using the domain adaptation technique.	The sensitivity for HCC identification of the model was 100%. The median Dice Score for HCC segmenting between radiologists and the CNN model was 0.81 for the test set.
Search terms included 'Hepatocellular Carcinoma' AND 'Artificial Intelligence' OR 'AI' OR 'Machine Learning' OR 'Neural Networks' AND 'CT' OR 'CT' on the Pubmed library. Search results were manually filtered to only include studies discussing the diagnosis of HCC. AI, artificial intelligence; AUC, area under the curve; CE-CT, contrast-enhanced CT; CNN, convolutional neural network; DL, deep learning; FLL, focal liver lesions; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; ML, machine learning.				

may be possible and in particular for detecting early HCC. Current AI models have shown promise in differentiating common liver lesions such as HCC, haemangioma and metastatic tumours through feature extraction of MRIs. In a study by Oyama *et al*, non-contrast-enhanced fat-suppressed 3D T1-weight gradient echo MRIs of 150 hepatic tumours (50 HCC, 50 haemangiomas and 50 metastatic tumours) were fed into an inhouse ML model. There, the images were ranked based on their likely identity (ie, HCC, haemangioma or metastatic tumour). Overall, the ML model was able to differentiate HCCs and metastatic tumours with an accuracy of 92% (92/100 lesions), sensitivity of 100% (50/50 lesions), specificity of 84% (42/50 lesions) and an AUC of 95%. Similarly, the model differentiated HCCs with haemangiomas with an accuracy of 90% (90/100 lesions), sensitivity of 96% (48/50 lesions), specificity of 84% (42/50) and AUC of 95%. Despite the limited analysis, the results support the potential of AI systems in the interpretation of MRIs.⁶⁸ Another study by Oestmann *et al* evaluating the use of CNNs in identifying HCCs on contrast-enhanced MRIs identified key factors affecting the diagnostic accuracy of ML models.⁶⁹ In the study, 118 patients (73 HCC and 45 non-HCC) were evaluated using contrast-enhanced MRIs and the images were analysed in a CNN. Atypical lesions, defined as lesions not displaying typical MRI appearances based on the Liver Imaging Reporting and Data System criteria (ie, arterial hyperenhancement, washout and enhancing rim/pseudocapsule),⁷⁰ were included to determine the correlation between lesion grading and CNN performance. The study concluded with the CNN demonstrating an overall accuracy of 87.3% and sensitivities/specificities for HCC and non-HCC lesions of 92.7%/82.0% and 82.0%/92.7%, respectively. CNN performance was also directly correlated with the lesion grading system, becoming less accurate the more atypical imaging features the lesion showed. This study highlighted the potential pitfall of an ML system as their strict

algorithm lacked the flexibility to consistently identify atypical lesions. Table 3 summarises the key studies evaluating the use of AI in the interpretation of liver MRIs.

OTHER APPLICATIONS OF AI IN HCC DIAGNOSIS

Apart from directly aiding in the interpretation of radiological images to diagnose HCC, several other applications have emerged to assist in the surveillance and classification of HCC. One such example would be the development of integrated ML models for risk-score prediction.³⁹ A recent model developed by Kim *et al* aimed to predict the risk of HCC in Korean and Caucasian patients with hepatitis B. The ML model was developed which incorporated ultrasound images and 10 baseline parameters to assess the risk of HCC in patients with hepatitis B virus (HBV). These parameters were the presence of cirrhosis, age, platelet count, antiviral agent used, sex, serum alanine aminotransferase levels, baseline serum levels of HBV DNA, serum levels of albumin and bilirubin, and hepatitis B e-antigen (HBeAg) status, which are listed in the order of importance. The results were prospectively validated against a Korean (n=5817) and Caucasian (n=1640) cohort and assessed with a Concordance Index (c-index).⁷¹ The model achieved a c-index of 0.79 (95% CI 0.78 to 0.80) and 0.81 (95% CI 0.79 to 0.83), respectively, in the Korean and Caucasian cohort and highlights the potential of AI models in the integration of patient parameters and for the risk stratification of HCC.

