Xu *et al. Art Int Surg* 2023;3:48-63
DOI: 10.20517/ais.2022.33

Artificial Intelligence Surgery

Review

Open Access

Augmenting care in hepatocellular carcinoma with artificial intelligence

Flora Wen Xin Xu[1,#], Sarah S Tang[1,#], Hann Natalie Soh[2], Ning Qi Pang[3], Glenn Kunnath Bonney[3]

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119077, Singapore.
[2]Department of Internal Medicine, Singapore General Hospital, Singapore 119077, Singapore.
[3]Department of Hepatopancreaticobiliary Surgery and Liver Transplantation, National University Hospital, Singapore 119077, Singapore.
#Authors contributed equally.

Correspondence to: Prof. Glenn Kunnath Bonney, Department of Hepatopancreaticobiliary Surgery and Liver Transplantation, National University Hospital, 5 Lower Kent Ridge Rd, Singapore 119077, Singapore. E-mail: glenn_bonney@nuhs.edu.sg

How to cite this article: Xu FWX, Tang SS, Soh HN, Pang NQ, Bonney GK. Augmenting care in hepatocellular carcinoma with artificial intelligence. *Art Int Surg* 2023;3:48-63. https://dx.doi.org/10.20517/ais.2022.33

Received: 1 Nov 2022 **First Decision:** 31 Jan 2023 **Revised:** 22 Feb 2023 **Accepted:** 16 Mar 2023 **Published:** 29 Mar 2023

Academic Editors: Derek O'Reilly, Andrew A. Gumbs **Copy Editors:** Ke-Cui Yang, Yanbing Bai **Production Editors:** Ke-Cui Yang, Yanbing Bai

Abstract

Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related death worldwide and prognosis remains poor. The recent paradigm shifts in management algorithms of such patients have resulted in unique challenges in the early identification of HCC, prognosis, surgical outcomes, prioritization of potential transplant recipients, donor-recipient matching, and so on. In recent years, advancements in artificial intelligence (AI) capabilities have shown potential in HCC treatment.

In this narrative review, we outline first the different types of AI models that are applied in clinical practice and then focus on the frontiers of AI research in the diagnosis, prognostication, and treatment of HCC, particularly in classification of indeterminate liver lesions, tumor staging, survival prediction, improving equity in transplant recipient selection, prediction of treatment response and prognosis. We show that US coupled with AI-driven predictive models can provide accurate noninvasive screening tools for early disease. While AI models applied to contrast-enhanced CT, MRI and PET studies may appear to have limited clinical utility in disease diagnosis and differentials, owing to their accuracy, we highlighted the importance of such models in predicting pathological findings preoperatively. Despite the availability of many accurate, sensitive, and specific AI algorithms that outperform traditional scoring systems, they have not been widely used in clinical practice. The challenges in AI application, including distributional shift and imbalanced data, lack of standardization, and the 'black box'

phenomenon, are addressed here. The importance of AI in the future of HCC makes it important for clinicians to have a good understanding of different AI techniques, their benefits, and potential pitfalls.

Keywords: Hepatocellular cancer, liver cancer, liver imaging, liver surgery, artificial intelligence, machine learning, neural network

INTRODUCTION

Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related death worldwide, with an estimated incidence of 700,000 annually[1]. While the past decade has seen paradigm shifts in the way HCC is diagnosed and treated[2,3], prognosis remains poor (5-year overall survival stands at less than 20%[4]) owing to multiple factors including the difficulty in identifying HCC in its early stages[5]. Early surveillance of high-risk patients is done via six-monthly abdominal ultrasound and serum α-fetoprotein (AFP) measurements[6], but both confer limited accuracy in identifying early-stage HCC, where nodules are small and indeterminate[7,8]. Sensitivity is particularly low in patients with underlying cirrhosis, steatosis, or obesity[7,8]. Moreover, the success of liver resection and transplantation for HCC is primarily dependent on patient selection, for which existing clinical scores rely heavily on rudimentary quantitative measures such as the size and multicentricity of the main nodule[2,9,10]. With mounting evidence to suggest that early diagnosis, biological stratification and treatment of HCC drastically improves survival outcomes[5,11,12], it is paramount that clinicians identify better tools for such purposes and rethink the way we approach diagnostication.

In recent years, advancements in artificial intelligence (AI) capabilities have shown great potential to redefine the way we navigate clinical care for HCC patients. AI has the capacity to improve risk prediction in chronic hepatitis patients[13], accelerate the diagnostic process with early identification of HCC[14-16], increase accuracy in the classification of liver lesions and HCC subtypes[17-20], tumor staging[21], and survival prediction[22,23]. Decisions regarding candidate selection and optimal treatment methods may also utilize AI in the prediction of treatment response, progression-free and overall survival[24,25] and risk of HCC recurrence[26].

Broadly, AI comprises machine learning (ML), deep learning (DL), and neural networks (NN). Each differs in terms of how the predictive model is built, the type of input data required, and the interpretability of the model itself. ML models are primarily built with the intent of improving predictions and decision-making accuracy. These models can be further distinguished into supervised and unsupervised learning[27]. Supervised learning algorithms train on sample input data with labeled outcome data, and their goal is to learn the relationship between the input data and the outcomes to make accurate predictions about the outcome when provided with a new set of input data[28]. Examples of supervised learning algorithms include traditional techniques such as linear regression and logistic regression, as well as more sophisticated techniques including support vector machines, random forest, and gradient boosting[28]. Unsupervised learning algorithms train on unlabeled sample data and analyze the underlying structure or distribution within the data to discover new clusters or patterns[29]. Examples of unsupervised learning algorithms include various other techniques such as K-means and principal component analysis[29].

Deep Learning (DL) aims to form computing systems that emulate biological neural networks. DL methods include the use of multilayered artificial neural networks (ANNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs)[30]. ANNs are formed by a network of perceptrons or neurons processed in a feed-forward fashion and are good for mapping nonlinear functions in text, tabular, or image

data[31,32]. CNNs have connective patterns resembling the visual cortex and can detect inherent spatial features of high-dimensional images[33]. RNNs have connections forming a graph over a temporal sequence, thus being useful in time series prediction[34]. In DL models, a significant "black box problem" remains as the programs have low interpretability and users may not completely understand how they work[35].

In this narrative review, we will outline the frontiers of AI research in the diagnosis, prognostication, and treatment of HCC.

