Review

JOURNAL OF HEPATOLOGY

Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma

Julien Calderaro[1,2], Tobias Paul Seraphin[3], Tom Luedde[3], Tracey G. Simon[4,5,*]

Keywords: Artificial intelligence; Machine learning; Deep learning; Liver cancer.

Received 7 November 2021; received in revised form 26 December 2021; accepted 14 January 2022

Summary

Hepatocellular carcinoma (HCC) currently represents the fifth most common malignancy and the third-leading cause of cancer-related death worldwide, with incidence and mortality rates that are increasing. Recently, artificial intelligence (AI) has emerged as a unique opportunity to improve the full spectrum of HCC clinical care, by improving HCC risk prediction, diagnosis, and prognostication. AI approaches include computational search algorithms, machine learning (ML) and deep learning (DL) models. ML consists of a computer running repeated iterations of models, in order to progressively improve performance of a specific task, such as classifying an outcome. DL models are a subtype of ML, based on neural network structures that are inspired by the neuroanatomy of the human brain. A growing body of recent data now apply DL models to diverse data sources – including electronic health record data, imaging modalities, histopathology and molecular biomarkers – to improve the accuracy of HCC risk prediction, detection and prediction of treatment response. Despite the promise of these early results, future research is still needed to standardise AI data, and to improve both the generalisability and interpretability of results. If such challenges can be overcome, AI has the potential to profoundly change the way in which care is provided to patients with or at risk of HCC.

[1]*Assistance Publique-Hôpitaux de Paris, Henri Mondor University Hospital, Department of Pathology, Créteil, France;* [2]*Inserm U955 and Univ Paris Est Creteil, INSERM, IMRB, 94010, Creteil, France;* [3]*Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty at Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany;* [4]*Liver Center, Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA;* [5]*Clinical and Translational Epidemiology Unit (CTEU), Massachusetts General Hospital, Boston, MA, USA*

* Corresponding author. Address: Liver Center, Division of Gastroenterology, Massachusetts General Hospital, 55 Fruit Street, Wang 5th Floor Boston, MA 02114, USA; Tel.: 617-724-2401, fax: 617-724-5997.

E-mail address: tgsimon@mgh.harvard.edu (T.G. Simon).

https://doi.org/10.1016/j.jhep.2022.01.014

Introduction and definitions

With a global incidence of approximately 500,000 cases per year, hepatocellular carcinoma (HCC) represents the fifth most common malignancy and the third-leading cause of cancer-related death worldwide.[1,2] The vast majority of HCC tumours arise on a background of cirrhosis, which in turn is most commonly caused by non-alcoholic fatty liver disease (NAFLD), alcohol-related liver disease, or HBV/HCV infection. Despite recent advances in treatment, including the use of atezolizumab plus bevacizumab for unresectable HCC, prognosis remains poor, with a 5-year survival rate of just 15%, due to delays in diagnosis and the limited efficacy of existing therapies.[3,4] While liver transplantation can be curative for HCC in selected cases, this represents a limited and resource-intensive solution, and the vast majority of patients are not eligible for transplantation. Thus, identifying novel approaches to improve the early diagnosis of HCC and to predict therapeutic response and survival among patients with established HCC is of paramount importance.

Owing to the broad heterogeneity in HCC risk factors and pathogenesis, established strategies for prediction and prognostication are still limited. Recently, artificial intelligence (AI) has emerged as a unique opportunity to improve the full spectrum of HCC clinical care, by: i) improving the prediction of future HCC risk in patients with established liver disease; ii) improving the accuracy of HCC diagnosis in patients undergoing surveillance imaging or liver biopsies; and iii) improving prognostication in patients with established HCC.

AI is a broad field that includes computational search algorithms, machine learning (ML) and deep learning (DL) models (Fig. 1). ML consists of a computer running repeated iterations of models in order to progressively improve performance of a specific task, such as classifying an outcome. ML models are designed to improve with time, by incorporating additional input training data and thereby optimising the parameters of an algorithm. With time and training, the desired output becomes increasingly accurate. Based on how the training process is conducted, ML may be classified as supervised or unsupervised. Supervised ML algorithms perform training on a dataset that is labelled in relation to the class of interest, and this label is available to the algorithm while the model is being created, trained, and optimised. In contrast, unsupervised ML involves training on a dataset that lacks class labels, yielding clusters of output data that subsequently require additional interpretation.