CURRENT CHALLENGES AND FUTURE DIRECTIONS

The recent boom of AI models into mainstream popularity has placed increased focus on their applications in healthcare. However, despite constant improvements in hardware and software capabilities, several challenges remain unaddressed. First, the standardisation of study design. Due to the relative infancy of ML and

Table 3 Summary of studies on the application of AI in HCC MRIs as of January 2023

Author	Model type	Data set	Model description	Results
Kim <i>et al</i> ¹⁰²	CNN	The preoperative gadoxetic acid-enhanced liver MRIs of 549 patients from 2010 to 2014 who were confirmed to have HCC after surgical resection were reviewed. An external data set of 54 patients with HCC was also used. There was a total of 92 645 hepatobiliary phase MRIs. They were categorised 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, validation and test data set, respectively.	The model used a fine-tuned CNN based on the best combination of the chosen hyperparameters, including batch normalisation, dropout, activation function, kernel size and optimiser.	The model achieved 87% sensitivity and 93% specificity for the detection of HCCs with an external validation data set (54 patients).
Zheng <i>et al</i> ¹⁰³	DL	A total of 290 consecutive patients with pathologically confirmed HCCs were recruited for liver gadolinium-DTPA-enhanced MRI. The retrospective study included 5376 image sets from 56 patients with cirrhosis in the training-validation cohort to build and validate the model. An external test cohort including 6144 image sets from 64 patients with cirrhosis was applied to further verify the generalisation ability of the model.	The proposed pattern matching DL model consisted of three main steps: 3D co-registration and liver segmentation, screening of suspicious lesions on diffusion-weighted images based on pattern matching algorithm, and identification/segmentation of small HCC lesions on dynamic contrast-enhanced images with CNN	The pattern matching DL model achieved a sensitivity of 89.74% and a positive predictive value of 85.00% in the external test cohort for per-lesion analysis. The model outperformed Liver Imaging Reporting and Data System in sensitivity (probable HCCs: LR-5 or LR-4, $p=0.18$; definite HCCs: LR-5, $p<0.001$), with a similar high specificity for per-patient analysis.
Oestmann <i>et al</i> ⁶⁹	CNN	This study included 118 patients with HCC ($n=73$) and non-HCC lesions ($n=45$). A total of 93 HCC lesions and 57 non-HCC lesions were analysed. The non-HCC group consisted of 19 (33%) ICCs, 16 (28%) haemangiomas, 15 (26%) cysts, 2 (4%) regenerative nodules, 2 (4%) dysplastic nodules, 2 (4%) FNHs and 1 (2%) bile duct adenoma.	The CNN was trained on 70 HCC and 70 non-HCC examples, drawn randomly from the augmented data set. As the data set contained lesions with atypical appearances on MRI, a lesion grading system was developed based on the established Liver Imaging Reporting and Data System major imaging criteria using imaging features typical of HCC: arterial hyperenhancement, washout, and enhancing rim/pseudocapsule	The CNN demonstrated an overall accuracy of 87.3%. Sensitivities/specificities for HCC and non-HCC lesions were 92.7%/82.0% and 82.0%/92.7%, respectively. The area under the receiver operating curve was 0.912. Performance of CNNs was correlated with the lesion grading system, becoming less accurate the more atypical imaging features the lesions showed.
Bousabarah <i>et al</i> ¹⁰⁴	CNN	A set of multiphasic MRI series from 174 patients with 231 lesions in total were used. These data were randomly split into a 70% training, a 15% test and a 15% validation set on a patient level.	A single CNN was trained for simultaneous liver and HCC segmentation. The input to the U-Net was a stack of nine adjacent uncropped axial slices across three channels (arterial, portal venous and delayed phase time points). From those, the model learnt and output 2D segmentations of the liver and HCC for the middle slice using the corresponding slice from the volumetric ground truth segmentation.	73% and 75% of HCCs were detected on validation and test sets, respectively.
Stollmayer <i>et al</i> ¹⁰⁵	DL	69 patients with FNHs, HCCs or liver metastases, that were acquired using Primovist (gadoxetate disodium) were retrospectively collected in multiphasic MRI studies. After volume registration, the same three most representative axial slices from all sequences were combined into four-channel images to train the 2D-DenseNet264 network.	The DenseNet264 models from the Pytorch-based open-source Medical Open Network for Artificial Intelligence (MONAI) framework was used. The study used categorical cross-entropy loss to measure the prediction error of the network during training and an Adam optimizer to update model parameters.	The average AUC value of the 2D model (0.98) was slightly higher than that of the 3D model (0.94). Mean positive predictive value (PPV), sensitivity, negative predictive value (NPV), specificity and F1 Scores (0.94, 0.93, 0.97, 0.97 and 0.93) of the 2D model were also superior to metrics of the 3D model (0.84, 0.83, 0.92, 0.92 and 0.83).
Cho <i>et al</i> ¹⁰⁶	CNN	This retrospective study reviewed 324 consecutive cases (298 patients) who underwent surgery for suspected HCC. 198 nodules from 183 patients were included. Regions of interests were drawn on the whole HCC volume on hepatobiliary phase (HBP), T1-weighted (T1WI), T2-weighted (T2WI) and portal venous phase (PVP) images.	The computer-aided detection system was developed from the HBP images using customised-nnUNet and postprocessed for false-positive reduction. Internal and external validation was performed with HBP, T1WI, T2WI and PVP images.	To evaluate the performance of computer-aided detection for HCC detection on multisequence MRI in the internal data set, the cut-off threshold (0.5) was determined using recall and average false-positive rate in the algorithm. At this cut-off threshold (0.5), the recalls of computer-aided detection in the internal test set were 87.0%, 73.3%, 13.3% and 66.7% in HBP, T1WI, T2WI and PVP, respectively