DIAGNOSIS OF HCC

There have been remarkable advances in the application of AI to aid traditional diagnostic techniques for HCC in recent years. This is primarily due to the use of DL algorithms using CNN, which is a multilayer ANN interconnected such that all input data is processed through multiple layers to produce valuable output data[36]. CNN algorithms trained on various imaging modalities such as ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) have been shown to increase the diagnostic yield in terms of identification, classification, staging and survival prediction in HCC[37]. All findings are summarized in Table 1.

Prediction of cirrhosis and HCC development

HCC often occurs on a backdrop of longstanding cirrhosis[49], yet cirrhotic changes can remain elusive until its later stages[50]. Standard radiological features on imaging include a nodular hepatic contour, changes in volume distribution with enlargement of the caudate lobe and the left lateral segment, atrophy of the right and left lobe medial segments, widening of the fissures and the porta hepatis, and the formation of regenerative nodules[50]. In response to this problem, Liu *et al.* designed an algorithm to determine the presence or absence of cirrhosis in US images with an area under the curve (AUC) of 0.968[38]. Using their analysis of liver capsule morphology, the DL program could identify early cirrhotic changes often invisible to the human eye. Expanding on this, a novel ML model by Ksiazek *et al.* forecasted the risk of HCC development based on 23 quantitative and 26 qualitative features gleaned from biochemistry, and clinical factors like viral status and comorbidities, ultimately achieving 88.5% accuracy[39]. Such predictive models, when coupled with other noninvasive methods in predicting fibrosis and cirrhosis, are likely to be developed further and be seen routinely in clinical practice in early disease detection.

Radiological identification

Ultrasound

Current clinical guidelines recommend regular abdominal US surveillance for the identification of HCC in high-risk patients with chronic viral hepatitis or cirrhosis[51]. US is, therefore, usually the primary tool to evaluate early liver disease and detect new lesions. However, image interpretation is subject to limitations such as inter-observer variability and patient body habitus, resulting in a sensitivity of only 63%[51]. For example, liver neoplasms can be difficult to distinguish from liver parenchyma, particularly with small indeterminate lesions[52] or diffuse HCC in the setting of cirrhosis[53]. To address this, several studies have proposed AI algorithms with data from various imaging modalities to improve the diagnostic accuracy of HCC.

To delineate HCC from background cirrhosis, Bharti *et al.* devised an ANN to classify US images into four stages of liver disease (normal liver, chronic liver disease, cirrhosis, and HCC) with an accuracy of 96.6%[14]. More recently, Brehar *et al.* also proposed a CNN model built on two independent datasets of US images that outperformed conventional ML methods (SVM, RF, multilayer perceptron, and AdaBoost)[40].

Table 1. Diagnosis of HCC

Study	Title	Study aim	Diagnostic technique	AI tool	Performance
Liu *et al.*[38]	Learning to diagnose Cirrhosis with liver capsule Guided ultrasound image classification	Early identification of cirrhosis	US	ML	AUC: 0.968
Ksiazek *et al.*[39]	A novel machine learning Approach for early detection of hepatocellular carcinoma Patients	Prediction of HCC risk	US	ML	Accuracy: 88.5%
Bharti *et al.*[14]	Preliminary study of chronic liver classification on ultrasound images using an ensemble model	Classification of liver disease into four stages (normal liver, chronic liver disease, cirrhosis and HCC)	US	ANN	Accuracy: 96.6%
Brehar *et al.*[40]	Comparison of deep-learning and conventional machine-learning methods for the automatic recognition of the hepatocellular carcinoma areas from ultrasound Images	Differentiate HCC from cirrhotic parenchyma	US	CNN	AUC: 0.95 Accuracy: 0.91 Sensitivity: 94.4% Specificity: 88.4%
Schmauch *et al.*[15]	Diagnosis of focal liver lesions from ultrasound using deep learning	Classification of liver lesions as benign or malignant	US	DL	AUC: 0.93 for benign lesions, 0.92 for malignant lesions
Guo *et al.*[41]	A two-stage multi-view learning framework-based computer-aided diagnosis of liver tumors with contrast enhanced ultrasound images	Classification of liver lesions as benign or malignant	CEUS	ML	Accuracy: 90.41% Sensitivity: 93.56% Specificity: 86.89% Youden index: 79.44% False positive rate: 13.11% False negative rate: 6.44%
Yang *et al.*[42]	Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: A multi-center study	Classification of liver lesions as benign or malignant	US	CNN	AUC: 0.924 (external validation)
Streba *et al.*[43]	Contrast-enhanced ultrasonography parameters in neural network diagnosis of liver tumors	Classification of focal liver lesions	US	ANN	Accuracy: 87.12% Sensitivity: 93.2% Specificity: 89.7%
Hassan *et al.*[44]	Diagnosis of focal liver diseases based on deep learning technique for ultrasound images	Classification of focal liver lesions	US	Auto-encoder	Accuracy: 97.2% accuracy Sensitivity: 98% Specificity: 95.70%
Shi *et al.*[45]	Deep learning assisted differentiation of hepatocellular carcinoma from focal liver lesions: choice of four-phase and three-phase CT imaging protocol	Classification of focal liver lesions	CT	CNN	AUC: 0.925
Yasaka *et al.*[46]	Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A Preliminary Study	Classification of focal liver lesions	CT	CNN	AUC: 0.92
Sun *et al.*[21]	LiSNet: An artificial intelligence -based tool for liver imaging staging of hepatocellular carcinoma aggressiveness	Prediction of MVI in HCC, and scoring HCC aggressiveness	CT	ML	AUC: 0.668 for predicting histopathological MVI Agreement rate of LiSNet with subspecialists: 0.658, 0.595 and 0.369 for scoring HCC aggressiveness grades I, II, and III
Hamm *et al.*[47]	Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multiphasic MRI	Classification of focal liver lesions	MRI	CNN	AUC: 0.992 for HCC identification Sensitivity: 90% for classifying FLLs Specificity: 98% for classifying FLLs

Preis *et al*.[48]	Neural network evaluation of PET scans of the liver: a potentially useful adjunct in clinical interpretation	Identify metastatic liver disease	PET	CNN	AUC: 0.905 for CNN incorporating lesion data, compared to 0.786 for blinded observers; 0.896 for CNN independent of lesion data, compared to 0.796 for blinded observers

AUC: area under the curve; ANN: artificial neural network; CT: computed tomography; CNN: convolutional neural networks; CEUS: contrast-enhanced US; DL: deep learning HCC: Hepatocellular carcinoma; MVI: microvascular invasion; MRI: magnetic resonance imaging; ML: machine learning; PET: positron emission tomography; US: ultrasound.