DL represents a subtype of ML models which are constructed using neural networks (NNs) inspired by the neuroanatomy of the human brain. NNs consist of a network of interconnected computing units – termed "neurons" – that are organised in layers, such that signals travel from the first layer

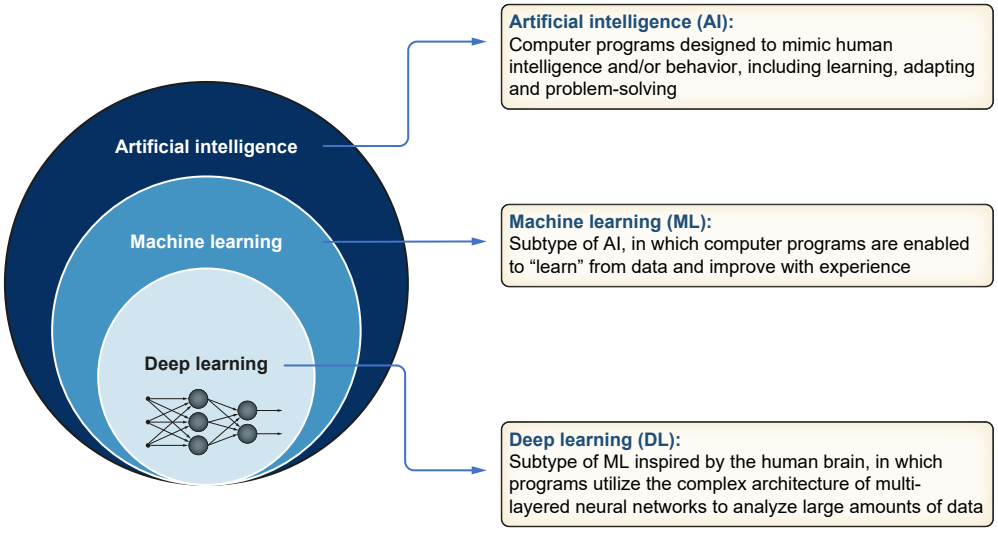

Fig. 1. Definitions of artificial intelligence (AI), machine learning (ML) and deep learning (DL).

(*i.e.* input data) to the last layer (*i.e.* output data) after passing through multiple, intervening hidden layers (Fig. 2). To train an NN, data are divided into a training set and a testing set. The training set characterises the architecture of the network and defines and adjusts the weights between neurons, in order to improve classification of the desired output. The testing set then evaluates the utility of the NN for identifying or predicting that output. This validation can be conducted internally or externally. Internal validation is commonly performed by k-fold cross validation within one dataset, by splitting that dataset into *k* parts and then training *k* times on *k-1* parts, and then subsequently testing on the remaining part of the dataset. External validation is typically considered more robust, as it demonstrates model generalisability across populations.

Current limitations of DL approaches include overfitting of data, limited 'explainability' of data, and the possibility of poor generalisability, due to the inherent reliance of DL models on the size and diversity of their training dataset. In this review, we will outline the rapidly evolving role and challenges for AI in the prediction, diagnosis, and prognostication of HCC.

AI for predicting incident HCC

Several previous case-control and cohort studies have developed predictive models for the development of HCC using clinical, demographic and/or laboratory risk factors, selected using conventional statistical approaches. However, these models have largely been criticised for their limited generalisability, modest accuracy, and lack of broad external validity. Moreover, HCC risk is notoriously

Key point

Due to the broad heterogeneity in risk factors for HCC and the lack of established strategies for prediction or prognostication, AI has recently emerged as a unique opportunity to improve the full spectrum of HCC clinical care.