Continued

Table 3 Continued

Author	Model type	Data set	Model description	Results
Liu <i>et al</i> ¹⁹	ML	This retrospective study included 85 patients with 86 histopathology-proven liver cancers: 24 cHCC-CC, 24 CC and 38 HCC who had MRI between 2004 and 2018.	Following tumour segmentation, 1419 radiomics features were extracted using PyRadiomics library and reduced to 20 principle components by principal component analysis. Support vector machine classifier was used to evaluate MRI and CT radiomics features for the prediction of cHCC-CC versus non-cHCC-CC and HCC versus non-HCC.	Contrast-enhanced MRI phases showed excellent performance for the differentiation of HCC from non-HCC (AUC of 0.79 (SD 0.07) to 0.81 (SD 0.13)). The misdiagnosis of cHCC-CC as HCC or CC using radiologists read was 58%.

Search terms included 'Hepatocellular Carcinoma' AND 'Artificial Intelligence' OR 'AI' OR 'Machine Learning' OR 'Neural Networks' AND 'MRI' on the Pubmed library. Search results were manually filtered to only include studies discussing the diagnosis of HCC.
 AI, artificial intelligence; AUC, area under the curve; cHCC-CC, hepatocellular carcinoma; CNN, convolutional neural network; DL, deep learning; DTPA, diethylenetriamine penta-acetate; FNH, focal nodular hyperplasia; HCC, hepatocellular carcinoma; ML, machine learning.

CNN models in healthcare, limited regulations exist governing their application. This holds true for their use in HCC diagnostics, where there is currently scarce data on their efficacy in actual patient settings. As of writing this paper, there is currently no consensus on the feasibility of AI in large-scale HCC radiomics nor standardised approaches on how results should be reported or interpreted. A recent meta-analysis on the quality assessment standards in systematic reviews of AI accuracy have also shown studies to be afflicted by biasness.⁷² These findings included 243 out of 423 (57.5%) studies reporting a high or unclear risk of patient selection bias, 110 studies (26.0%) reporting a risk of bias in index test selection, 121 studies (28.6%) reporting a risk of bias in reference standard and 157 studies (37.1%) reporting a risk of bias in study flow. The study highlights the need for AI-specific quality assessment tools and guidelines to facilitate the safe use of AI tools into clinical practice. Several guidelines have been proposed such as the Standard Protocol Items: Recommendations for Interventional Trials - Artificial Intelligence (SPIRIT-AI) and Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT AI) guidelines which propose a framework for designing and reporting AI-related clinical trials to ensure robustness and accuracy of results.⁷³ Additionally, while several studies exist evaluating the performance of AI systems, the vast majority of these have been retrospective in nature and prospective evidence in the field remains limited.