Beyond distinguishing liver lesions from background tissue, AI also has demonstrable utility in classifying these lesions as benign or malignant. Schmauch *et al*. built a supervised DL model using a training dataset of 367 US images with their corresponding radiological reports, achieving a mean AUC of 0.93 and 0.92, respectively, in determining benign versus malignant masses[15]. Guo *et al*. also established that DL can be applied to contrast-enhanced US (CEUS) to discriminate benign and malignant liver neoplasms[41]. Recently, Yang *et al*. designed a large CNN incorporating clinical features and radiomic features like lesion size and liver background echo. Their model is one forerunner in AI-based US interpretation, achieving an AUC of 0.924 in an external validation cohort, with diagnostic capabilities comparable to contrast-enhanced CT (CECT) and exceeding that of skilled radiologists with 15 years of experience in diagnosing focal liver lesions[42].

The preoperative pathological classification of HCC and liver parenchyma is important to the determination of tumor extent and treatment planning. Streba *et al*. prospectively studied CEUS images of 112 patients to train an ANN that classified five different types of liver tissue (HCC, hypervascular or hypovascular liver metastasis, hepatic hemangioma, or focal fatty changes) and achieved promising results. Their automatic classification process achieved 93.2% sensitivity, 89.7% specificity, 94.42% positive predictive value, and 87.57% negative predictive value, which was comparable to human interpretation[43]. Hassan *et al*. reported using an unsupervised DL technique, the stacked sparse auto-encoder, to segment and classify liver lesions on US images with a classification accuracy of 97.2%[44]. Optimizing an AI solution on US findings in accurately detecting HCC will prove a less invasive manner in which screening could be meaningful (negating the use and access to CECT).

CT, MRI, PET

A noteworthy advancement in CT imaging is the creation of a CNN by Shi *et al*. that enabled accurate HCC identification using a three-phase CT protocol. Their model achieved similar diagnostic accuracy when compared to a four-phase protocol, potentially allowing patients to receive lower doses of radiation[45]. To categorize liver lesions identified on CT, Yasaka *et al*. designed a model to differentiate liver lesions on CT into five categories: HCC, other malignant tumors, indeterminate masses, hemangiomas, and cysts, with a median AUC of 0.92[46]. Most recently, the LiSNet AI tool was developed for staging of HCC aggressiveness using CT images, where Sun *et al*. showed results comparable to subspecialist analysis[21]. A human-AI partnered diagnosis was also attempted, combining experience-based binary diagnosis and LiSNet, resulting in the best predictive ability for certain parameters such as microvascular invasion (MVI) with AUC 0.705[21].

Hamm *et al.* used MRI images from 494 patients to train a CNN which can classify hepatic lesions into six different categories (benign cysts, cavernous hemangiomas, focal nodular hyperplasia, HCC, intrahepatic cholangiocarcinoma, and colorectal metastasis, even outperforming expert radiologists in HCC detection (90% *vs.* 60%-70% sensitivity)[47]. Preis *et al.* improved this study and reported that incorporating lesion data from PET-CT into an ANN achieved high sensitivity and specificity in detecting liver cancer unidentified visually, with an AUC of 0.905[48]. While such endeavors in AI models for CT, MRI and PET are laudable, the real-world clinical utility of this is likely to be limited for a clinician as a combination of these scans already achieves high accuracy in diagnosis. However, the human-AI algorithms, such as LiSNet (highlighted above), that can predict biology better (microvascular invasion in this instance) would be of important clinical utility and we highlight this below.

PROGNOSTICATION

Staging

Besides serving as efficient tools in the detection and classification of liver tumors, AI models can utilize data for staging and prognostication. One of the key prognostic factors in HCC is vascular invasion[54]. Jiang *et al.* developed two predictive models using DL and XGBoost, a distributed gradient-boosted decision tree ML library, to detect MVI using CT images from 405 patients, with an AUC of 0.952-0.980[55]. Zhang *et al.* also developed a 3D-CNN model to predict MVI in HCC, with an AUC of 0.81[56]. Findings are summarized in Table 2. However, in a real-world context, the prediction of MVI preoperatively in resectable or transplantable (within criteria) HCC remains a contentious one. The rapidly expanding neoadjuvant and peri-operative systemic treatment options in the field may result in better case selection and preoperative treatment of patients with MVI prior to resection or transplantation.

Liver segmentation

Many developed imaging modalities such as CT, MRI, PET and US are used for the liver's morphological and volumetric analysis and diagnosis of associated diseases[59]. They are useful for their capability of giving surgeons insights into the current state of organs non-invasively. With the existence of such modalities, computer-aided detection (CAD) systems have become significantly more important[60]. Furthermore, CT, MRI and PET can generate 3-dimensional (3D) holistic organ volumes for more informative image slices with accurate anatomical information. These modalities are utilized extensively for clinical applications including cancer diagnosis, tumor burden quantification, surgical planning and organ transplantation[60]. Additionally, such modalities are used for adaptive radiation therapy, which is a radiation treatment plan that is customized based on the patient's functional changes during a course of radiation[61]. In another clinical procedure, a pre-procedural CT or MRI scan can help in interventional endoscopy for pancreatic and biliary diseases, as image guidance can be supportive in intra-procedural navigation to specific gastrointestinal positions[62]. All the aforementioned reasons demonstrate the importance of segmenting the liver to aid in disease diagnosis and prognosis.

Survival prediction

Beyond detecting HCC on imaging, several studies have proposed AI algorithms for survival prediction. Using CEUS images taken prior to treatment, Liu *et al.* devised a DL radiomics model to project post-treatment progression-free survival (PFS) in HCC patients as a future selection tool between treatment options (see section 4.4)[57]. Zhang *et al.* built a DL-based model predicting overall survival using CT images from 201 patients with unresectable HCC treated with TACE and sorafenib, which achieved superior predictive performance compared to the clinical nomogram (C-index 0.730)[58].