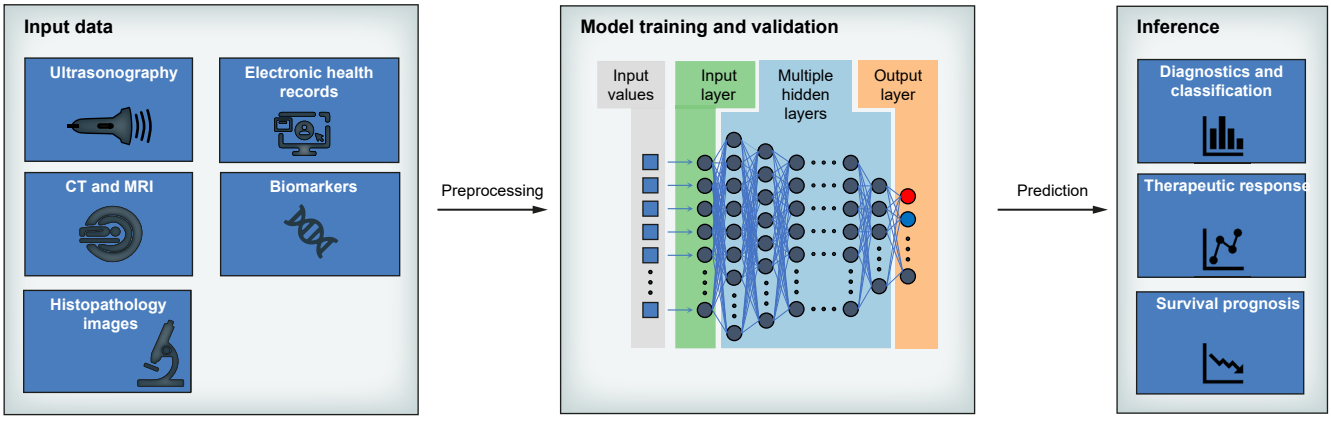

Fig. 2. General concept of pipelines using neural networks. Different input data are pre-processed in such a way that they can be used as input values for the training of a neural network. The neural network consists of one input layer, multiple hidden convolutional and/or multiple fully connected layers extracting features from the input data, and one output layer with nodes that refer to different labels. These networks can then – among others – be used to classify data or to predict therapeutic response or prognosis.

Table 1. Selected prior studies utilising AI to predict incident HCC.

Author, year	Population	AI classifier	Validation method	HCC cases (n)/total cohort (n)	Accuracy	Sensitivity/specificity	Improvement over traditional methods
Singal AG, 2013	Cirrhosis	Random forest	External validation (HALT-C trial)	Training: 41/442 Validation: 88/1,050	C-statistic 0.64	80.5% (57.9% in the training set)/80.7% (46.8% in the validation set)	Outperformed HALT-C model for predicting HCC (IDI = 0.01, p = 0.04; NRI = 0.39, p <0.001)
Reddy R, 2017	Cirrhosis	Artificial neural network	n.a.	Training: 165/6,092	AUROC 0.96	83.6%/99.9%	n.a.
Ioannou GN, 2019	Chronic HCV	Recurrent neural network	n.a.	Training: 10,741/ 48,151	AUROC 0.759	Proportion testing positive at 90%: sensitivity = 0.663	Outperformed conventional logistic regression models
Nam JY, 2020	Cirrhosis (HBV) on entecavir	Deep neural network	External validation	Training: 86/424 Validation: n = 316	C-statistic 0.719 (training); 0.782 (validation)	n.a.	Outperformed numerous conventional algorithms (PAGE-B, CU-HCC, ADRESS-HCC and THRI; all p <0.001)
An C, 2021	General population (Korea)	Random forest	Internal validation	Training: 1,799/ 331,694 Validation: 390/ 85,692	C-statistic (validation) 0.857; AUROC 0.873	71.8%/88.4%	n.a.

AI, artificial intelligence; AUROC, area under the receiver-operating characteristic curve; HCC, hepatocellular carcinoma; IDI, integrated discrimination index; NRI, net reclassification index.

challenging to model because this risk can fluctuate widely in an individual over time, and such non-linear changes are difficult to estimate using rigid, conventional regression models. Recently, the rapid expansion of available electronic health record (EHR) data has provided an opportunity to leverage large-scale, longitudinal data elements for automatic feature selection over long-term follow-up, and thereby improve HCC risk prediction. To that end, several recent studies have applied AI approaches to longitudinal EHR data to improve prediction of incident HCC (Table 1). For example, in 2013, a supervised ML algorithm was found to have a c-statistic of 0.64 for predicting incident HCC in patients with cirrhosis of any aetiology, and this significantly outperformed a conventional system for HCC risk prediction.[5] More recently, another model developed in patients with chronic hepatitis C infection in the U.S. Veterans Affairs cohort demonstrated an AUROC of 0.759 for incident HCC.[6] In all cases, the models constructed by AI approaches significantly outperformed traditional regression models.