Another challenge is the quality assessment of data sets. Proper training and validation of ML models are crucial to ensure accuracy which necessitates quality data sets.⁷⁴ Currently, most published studies on the applications of AI in HCC diagnostics have focused on improving the quality of their models through additions and optimisation on the model itself (such as optimising feature selection and increasing CNN layers). However, limited studies have been published focusing on the impact of data set quality on ML training and the ideal features of a good data set.^{75 76} Differences in diagnostic equipment and workflows have also made it difficult to accurately identify an optimal imaging workflow for training AI models. A systematic review by Fatania *et al* provides insight into how the optimal radiological features could

be evaluated.⁷⁷ In the paper, the intensity standardisation techniques of 12 studies were classified and evaluated to assess reliability based on the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) quality-assessment criteria.⁷⁸ Despite the small sample size (n=12) and focus on gliomas, the paper by Fatania *et al* serves as a helpful framework for developing future evaluative studies for AI-related image acquisition methods in HCC. Another interesting approach has been the use of a supplementary AI model to automatically detect and exclude poor-quality data in a training set.⁷⁹ The proposed model, based on a novel technique of untrainable data cleansing, identifies and removes a subset of the data that AI models are unable to correctly label (classify) during the AI training process. The feasibility of their model was evaluated using a publicly available data set of 624 paediatric X-rays, of which 200 images were flagged as 'noisy' by the system. From there, a trained radiologist was given a subset of the data (100 correct images and 100 'noisy' images based on the AI's classification) for interpretation. Results showed that the reader consensus between the radiologist's label and the original label was significantly higher for the correct images (Cohen's $\kappa=0.65$) compared with the noisy images (Cohen's $\kappa=0.05$). These results indicate a high discordance in interpretation of the 'noisy' images that would make them suboptimal for a training data set. However, the feasibility of such a model remains unevaluated in HCC radiomics.

Lastly, the availability of large data sets remains scarce. In addition to the quality of data, sufficient images are also required to optimise ML performance. Most published studies have relied solely on internal data sets for the training and validation of their models. Although the use of internal validation is acceptable, the development of new and existing public databases for external validation would be a core for the future of the ML landscape. The benefit of a large, public data set is multifold. First, it improves the accuracy of feature extraction, improving the diagnostic performance of the AI model.^{80 81} The fostering of a culture of cooperation would assist in the creation of databases of sufficient size and detail to adequately train new AI models.^{82 83} Second, it provides the opportunity for smaller hospitals, which may lack the

experience or resources to create custom ML systems, to incorporate AI into their diagnostic frameworks. To that end, several sources have already made their data sets publicly available,^{81 84 85} marking a good first step towards an integrative data network. However, key concerns such as the cross-compatibility of data sets between hospital information systems, the safety and security of patient data, and the economic cost of developing the system must be addressed.

CONCLUSION

The recent surge of AI into the mainstream has raised discussion about their use in healthcare. To that extent, many hospitals have begun evaluating their feasibility. In this paper, we discussed the current landscape of AI application in the radiological diagnosis of HCC and its future directions. Current research has been optimistic, with AI models consistently reported to outperform clinicians in the interpretation of ultrasound, CT and MRI. However, rather than replace the role of clinicians, AI systems should be seen as tools bolstering the confidence of clinicians (in particular inexperienced or non-radiology trained clinicians) in their diagnosis of HCC. Despite the positive findings, it is prudent that we also acknowledge that the implementation of AI into healthcare will come with new challenges such as the financial feasibility, standardisation of use and security of data systems. All in all, the future of AI systems in diagnosing HCC looks bright but large-scale prospective is still needed.

Contributors All authors have made substantial contributions to all of the following: (1) The conception and design of the study, or acquisition of data, or analysis and interpretation of data, (2) Drafting the article or revising it critically for important intellectual content, (3) Final approval of the version to be submitted. All authors approve the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests CHN has served as a consultant for Boxer Capital.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Benjamin Koh <http://orcid.org/0000-0001-8334-0462>