Table 2. Prognostication of HCC

Study	Title	Study aim	Diagnostic technique	AI tool	Performance
Jiang *et al.*[55]	Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning	Identification of MVI in HCC	CT	CNN	AUC: 0.906
Zhang *et al.*[56]	Deep learning with 3d convolutional neural network for noninvasive prediction of microvascular invasion in hepatocellular carcinoma	Prediction of MVI in HCC	MRI	CNN	AUC: 0.72 Sensitivity: 55% Specificity: 81%
Liu *et al.*[57]	Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. liver cancer	Prediction of 2-year progression-free survival (PFS) of radiofrequency ablation (RFA) and liver resection prior to treatment; Optimize treatment selection for patients with very early and early-stage HCC	CEUS	DL	C-index: 0.726 for RFA, 0.741 for liver resection
Zhang *et al.*[58]	Deep Learning predicts overall survival of patients with unresectable hepatocellular carcinoma treated by transarterial chemoembolization plus Sorafenib	Prediction of overall survival in HCC after treatment with TACE and Sorafenib	CT	CNN	C-index: 0.717 in training set, 0.714 in validation set
Simsek *et al.*[23]	Artificial intelligence method to predict overall survival of hepatocellular carcinoma	Prediction of overall survival in HCC	Clinical, Biochemical	ML	AUC: 0.92 for >6 months, 0.81 for >1 year, 0.78 for >2 years, 0.81 for >3 years, 0.82 for >5 years, 0.81 for >8 years, and 0.66 for >10 years

AUC: Area under the curve; CT: computed tomography; CNN: convolutional neural networks; CEUS: contrast-enhanced US; DL: deep learning HCC: hepatocellular carcinoma; MVI: microvascular invasion; MRI: magnetic resonance imaging; ML: machine learning; TACE: transarterial chemoembolization.

Recently, Simsek *et al*. reported a DL model studying non-radiological features (age, bilirubin, AFP, smoking status, alcoholic liver disease etiology, and GGT) predicted overall survival of HCC patients at short and long-term intervals (AUC 0.92)[23]. With the established role of immunotherapy in the management algorithm of HCC[63], these studies at present may have limited clinical applicability. However, at present, it must be noted that standard molecular markers of sensitivity to immunotherapy, such as microsatellite instability, tumor mutational burden and mismatch repair, have a limited role in predicting responders to immunotherapy in HCC[64,65]. The principles of radiomic and DL methods, as described above, may indeed prove to be the mainstay of such predictions prior to HCC treatment in the future. All findings are summarized in Table 2.

TREATMENT OF HCC

Liver resection

Survival outcomes after resection

Liver resection is recommended as first-line therapy for patients with HCC, but there is a paucity of outcome prediction models to aid in patient selection and postoperative tumor recurrence remains high. Traditionally, the decision for surgery is guided by treatment pathways such as the Barcelona Clinic Liver Cancer (BCLC) algorithm[2]. With the emergence of AI tools that combine clinical, biochemical, and multimodal radiological features, there is potential for more accurate preoperative identification of HCC patients at higher risk of recurrence.

Ji *et al.* designed an ML framework that identified a three-feature radiomic signature of contrast-enhanced CT images. To further boost prediction performance, clinical factors and biochemical measures like the serum AFP level and albumin-bilirubin grade were included. Their model achieved a C-statistic of 0.73 and outperformed conventional metrics of prognostication like BCLC scoring[66]. Wang *et al.* devised a similar combined model using multiphasic CT features and clinical factors, yielding promising results with an AUC of 0.82. In a similar vein, Saillard *et al.* employed a DL model based on digitized histological slides that could predict post-resection survival more accurately than relevant clinical, biological, and pathological factors[67]. However, these findings were not upheld when subjected to external validation. Post-resection features predicting survival have had limited clinical impact due to the lack of adjuvant treatment options in HCC previously. With continued expansions and trials in adjuvant treatment in HCC, such features may have relevance when incorporated into survival prediction post-resection.

Liver transplantation

Recipient selection

The Model for End-Stage Liver Disease (MELD) score, originally devised to prognosticate patients after a transjugular intrahepatic portosystemic shunt (TIPS) procedure for portal hypertension, has been used since 2002 for prioritizing donor liver allocation in liver transplantation in a "sickest-first" approach[68]. This logarithmic score comprises biochemical factors like the International Normalized Ratio (INR), serum creatinine, and total serum bilirubin. While regional allocation policies may differ, the final MELD score given to a patient on the waiting list usually gives additional 'exception points' after considering the etiology of cirrhosis as well[69]. This model has served patients around the world well for many years, but it is gradually being superseded by more updated listing criteria. The MELD score has been critiqued for being disadvantageous to female patients because of its inclusion of serum creatinine (typically lower in females) without correction for gender. While the new MELD 3.0 score promises to correct for gender bias, the question remains – could AI-based models outperform this, either supervised on unsupervised?

The Optimized Prediction of Mortality (OPOM) model employs ML optimal classification tree models to more accurately predict three-month mortality compared to the MELD score. Specifically, a model was calibrated based on optimal classification trees (or OCTs), which represented a ML prediction method that afforded interpretability and high prediction accuracy. This predictive model was trained on historical data of patients in the United States from 2002 to 2016 (comprising 1,618, 966 patient observations) obtained from the Scientific Registry of Transplant Recipients (SRTR) in a Liver Simulation Allocation Model (LSAM). The end product was a classification tree that predicted the probability of a patient dying or becoming unsuitable for transplant within 3 months (the dependent variable), given observations of certain patient characteristics (the independent variables). Bertsimas *et al.* showed that OPOM allocation reduced mortality by 417.96 deaths per year compared to MELD[70]. Indeed, although a simple method to stratify candidates awaiting liver transplantation, the MELD score is a linear regression method that does not accurately predict mortality for all candidates who can benefit from liver transplantation. This is especially demonstrated in the significant deterioration in MELD predictive capabilities with increasing disease severity compared to OPOM. In contrast to MELD, which demonstrated decreasing AUC values as sicker patient strata are considered, OPOM maintained significantly higher AUCs, especially within the sickest candidate population, thus allowing for a more accurate prediction of waitlist mortality. A recent study by Yu *et al.* using ML in a Korean cohort also showed superior outcomes of its random forest model (AUC 0.80-0.85) compared to using the MELD score (AUC 0.70)[71].

Unfortunately, the OPOM experimental model has yet to be validated in other centers with HCC patient cohorts. It should be noted that LSAM analysis is also limited in that it only allows for an accurate assessment of waitlist deaths, as waitlist removals include not only candidates with deterioration in their

condition, but also those removed due to improvement in their condition. It should be noted that OPOM allocation does not address the issues in liver distribution, nor the resultant geographic disparity that exists between the united network for organ sharing (UNOS) regions and donor service areas (DSAs)[66]. Similarly, for the Korean random forest ML model, despite its superior outcomes, organ shortage is the main hurdle for organ transplantation and liver allocation remains a major issue[71].