It has been posited that improved HCC risk prediction models leveraging AI techniques could be used to personalise HCC surveillance strategies by improving risk stratification of patients with chronic liver disease. For example, Ioannou and colleagues found that targeting patients with the uppermost 51% of their NN-derived HCC risk score would include 80% of patients who would develop HCC within the subsequent 3 years.[6] Such an approach could be useful in resource-limited settings that do not have sufficient capacity for regular HCC surveillance in all at-risk patients. However, to date, the clinical utility of this and other AI-based scores for predicting risk of HCC is unclear, particularly as these data have limited generalisability, given their reliance on the size and diversity of the training dataset.

AI for diagnosing HCC: radiomics, histopathology and biomarkers

Numerous studies have tested the utility of AI for accurately detecting existing HCC, based on imaging modalities or biomarkers.

Radiomics: ultrasound

Current clinical guidelines recommend regular B-mode abdominal ultrasound surveillance for the identification of HCC in patients with cirrhosis.[7–9] However, ultrasound has several well-described limitations when it comes to detecting focal liver lesions, including a high degree of dependence on operator experience, equipment quality, and patient body habitus, among others. For detection of HCC, the sensitivity of B-mode ultrasound is only 46-63%.[9–11] To address this, several recent studies have tested the ability of AI frameworks to improve the diagnostic accuracy of ultrasound in this setting.

Schmauch and colleagues designed a supervised DL model, using a training dataset of 367 ultrasound images together with their corresponding radiological reports, that could identify liver lesions as benign or malignant with a mean AUROC of 0.93 and 0.92, respectively.[12] More recently, Yang and colleagues developed and externally validated a deep convolutional neural network (DCNN), using a large, multicentre, ultrasound imaging database from 13 hospital systems. The final model demonstrated an AUROC of 0.92 for distinguishing benign from malignant liver lesions, and showed comparable a) performance to the judgment of clinical radiologists (diagnostic accuracy, both 76.0%) and b) accuracy to contrast-enhanced CT (diagnostic accuracy, both 84.7%) that was only slightly inferior to MRI (87.9%).[13]

Similar approaches have also been applied to contrast-enhanced ultrasound (CEUS) imaging for the detection of HCC. For example, Guo and colleagues recently demonstrated that a DL algorithm applied to liver lesions seen by CEUS could increase the sensitivity, specificity, and overall accuracy of CEUS for detecting HCC.[14] Others have used AI to apply additional pattern recognition classifiers to CEUS DCNN algorithms, to improve diagnosis of indeterminate focal liver lesions.[15] However, to date, most prior CEUS studies have had small sample sizes and lacked standardised imaging data or external validation cohorts (to confirm the generalisability of models across populations).

CT and MRI

Another rapidly growing area of research is focused on improved characterisation of indeterminate liver lesions. In clinical practice, when an abdominal ultrasound shows a new liver lesion, a patient is typically referred for further imaging, with contrast-enhanced CT or MRI. Based on the fulfilment of specific radiologic criteria, certain liver lesions may be considered as having pathognomonic features of HCC, and thus do not require liver biopsy for further histological confirmation. However, liver nodules imaged by CT or MRI often demonstrate indeterminate features, for which current recommendations include either liver biopsy or close interval follow-up with serial imaging.[7,9] This practice is sub-optimal, resulting in numerous imaging studies, patient stress, and the potential for delayed diagnoses of liver cancer. For this reason, a growing body of recent literature has explored AI approaches to improve risk stratification of indeterminate liver lesions, to facilitate earlier and more accurate detection of HCC.