REFERENCES

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
- Balogh J, Victor D, Asham EH, et al. Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma* 2016;3:41–53.
- Mazzaferro V, Regalia E, Doci R, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med* 1996;334:693–9.
- Bruix J, Cirera I, Calvet X, et al. Surgical resection and survival in Western patients with hepatocellular carcinoma. *J Hepatol* 1992;15:350–5.
- Bismuth H, Chiche L, Adam R, et al. Liver resection versus transplantation for hepatocellular carcinoma in cirrhotic patients. *Ann Surg* 1993;218:145–51.
- Ringe B, Pichlmayr R, Wittekind C, et al. Surgical treatment of hepatocellular carcinoma: experience with liver resection and transplantation in 198 patients. *World J Surg* 1991;15:270–85.
- Cucchetti A, Elshaarawy O, Han G, et al. Potentially curative therapies for hepatocellular carcinoma: how many patients can actually be cured? *Br J Cancer* 2023;128:1665–71.
- Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol* 2022;76:681–93.
- Aljumah AA, Kuriry H, AlZunaitan M, et al. Clinical presentation, risk factors, and treatment modalities of hepatocellular carcinoma: a single tertiary care center experience. *Gastroenterol Res Pract* 2016;2016:1989045.
- Patel N, Yopp AC, Singal AG. Diagnostic delays are common among patients with hepatocellular carcinoma. *J Natl Compr Canc Netw* 2015;13:543–9.
- Park J-W, Chen M, Colombo M, et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE study. *Liver Int* 2015;35:2155–66.
- Stravitz RT, Heuman DM, Chand N, et al. Surveillance for hepatocellular carcinoma in patients with cirrhosis improves outcome. *Am J Med* 2008;121:119–26.
- Bodzin AS, Busuttill RW. Hepatocellular carcinoma: advances in diagnosis, management, and long term outcome. *World J Hepatol* 2015;7:1157–67.
- Wang W, Wei C. Advances in the early diagnosis of hepatocellular carcinoma. *Genes & Diseases* 2020;7:308–19.
- Lim J, Singal AG. Surveillance and diagnosis of hepatocellular carcinoma. *Clin Liver Dis (Hoboken)* 2019;13:2–5.
- European Association For The Study Of The Liver, European Organisation For Research And Treatment Of Cancer. Treatment of C. EASL–EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 2012;56:908–43.
- Xie DY, Ren ZG, Zhou J, et al. Chinese clinical guidelines for the management of hepatocellular carcinoma: updates and insights. *Hepatobiliary Surg Nutr* 2020;9:452–63.
- Gao R, Zhao S, Aishanjiang K, et al. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. *J Hematol Oncol* 2021;14:154.
- Liu X, Khalvati F, Namdar K, et al. Can machine learning radiomics provide pre-operative differentiation of combined hepatocellular cholangiocarcinoma from hepatocellular carcinoma and cholangiocarcinoma to inform optimal treatment planning? *Eur Radiol* 2021;31:244–55.
- Li W, Lv X-Z, Zheng X, et al. Machine learning-based ultrasonics improves the diagnostic performance in differentiating focal nodular hyperplasia and atypical hepatocellular carcinoma. *Front Oncol* 2021;11:544979.
- McCarthy J. *What is artificial intelligence*. 2007.
- IBM. What is artificial intelligence (AI)? 2023. Available: <https://www.ibm.com/topics/artificial-intelligence>
- Fjelland R. Why general artificial intelligence will not be realized. *Humanit Soc Sci Commun* 2020;7.
- Bishop CM. *Pattern recognition and machine learning*. 2006.
- Wang H, Ma C, Zhou L. A brief review of machine learning and its application. 2009 International Conference on Information Engineering and Computer Science. ICIECS 2009; Wuhan, China.
- Shinde PP, Shah S. A review of machine learning and deep learning applications. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA); Pune, India.
- Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021;8:53.
- Guresen E, Kayakutlu G. Definition of artificial neural networks with comparison to other networks. *Procedia Computer Science* 2011;3:426–33.
- Walczak S, Cerpa N. Heuristic principles for the design of artificial neural networks. *Information and Software Technology* 1999;41:107–17.