Donor matching

Liver transplantation has traditionally relied on MELD score and (in living donors) volumetric matching between donor and recipient to achieve an ideal pairing. Beyond simply using AI algorithms to derive a "better MELD score", there has been a fundamental shift away from recipient selection and ranking alone to donor-recipient (D-R) matching models. One of the most widely debated models for D-R matching is an ANN by Briceno *et al.* analyzing 64 different variables and their effects on the probability of graft survival and reduction of graft loss[72]. They found that utilizing their ANN yielded superior results compared to current validated scores, including MELD, D-MELD, DRI, P-SOF, SOFT, and BAR[72].

However, the use of AI in D-R matching is also not without its limitations. A recent 2021 study by Gujio-Rubio *et al.* compared modeling techniques using standard statistical methods (including logistic regression and naive Baynes) to standard machine learning methods (including multilayer perceptron, random forest, gradient boosting and support vector machines) and standard scores (MELD, SOFT and BAR)[73]. Of note, the study concluded that logistic regression (AUC 0.654) outperformed ML techniques (AUC 0.599-0.644) and also outperformed standard scores[73]. This adds further uncertainty to the true utility of AI techniques in liver transplantation, which will be discussed below.

Transarterial chemoembolization

Transarterial Chemoembolization (TACE) is typically used to treat Stage B HCC following the BCLC guidelines. Patient selection is key to ensuring that patients suitable for upfront resection do not delay definitive curative treatment. Several models have been developed based on clinical data and CT or MRI imaging features. These include the ML and DL models developed by Peng *et al.*[74], Morshid *et al.*[75], Liu *et al.*[76] amongst others - these models have produced fairly satisfactory results, with an AUC of 0.93-0.97 for predicting TACE response.

Radiofrequency ablation

RFA is used to treat both early-stage HCC and unresectable diseases. In selected patients, this treatment modality aims for curative treatment that confers lower morbidity than traditional liver resection and/or transplantation would. Liang *et al.* proposed an ML model in 2014 looking at recurrence after RFA, attaining an AUC of 0.69. In this study, high-risk patients could be identified and followed up closely after RFA treatment for surveillance. In 2020, Liu *et al.* further developed a novel DL-based radiomic strategy to predict 2-year PFS among 419 patients with very early and early-stage HCC, using CEUS images taken one week prior to liver resection (n = 205) or RFA (n = 214). Their updated model achieved accurate pre-treatment predictions of future PFS (C-index 0.741 for liver resection, 0.726 for RFA), potentially serving as a future tool for patient selection between the two options[57]. All findings are summarized in Table 3.

CURRENT CHALLENGES IN THE APPLICATION OF AI

In his celebrated thesis 'The Critique of Pure Reason', Immanuel Kant asks: "What can we know?" "What should we do?" "What is reasonable to hope for[77]?" In the application of AI to clinical practice, this is a relevant framing for us to consider its further development and its applicability. The current exponential development of AI and its accompanying hardware has resulted in landmark scientific discoveries to date,

Table 3. Management of HCC

Study	Title	Study aim	Diagnostic technique	AI tool	Performance
Ji *et al.*[66]	Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: a multi-institutional study	Prediction of HCC recurrence	CECT	ML	C-index: 0.733-0.801 Integrated Brier score: 0.147-0.165
Saillard *et al.*[67]	Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides	Prediction of survival in HCC patients after surgical resection	Histopathology	CNN	C-index: 0.75-0.78
Bertsimas *et al.*[70]	Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation	Prediction of candidate's 3-month waitlist mortality or removal	Standard Transplant Analysis and Research (STAR) dataset	ML	Compared to MELD, OPOM allocation reduced mortality by 417.96 deaths per year
Yu *et al.*[71]	Artificial intelligence for predicting survival following deceased donor liver transplantation: retrospective multicenter study	Prediction of survival following liver transplantation using traditional statistical models versus ML approaches	Deceased donor liver transplant recipients variables	ML	AUC: 0.80-0.85
Briceño *et al.*[72]	Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter spanish study	Donor-recipient (D-R) matching in liver transplantation, comparison of ANN accuracy with validated scores of graft survival	D-R variables	ANN	Prediction of probability of graft survival (90.79%) and -loss (71.42%)
Gujio-Rubio *et al.*[73]	Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation	Analyze how several ML techniques behave in the largest liver transplant database	United Network for Organ Sharing database	ML	AUC: 0.654 for logistic regression AUC: 0.599-0.644 for ML
Peng *et al.*[74]	Residual convolutional neural network for predicting the response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging	Prediction of response to TACE	CT	CNN	AUC: 0.97 Accuracy: 84.3%
Morshid *et al.*[75]	A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. radiology artificial intelligence	Prediction of response to TACE	CT	ML	Accuracy: 74%
Liu *et al.*[76]	Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound	Prediction of response to TACE	CEUS	DL	AUC: 0.93

AUC: area under the curve; ANN: artificial neural network; CT: computed tomography; CNN: convolutional neural networks; CEUS: contrast-enhanced US; CECT: contrast-enhanced CT; DL: deep learning HCC: Hepatocellular carcinoma; MELD: the model for end-stage liver disease; OPOM: the optimized prediction of mortality; TACE: transarterial chemoembolization.

including the discovery of a novel protein folding structure and a new clinically approved antibiotic, firmly establishing its role in translational sciences[78,79]. However, the "AI chasm", a term coined to reflect the gulf between AI development and deployment[80], remains an important practical challenge in clinical utility. Despite the multifold benefits of using AI as an adjunct in clinical decision-making, its application has been relatively slow to be adopted across the clinical arenas.

Existing barriers to the use of AI approaches include the lack of standardized algorithms and software used across institutions, difficulty justifying AI-based predictions given the "black box" phenomenon, and poor generalizability outside the training set. ML algorithms require external validation in independent datasets with patient populations of substantial size and diversity for successful training[81,82]. There are also considerable differences between experimental algorithms written for proof-of-concept studies and those required for producing a marketable healthcare product. The latter must be done following Good Manufacturing Practice guidelines by the Food and Drug Administration[83], often requiring immense labor and experience.

Distributional shift and imbalanced data

Distributional shift is a critical problem in model creation[84]. ML models perform best when index cases and control cases are similar in the training set[85], but this is rarely the case with HCC. Disease patterns in cirrhosis and cancer also evolve drastically over time (such as the current epidemic of non-alcoholic fatty liver disease), resulting in mismatches between training and operational data. Imbalanced datasets can be "re-balanced" with under-sampling or over-sampling, but a failure to correct inherent biases will result in a model that over-diagnoses rare cases[86].

Lack of standardization

In pursuit of safety and efficacy in AI use, standardization is key. As described above, comparability and reproducibility remain poor across studies due to gross inconsistencies in data management, imaging and data processing equipment used, and the reporting of methods and results. Common metrics used in reporting the results of AI prediction, such as area under the curve, sensitivity and specificity, do not reliably show clinical efficacy[87]. Biomedical researchers should strongly consider following standardized guidelines for reporting published by Luo *et al.* in 2016[88]. Their seminal work highlights how most pitfalls of applying ML in medicine originate from a small set of common issues like data leakage and overfitting. They have thus generated guidelines for developing predictive models and a minimum list of reporting items, including information on independent variables, negative or positive examples and modeling technique selection[89]. The majority of clinical studies reported here fail to reach such reporting standards. Scientific publications should stipulate such reporting standards in AI-based studies as part of quality assurance and, therefore, potential clinical consideration, something the scientific community "should do".

Overfitted data and generalizability

Following the initial success of various models trained and tested on small datasets, few have translated to any real-world impact because of problems with data overfitting and difficulty generalizing results to other patient populations[89]. The application of AI in HCC remains an emerging field and most algorithms require training on diverse datasets, as well as testing with external validation or prospective trials. Several studies discussed have managed to maintain high accuracy rates in independent external validation cohorts. For instance, the AI model for predicting HCC risk in chronic hepatitis B patients developed by Kim *et al.* using a Korean cohort (C-index: 0.79) remained accurate in testing against both an independent external Korean validation cohort (C-index: 0.79) and an independent external Caucasian validation cohort (C-index: 0.81)[13]. Notably, the training/derivation cohort, external Korean validation cohort and external Caucasian validation cohorts differed in their baseline characteristics and had significant differences in age and prevalence of cirrhosis[13]. Other AI models that have achieved similar results include the ML analysis of contrast-enhanced CT radiomics for HCC recurrence by Ji *et al*[66]. The inclusion of such external national and international cohorts would rapidly advance generalizability.

Black box phenomenon

With the use of "black-box" algorithms in NNs, even developers do not fully understand the underlying mechanisms for automated decision-making[90], thus making it difficult to explain results to doctors and patients. In HCC research, programs like DeepDream have been applied to aid NN visualization in tumor segmentation of CT liver images[91]. Nonetheless, such post-hoc models have been criticized out of concerns regarding the fidelity and logicality of explanations provided; Rudin *et al.* recommend the creation of inherently explainable models instead[92]. Accepting that AI has already demonstrated greater efficacy in recognizing novel patterns and relationships than supervised standard mathematical modeling, the question remains: is transparency ethically imperative in clinical decision making even if that model far outperforms any previous modeling? Is this what we should "reasonably hope for" in the future of NN studies in clinical practice?

Moving towards clinical use

The models developed have shown their potential to add great value to patient care. However, a concerted effort is required for meta-analyses to sieve out front-runner models and for clinicians to validate those models both locally and internationally. Secondly, when the models are mature enough, collaboration with ethics review boards and local government will be crucial for deployment into actual clinical practice. Lastly, the end-users of the product being clinicians, we should also seek to understand the science behind AI algorithms, overcome the 'black-box' uncertainty of AI, and be confident in using them in practice. As a community, this is something "we should do". In order to overcome this, more so in AI-based algorithms than standard formulae, there is a great necessity for external validation of such models with global collaborative studies. To this end, the opacity of the AI model requires stringent data entry and quality assurance that will require careful central control and data monitoring.

CONCLUSION

The utilization of AI in the care of HCC patients is a field that has grown exponentially in the past few years, with particular areas of care (e.g., liver transplantation and imaging in HCC) being more hotly debated and investigated than others. We summarize in this article that some AI solutions are also more acceptable than others - algorithmic approaches may be more easily grasped as compared to NN and DL models. In addressing the three questions posed by Kant mentioned above, it is clear that AI has established itself as a tool with limitless learning ability. However, addressing what we should do with this data and what is reasonable to hope for remains critical to its adoption. Efforts to establish collaborative datasets and sound external validation in the global scientific and clinical communities will be integral to this. With sound validation studies in well-curated clinical cohorts and clear reporting standards, some of the five concerns we put forward are likely to be allayed; thereby, AI application become mainstream in the care of HCC.

DECLARATIONS

Authors' contributions

Study concept and design: Bonney GK , Pang NQ
Review of literature: Xu FWX, Tang SS , Soh HN
Drafting of manuscript: Xu FWX , Tang SS , Soh HN
Critical review of manuscript: Xu FWX , Tang SS , Soh HN, Pang NQ, Bonney GK
All authors read and approved the final manuscript: Xu FWX, Tang SS, Soh HN, Pang NQ, Bonney GK

Availability of data and materials

Not applicable.

Financial support and sponsorship
None.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

REFERENCES

1. Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers* 2021;7:6. DOI
2. 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. DOI
3. Finn RS, Qin S, Ikeda M, et al; IMbrave150 investigators. atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *N Engl J Med* 2020;382:1894-905. DOI
4. Brar G, Greten TF, Graubard BI, et al. Hepatocellular carcinoma survival by etiology: a seer-medicare database analysis. *Hepatol Commun* 2020;4:1541-51. DOI PubMed PMC
5. Yang JD. Detect or not to detect very early-stage hepatocellular carcinoma? *Clin Mol Hepatol* 2019;25:335-43. DOI
6. SD; British society of gastroenterology. guidelines for the diagnosis and treatment of hepatocellular carcinoma (HCC) in adults. *Gut* 2003;52 Suppl 3:iii1-8. DOI
7. Best J, Sydor S, Bechmann LP, Canbay A. Evaluation and impact of different biomarkers for early detection of hepatocellular carcinoma. *HR* 2020:2020. DOI
8. Simmons O, Fetzer DT, Yokoo T, et al. Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis. *Aliment Pharmacol Ther* 2017;45:169-77. DOI PubMed PMC
9. 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. DOI PubMed
10. Massad E, Chaib E. Liver tumors and liver transplantation. Elsevier; 2020. pp. 97-115. DOI
11. Burak KW. Prognosis in the early stages of hepatocellular carcinoma: predicting outcomes and properly selecting patients for curative options. *Can J Gastroenterol* 2011;25:482-4. DOI PubMed PMC
12. Farinati F, Sergio A, Baldan A, et al. Early and very early hepatocellular carcinoma: when and how much do staging and choice of treatment really matter? *BMC Cancer* 2009;9:33. DOI PubMed PMC
13. 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. DOI PubMed
14. Bharti P, Mittal D, Ananthasivan R. Preliminary study of chronic liver classification on ultrasound images using an ensemble model. *Ultrason Imaging* 2018;40:357-79. DOI PubMed
15. Schmauch B, Herent P, Jehanno P, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. *Diagn Interv Imaging* 2019;100:227-33. DOI
16. Mokrane FZ, Lu L, Vavasseur 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. DOI PubMed
17. Jansen MJA, Kuijf HJ, Veldhuis WB, Wessels FJ, Viergever MA, Pluim JPW. Automatic classification of focal liver lesions based on MRI and risk factors. *PLoS One* 2019;14:e0217053. DOI PubMed PMC
18. Zhang F, Yang J, Nezami N, et al. Liver tissue classification using an auto-context-based deep neural network with a multi-phase training framework. In: Bai W, Sanroma G, Wu G, Munsell BC, Zhan Y, Coupé P, editors. patch-based techniques in medical imaging. cham: springer international publishing; 2018. pp.59-66. DOI PubMed PMC
19. Kiani A, Uyumazturk B, Rajpurkar P, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *NPJ Digit Med* 2020;3:23. DOI
20. Liao H, Long Y, Han R, et al. Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma. *Clin Transl Med* 2020;10:e102. DOI PubMed PMC
21. Sun SW, Xu X, Liu QP, et al. LiSNet: An artificial intelligence -based tool for liver imaging staging of hepatocellular carcinoma aggressiveness. *Med Phys* 2022;49:6903-13. DOI PubMed