- 30 Zeiler MD, Fergus R, eds. *Visualizing and understanding convolutional networks. Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014.
- 31 Albawi S, Bayat O, Al-Azawi S, et al. Social touch gesture recognition using convolutional neural network. *Comput Intell Neurosci* 2018;2018:6973103.
- 32 Attwa MH, El-Etreby SA. Guide for diagnosis and treatment of hepatocellular carcinoma. *World J Hepatol* 2015;7:1632–51.
- 33 Roberts LR, Sirlin CB, Zaiem F, et al. Imaging for the diagnosis of hepatocellular carcinoma: a systematic review and meta-analysis. *Hepatology* 2018;67:401–21.
- 34 Schraml C, Kaufmann S, Rempp H, et al. Imaging of HCC—current state of the art. *Diagnostics (Basel)* 2015;5:513–45.
- 35 Min JH, Lee MW, Park HS, et al. Interobserver variability and diagnostic performance of gadoxetic acid-enhanced MRI for predicting microvascular invasion in hepatocellular carcinoma. *Radiology* 2020;297:573–81.
- 36 Covert EC, Fitzpatrick K, Mikell J, et al. Intra- and inter-operator variability in MRI-based manual segmentation of HCC lesions and its impact on dosimetry. *EJNMMI Phys* 2022;9:90.
- 37 Shen J, Zhang CJP, Jiang B, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. *JMIR Med Inform* 2019;7:e10010.
- 38 Tiyyarattanachai T, Apiparakoon T, Marukatat S, et al. The feasibility to use artificial intelligence to aid detecting focal liver lesions in real-time ultrasound: a preliminary study based on videos. *Sci Rep* 2022;12:7749.
- 39 Kim HY, Lampertico P, Nam JY, et al. An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and caucasian patients with chronic hepatitis B. *J Hepatol* 2022;76:311–8.
- 40 Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278:563–77.
- 41 Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* 2020;20:33.
- 42 van Timmeren JE, Cester D, Tanadini-Lang S, et al. “Radiomics in medical Imaging—“How-to” guide and critical reflection”. *Insights Imaging* 2020;11:91.
- 43 Miranda J, Horvat N, Fonseca GM, et al. Current status and future perspectives of radiomics in hepatocellular carcinoma. *World J Gastroenterol* 2023;29:43–60.
- 44 Lewis S, Hectors S, Taouli B. Radiomics of hepatocellular carcinoma. *Abdom Radiol (NY)* 2021;46:111–23.
- 45 Hatt M, Parmar C, Qi J, et al. Machine (deep) learning methods for image processing and radiomics. *IEEE Trans Radiat Plasma Med Sci* 2019;3:104–8.
- 46 Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* 2018;2:36.
- 47 Chang Y, Lee J-H. Optimal modalities for HCC surveillance in a high-incidence region. *Clin Liver Dis (Hoboken)* 2020;16:236–9.
- 48 Tzartzeva K, Obi J, Rich NE, et al. Surveillance imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: a meta-analysis. *Gastroenterology* 2018;154:1706–18.
- 49 Lencioni R, Piscaglia F, Bolondi L. Contrast-enhanced ultrasound in the diagnosis of hepatocellular carcinoma. *J Hepatol* 2008;48:848–57.
- 50 Samoylova ML, Mehta N, Roberts JP, et al. Predictors of ultrasound failure to detect hepatocellular carcinoma. *Liver Transpl* 2018;24:1171–7.
- 51 Lin T-Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell* 2020;42:318–27.
- 52 Galle PR, Forner A, Llovet JM, et al. EASL clinical practice guidelines: management of hepatocellular carcinoma. *Journal of Hepatology* 2018;69:182–236.
- 53 Tsuchiya N, Sawada Y, Endo I, et al. Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J Gastroenterol* 2015;21:10573–83.
- 54 Zhang W, Hou S, Chen Y, et al. Deep learning for approaching hepatocellular carcinoma ultrasound screening dilemma: identification of A-Fetoprotein-negative hepatocellular carcinoma from focal liver lesion found in high-risk patients. *Front Oncol* 2022;12.
- 55 Chollet F. Xception: deep learning with depthwise separable convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, HI. CVPR: IEEE Computer Society, 2017:1800–7.
- 56 Andrew G. Howard MZ, Chen B, Kalenichenko D, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv* 2017.
- 57 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Sun J, ed. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA. 2016.
- 58 Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. Weinberger KQ, ed. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, HI. 2017.
- 59 Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. Wojna Z, ed. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA. 2016.
- 60 Lim JH, Kim MJ, Park CK, et al. Dysplastic nodules in liver cirrhosis: detection with triple phase helical dynamic CT. *Br J Radiol* 2004;77:911–6.
- 61 Ippolito D, Sironi S, Pozzi M, et al. Perfusion CT in cirrhotic patients with early stage hepatocellular carcinoma: assessment of tumor-related vascularization. *Eur J Radiol* 2010;73:148–52.
- 62 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.
- 63 Lee YJ, Lee JM, Lee JS, et al. Hepatocellular carcinoma: diagnostic performance of multidetector CT and MR imaging—a systematic review and meta-analysis. *Radiology* 2015;275:97–109.
- 64 Wang M, Fu F, Zheng B, et al. Development of an AI system for accurately diagnose hepatocellular carcinoma from computed tomography imaging data. *Br J Cancer* 2021;125:1111–21.
- 65 Arif-Tiwari H, Kalb B, Chundru S, et al. MRI of hepatocellular carcinoma: an update of current practices. *Diagn Interv Radiol* 2014;20:209–21.
- 66 Osho A, Rich NE, Singal AG. Role of imaging in management of hepatocellular carcinoma: surveillance, diagnosis, and treatment response. *Hepatoma Res* 2020;6:55.
- 67 Hussain SM, Reinhold C, Mitchell DG. Cirrhosis and lesion characterization at MR imaging. *Radiographics* 2009;29:1637–52.
- 68 Oyama A, Hiraoka Y, Obayashi I, et al. Hepatic tumor classification using texture and Topology analysis of non-contrast-enhanced three-dimensional T1-weighted MR images with a radiomics approach. *Sci Rep* 2019;9:8764.
- 69 Oestmann PM, Wang CJ, Savic LJ, et al. Deep learning-assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (HCC) versus non-HCC on contrast-enhanced MRI of the liver. *Eur Radiol* 2021;31:4981–90.
- 70 Lee YT, Wang JJ, Zhu Y, et al. Diagnostic criteria and LI-RADS for hepatocellular carcinoma. *Clin Liver Dis (Hoboken)* 2021;17:409–13.
- 71 Harrell FE, Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. *JAMA* 1982;247:2543–6.
- 72 Jayakumar S, Sounderajah V, Normahani P, et al. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. *Npj Digit Med* 2022;5.
- 73 Cruz Rivera S, Liu X, Chan A-W, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health* 2020;2:e549–60.
- 74 Jain A, Patel H, Nagalapatti L, et al. Overview and importance of data quality for machine learning tasks. KDD '20: Virtual Event CA USA. New York, NY, USA: Association for Computing Machinery, August 23, 2020:3561–2.
- 75 Diaz O, Kushibar K, Osuala R, et al. Data preparation for artificial intelligence in medical imaging: a comprehensive guide to open-access platforms and tools. *Phys Med* 2021;83:25–37.
- 76 Papadimitroulas P, Brocki L, Christopher Chung N, et al. Artificial intelligence: deep learning in oncological radiomics and challenges of interpretability and data harmonization. *Phys Med* 2021;83:108–21.
- 77 Fatania K, Mohamud F, Clark A, et al. Intensity standardization of MRI prior to radiomic feature extraction for artificial intelligence research in glioma—a systematic review. *Eur Radiol* 2022;32:7014–25.
- 78 Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- 79 Dakka MA, Nguyen TV, Hall JMM, et al. Automated detection of poor-quality data: case studies in healthcare. *Sci Rep* 2021;11:18005.
- 80 Althnani A, AlSaeed D, Al-Baity H, et al. Impact of dataset size on classification performance: an empirical evaluation in the medical domain. *Applied Sciences* 2021;11:796.
- 81 Quinton F, Popoff R, Presles B, et al. A tumour and liver automatic segmentation (ATLAS) dataset on contrast-enhanced magnetic resonance imaging for hepatocellular carcinoma. *Data* 2023;8:79.