22. Noh B, Park YM, Kwon Y, et al. Machine learning-based survival rate prediction of Korean hepatocellular carcinoma patients using multi-center data. *BMC Gastroenterol* 2022;22:85. DOI PubMed PMC
23. Simsek C, Can Guven D, Koray Sahin T, et al. Artificial intelligence method to predict overall survival of hepatocellular carcinoma. *Hepatol Forum* 2021;2:64-8. DOI PubMed PMC
24. Mähringer-Kunz A, Wagner F, Hahn F, et al. Predicting survival after transarterial chemoembolization for hepatocellular carcinoma using a neural network: a pilot study. *Liver Int* 2020;40:694-703. DOI PubMed
25. Saillard C, Schmauch B, Laifa O, et al. Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides. *Journal of Hepatology* 2020;73:S381. DOI
26. Liang JD, Ping XO, Tseng YJ, Huang GT, Lai F, Yang PM. Recurrence predictive models for patients with hepatocellular carcinoma after radiofrequency ablation using support vector machines with feature selection methods. *Meth Pro* 2014;117:425-34. DOI PubMed
27. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electron Markets* 2021;31:685-95. DOI
28. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther* 2020;51:675-87. DOI PubMed PMC
29. Ghahramani Z. Unsupervised learning. in: bousquet o, von luxburg u, rätsch g, editors. advanced lectures on machine learning. berlin: springer berlin heidelberg; 2004. pp.72-112. DOI
30. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci* 2021;2:420. DOI PubMed PMC
31. Han SH, Kim KW, Kim S, Youn YC. Artificial neural network: understanding the basic concepts without mathematics. *Dement Neurocogn Disord* 2018;17:83-9. DOI PubMed PMC
32. Pai A. CNN vs. RNN vs. ANN – analyzing 3 types of neural networks in deep learning. Available from: https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/ [Last accessed on 23 Mar 2023].
33. Indolia S, Goswami AK, Mishra S, Asopa P. Conceptual understanding of convolutional neural network- a deep learning approach. *Procedia Computer Science* 2018;132:679-88. DOI
34. Marhon SA, Cameron CJF, Kremer SC. Recurrent neural networks. in: bianchini m, maggini m, jain lc, editors. handbook on neural information processing. berlin: springer berlin heidelberg; 2013.p.29-65. DOI
35. Savage N. Breaking into the black box of artificial intelligence. *Nature* 2022. DOI PubMed
36. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. *Radiographics* 2017;37:2113-31. DOI PubMed
37. Azer SA. Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: a systematic review. *World J Gastrointest Oncol* 2019;11:1218-30. DOI PubMed PMC
38. Liu X, Song JL, Wang SH, Zhao JW, Chen YQ. Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification. *Sensors* 2017;17:149. DOI PubMed PMC
39. Książek W, Abdar M, Acharya UR, Pławiak P. A novel machine learning approach for early detection of hepatocellular carcinoma patients. *Csri* 2019;54:116-27. DOI
40. Brehar R, Mitrea DA, Vancea F, et al. Comparison of deep-learning and conventional machine-learning methods for the automatic recognition of the hepatocellular carcinoma areas from ultrasound images. *Sensors* 2020;20:3085. DOI PubMed PMC
41. Guo LH, Wang D, Qian YY, et al. A two-stage multi-view learning framework based computer-aided diagnosis of liver tumors with contrast enhanced ultrasound images. *Clin Hemorheol Microcirc* 2018;69:343-54. DOI PubMed
42. Yang Q, Wei J, Hao X, et al. Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: a multicentre study. *EBioMedicine* 2020;56:102777. DOI PubMed PMC
43. 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. DOI PubMed PMC
44. Hassan TM, Elmogy M, Sallam E. Diagnosis of focal liver diseases based on deep learning technique for ultrasound images. *Arab J Sci Eng* 2017;42:3127-40. DOI
45. 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* 2020;45:2688-97. DOI PubMed
46. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced ct: a preliminary study. *Radiology* 2018;286:887-96. DOI PubMed
47. Hamm CA, Wang CJ, Savic LJ, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol* 2019;29:3338-47. DOI PubMed PMC
48. Preis O, Blake MA, Scott JA. Neural network evaluation of PET scans of the liver: a potentially useful adjunct in clinical interpretation. *Radiology* 2011;258:714-21. DOI PubMed
49. Tunissiolli NM, Castanhole-Nunes MMU, Biselli-Chicote PM, et al. Hepatocellular carcinoma: a comprehensive review of biomarkers, clinical aspects, and therapy. *Asian Pac J Cancer Prev* 2017;18:863-72. DOI PubMed PMC
50. Yeom SK, Lee CH, Cha SH, Park CM. Prediction of liver cirrhosis, using diagnostic imaging tools. *World J Hepatol* 2015;7:2069-79. DOI PubMed PMC
51. association for the study of the liver. electronic address: easloffice@easloffice.eu, European association for the study of the liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 2018;69:182-236. DOI
52. Tanaka H. Current role of ultrasound in the diagnosis of hepatocellular carcinoma. *J Med Ultrason* 2020;47:239-55. DOI PubMed