- 82 Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. *JAMA* 2016;315:1237–9.
- 83 Calderaro J, Seraphin TP, Luedde T, *et al.* Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. *Journal of Hepatology* 2022;76:1348–61.
- 84 Xu Y, Zheng B, Liu X, *et al.* Improving artificial intelligence pipeline for liver malignancy diagnosis using ultrasound images and video frames. *Brief Bioinformatics* 2023;24.
- 85 Moawad AW, Morshid A, Khalaf AM, *et al.* Multimodality annotated hepatocellular carcinoma data set including pre- and post-TACE with imaging Segmentation. *Sci Data* 2023;10:33.
- 86 Shiraishi J, Sugimoto K, Moriyasu F, *et al.* Computer-aided diagnosis for the classification of focal liver lesions by use of contrast-enhanced ultrasonography. *Medical Physics* 2008;35:1734–46.
- 87 Streba CT, Ionescu M, Gheonea DI, *et al.* Contrast-enhanced ultrasonography parameters in neural network diagnosis of liver tumors. *World J Gastroenterol* 2012;18:4427–34.
- 88 Huang Q, Pan F, Li W, *et al.* Differential diagnosis of atypical hepatocellular carcinoma in contrast-enhanced ultrasound using Spatio-temporal diagnostic semantics. *IEEE J Biomed Health Inform* 2020;24:2860–9.
- 89 Tiyyarattanachai T, Apiparakoon T, Marukatat S, *et al.* Development and validation of artificial intelligence to detect and diagnose liver lesions from ultrasound images. *PLOS ONE* 2021;16:e0252882.
- 90 Virmani J, Kumar V, Kalra N, *et al.* Neural network ensemble based CAD system for focal liver lesions from B-mode ultrasound. *J Digit Imaging* 2014;27:520–37.
- 91 Mitrea D, Badea R, Mitrea P, *et al.* Hepatocellular carcinoma automatic diagnosis within CEUS and B-mode ultrasound images using advanced machine learning methods. *Sensors (Basel)* 2021;21:2202.
- 92 Weston AD, Korfiatis P, Kline TL, *et al.* Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology* 2019;290:669–79.
- 93 Nayak A, Baidya Kayal E, Arya M, *et al.* Computer-aided diagnosis of cirrhosis and hepatocellular carcinoma using multi-phase abdomen CT. *Int J Comput Assist Radiol Surg* 2019;14:1341–52.
- 94 Yamada A, Oyama K, Fujita S, *et al.* Dynamic contrast-enhanced computed tomography diagnosis of primary liver cancers using transfer learning of pretrained convolutional neural networks: is registration of Multiphasic images necessary *Int J Comput Assist Radiol Surg* 2019;14:1295–301.
- 95 Shi W, Kuang S, Cao S, *et al.* Deep learning assisted differentiation of hepatocellular carcinoma from focal liver lesions: choice of four-phase and three-phase CT imaging protocol. *Abdom Radiol (NY)* 2020;45:2688–97.
- 96 Mokrane F-Z, Lu L, Vavasour A, *et al.* Radiomics machine-learning signature for diagnosis of hepatocellular carcinoma in cirrhotic patients with indeterminate liver nodules. *Eur Radiol* 2020;30:558–70.
- 97 Kim DW, Lee G, Kim SY, *et al.* Deep learning-based algorithm to detect primary hepatic malignancy in multiphase CT of patients at high risk for HCC. *Eur Radiol* 2021;31:7047–57.
- 98 Khan AA, Narejo GB. Analysis of abdominal computed tomography images for automatic liver cancer diagnosis using image processing algorithm. *CMIR* 2019;15:972–82.
- 99 Xu X, Mao Y, Tang Y, *et al.* Classification of hepatocellular carcinoma and Intrahepatic cholangiocarcinoma based on radiomic analysis. *Comput Math Methods Med* 2022;2022:5334095.
- 100 Rocha BA, Ferreira LC, Vianna LGR, *et al.* Contrast phase recognition in liver computer tomography using deep learning. *Sci Rep* 2022;12:20315.
- 101 Duc VT, Chien PC, Huyen LDM, *et al.* Deep learning model with convolutional neural network for detecting and segmenting hepatocellular carcinoma in CT: A preliminary study. *Cureus* 2022;14:e21347.
- 102 Kim J, Min JH, Kim SK, *et al.* Detection of hepatocellular carcinoma in contrast-enhanced magnetic resonance imaging using deep learning Classifier: a multi-center retrospective study. *Sci Rep* 2020;10.
- 103 Zheng R, Wang L, Wang C, *et al.* Feasibility of automatic detection of small hepatocellular carcinoma (≤ 2 cm) in cirrhotic liver based on pattern matching and deep learning. *Phys Med Biol* 2021;66:085014.
- 104 Bousabarah K, Letzen B, Tefera J, *et al.* Automated detection and delineation of hepatocellular carcinoma on Multiphasic contrast-enhanced MRI using deep learning. *Abdom Radiol (NY)* 2021;46:216–25.
- 105 Stollmayer R, Budai BK, Tóth A, *et al.* Diagnosis of focal liver lesions with deep learning-based multi-channel analysis of hepatocyte-specific contrast-enhanced magnetic resonance imaging. *World J Gastroenterol* 2021;27:5978–88.
- 106 Cho Y, Han YE, Kim MJ, *et al.* Computer-aided hepatocellular carcinoma detection on the hepatobiliary phase of gadoteric acid-enhanced magnetic resonance imaging using a convolutional neural network: feasibility evaluation with multi-sequence data. *Comput Methods Programs Biomed* 2022;225:107032.