PMC

53. Bhogadi Y, Brown E, Lee SY. Contrast-enhanced ultrasound in the diagnosis of infiltrative hepatocellular carcinoma: a report of three cases. *Radiol Case Rep* 2021;16:448-56. DOI PubMed PMC
54. Shen J, Wen J, Li C, et al. The prognostic value of microvascular invasion in early-intermediate stage hepatocelluar carcinoma: a propensity score matching analysis. *BMC Cancer* 2018;18:278. DOI PubMed PMC
55. Jiang YQ, Cao SE, Cao S, et al. Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning. *J Cancer Res Clin Oncol* 2021;147:821-33. DOI PubMed PMC
56. Zhang Y, Lv X, Qiu J, et al. Deep learning with 3D convolutional neural network for noninvasive prediction of microvascular invasion in hepatocellular carcinoma. *J Magn Reson Imaging* 2021;54:134-43. DOI
57. Liu F, Liu D, Wang K, et al. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. *Liver Cancer* 2020;9:397-413. DOI
58. Zhang L, Xia W, Yan ZP, et al. Deep learning predicts overall survival of patients with unresectable hepatocellular carcinoma treated by transarterial chemoembolization plus sorafenib. *Front Oncol* 2020;10:593292. DOI PubMed PMC
59. Gotra A, Sivakumaran L, Chartrand G, et al. Liver segmentation: indications, techniques and future directions. *Insights Imaging* 2017;8:377-92. DOI PubMed PMC
60. Al-kababji A, Bensaali F, Dakua SP, Himeur Y. Automated liver tissues delineation techniques: a systematic survey on machine learning current trends and future orientations. *Eng Appl Artif Intell* 2023;117:105532. DOI
61. Liang F, Qian P, Su KH, et al. Abdominal, multi-organ, auto-contouring method for online adaptive magnetic resonance guided radiotherapy: An intelligent, multi-level fusion approach. *Artif Intell Med* 2018;90:34-41. DOI PubMed
62. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense v-networks. *IEEE Trans Med Imaging* 2018;37:1822-34. DOI PubMed PMC
63. Llovet JM, Castet F, Heikenwalder M, et al. Immunotherapies for hepatocellular carcinoma. *Nat Rev Clin Oncol* 2022;19:151-72. DOI
64. Muhammed A, D'Alessio A, Enica A, et al. Predictive biomarkers of response to immune checkpoint inhibitors in hepatocellular carcinoma. *Expert Rev Mol Diagn* 2022;22:253-64. DOI PubMed
65. He Y, Lu M, Che J, Chu Q, Zhang P, Chen Y. Biomarkers and future perspectives for hepatocellular carcinoma immunotherapy. *Front Oncol* 2021;11:716844. DOI
66. Ji GW, Zhu FP, Xu Q, et al. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. *EBioMedicine* 2019;50:156-65. DOI PubMed PMC
67. Saillard C, Schmauch B, Laifa O, et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. *Hepatology* 2020;72:2000-13. DOI
68. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 2000;31:864-71. DOI PubMed
69. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001;33:464-70. DOI
70. Bertsimas D, Kung J, Trichakis N, Wang Y, Hirose R, Vagefi PA. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. *Am J Transplant* 2019;19:1109-18. DOI PubMed
71. Yu YD, Lee KS, Man Kim J, et al; Korean organ transplantation registry study group. Artificial intelligence for predicting survival following deceased donor liver transplantation: Retrospective multi-center study. *Int J Surg* 2022;105:106838. DOI PubMed
72. Briceño J, Cruz-Ramírez M, Prieto M, et al. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. *J Hepatol* 2014;61:1020-8. DOI PubMed
73. Guijo-Rubio D, Briceño J, Gutiérrez PA, Ayllón MD, Ciria R, Hervás-Martínez C. Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation. *PLoS One* 2021;16:e0252068. DOI PubMed PMC
74. Peng J, Kang S, Ning Z, et al. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur Radiol* 2020;30:413-24. DOI PubMed PMC
75. Morshid A, Elsayes KM, Khalaf AM, et al. A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. *Radiol Artif Intell* 2019;1:e180021. DOI PubMed PMC
76. Liu D, Liu F, Xie X, et al. Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. *Eur Radiol* 2020;30:2365-76. DOI PubMed
77. Kant I. Critique of Pure Reason. Available from: https://play.google.com/store/books [Last accessed on 23 Mar 2023].
78. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583-9. DOI PubMed PMC
79. Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. *Cell* 2020;180:688-702.e13. DOI
80. Marwaha JS, Kvedar JC. Crossing the chasm from model performance to clinical impact: the need to improve implementation and evaluation of AI. *NPJ Digit Med* 2022;5:25. DOI PubMed PMC
81. Jiang L, Wu Z, Xu X, et al. Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. *J Int Med Res* 2021;49:3000605211000157. DOI PubMed PMC
82. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. *JMIR Med Educ* 2019;5:e13930. DOI PubMed PMC
83. FDA. Current good manufacturing practice (CGMP) regulations. Available from: https://www.fda.gov/drugs/pharmaceutical-quality-

resources/current-good-manufacturing-practice-cgmp-regulations [Last accessed on 23 Mar 2023].

84. Dockès J, Varoquaux G, Poline JB. Preventing dataset shift from breaking machine-learning biomarkers. *Gigascience* 2021:10. DOI PubMed PMC
85. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications* 2017;73:220-39. DOI
86. Storkey AJ. When training and test sets are different: characterising learning transfer. Available from:https://homepages.inf.ed.ac.uk/amos/publications/Storkey2009TrainingTestDifferent.pdf [Last accessed on 23 Mar 2023].
87. Shah NH, Milstein A, Bagley PhD SC. Making machine learning models clinically useful. *JAMA* 2019;322:1351-2. DOI PubMed
88. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016;18:e323. DOI PubMed PMC
89. Mathrani A, Susnjak T, Ramaswami G, Barczak A. Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. *Comput Educ* 2021;2:100060. DOI
90. Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. *Can J Cardiol* 2022;38:204-13. DOI PubMed
91. Suzuki K, Reyes M, Syeda-Mahmood T, Konukoglu E, Glocker B, Wiest R, et al. Interpretability of machine intelligence in medical image computing and multimodal learning for clinical decision support: second international workshop, iMIMIC 2019, and 9th international workshop, ML-CDS 2019, held in conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings. Available from: https://play.google.com/store/books/details?id=Vvm4DwAAQBAJ [Last accessed on 23 Mar 2023].
92. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019;1:206-15. DOI PubMed PMC