In an early study focused on this issue, Preis and colleagues developed a NN to assess focal liver lesions identified by ^{18}F-FDG-PET/CT (fluorine-18 fluorodeoxyglucose positron emission tomography/CT) evaluations, together with patient demographics and clinical characteristics of 98 patients; their model had an AUROC of 0.896 for the identification of focal liver lesions, outperforming the results of blinded radiologists.[16] Mokrane and colleagues conducted a small retrospective study (n = 178) of patients with cirrhosis and indeterminate liver lesions, for whom diagnostic liver biopsy was recommended. Applying DL approaches, the authors constructed a radiomics signature based on 13,920 CT imaging classifiers, that achieved an AUROC of 0.70 for distinguishing HCC from non-HCC lesions. Importantly, the authors demonstrated that the signature was not influenced by segmentation or by contrast enhancement, which adds to its putative generalisability.[17] Another retrospective study, by Yasaka *et al.* (n = 460), utilised CT imaging classifiers from 3 phases (non-contrast-enhanced, arterial, and delayed) to construct a 3-layer CNN for distinguishing a) HCC and non-HCC liver cancers from (b) indeterminate liver lesions, haemangiomas and cysts; their CNN had a diagnostic accuracy of 0.84 with a median AUROC of 0.92.[18] More recently, Shi and colleagues compared the performance of a triple-phase contrast-enhanced CT protocol coupled with a DL model, to a four-phase CT protocol, for distinguishing HCC from other focal liver lesions.[19] The authors found that a DL model combined with triple-phase CT protocol without pre-contrast yielded similar diagnostic accuracy (85.6%) to a four-phase protocol (83.3%; $p = 0.765$). These findings suggest that reducing a patient's radiation dose with a triple-phase CT protocol may not compromise accuracy, and thereby brings the field one step closer to optimising CT protocols for the accurate classification of liver lesions.

Given the wide variability of radiographic features of the liver and liver lesions, manual segmentation for radiomics-based assessments of HCC is both difficult and time-consuming. In 2017, the Liver Tumor Segmentation (LiTS) Challenge called upon investigators to develop AI-based algorithms that could automatically segment liver tumours, using a multinational dataset of 200 CT scans (130 training, 70 validation scans).[20,21] All of the top-scoring automatic methods used fully convolutional NNs that separately segmented the liver and liver tumours. Segmentation quality was evaluated using Dice scores, and the best-scoring algorithm achieved a Dice score of 0.96, whereas for liver tumour segmentation the best algorithm achieved Dice scores between 0.67 and 0.70. While these findings are promising, there was notable variability in both the imaging characteristics of liver tumours and in their annotation, underscoring the need for universal, standardised methods for liver tumour segmentation.[20,21]

To date, AI has been applied less frequently to MRI imaging of HCC tumours, and given the technical difficulty and expense associated with manually designing MRI features, the majority of published studies have been conducted in relatively small populations. Nevertheless, a prior

study combined clinical data with MRI-based classifiers to distinguish HCC from metastases and from liver adenomas, cysts or haemangiomas, and demonstrated a sensitivity of 0.73 for identifying HCC, albeit with a specificity of just 0.56.[22] Additionally, Hamm *et al.* developed a NN algorithm that successfully classified MRI liver lesions with a sensitivity of 92%, a specificity of 98%, and an overall accuracy of 92%.[23] Zhang and colleagues tested an automated approach to segmentation of multi-parameter MR images in 20 patients with HCC, and demonstrated the feasibility of bypassing the time-consuming process of manually designing MRI-based features.[24]

Key point

AI reflects a broad and rapidly evolving field that includes ML and DL computational algorithms, which are iteratively repeated, in order to progressively improve model performance and classification over time.

More recently, Zhen *et al.* used CNNs to develop a novel DL system that incorporated enhanced MR images, unenhanced MR images and both structured and unstructured clinical data, from 1,210 patients with liver tumours, and an external validation set (n = 201).[25] This DL system demonstrated excellent performance for classifying liver tumours – including HCC – with sensitivity and specificity on a par with that observed for experienced radiologists. Importantly, this DL model also showed excellent performance when combining unenhanced MR imaging with clinical data, suggesting that, with further validation, these models may permit patients to avoid contrast-related complications of MRI. Finally, Wang and colleagues recently described a DL model designed to address the limited interpretability of AI-based radiomics assessments of HCC.[26] This innovative model provides feedback on the relative importance of various radiological input features, and thereby serves as an important proof of concept, demonstrating that "interpretable" DL models could one day be used to improve standardised HCC reporting systems and thereby clinical outcomes.

To date, published AI algorithms for radiomics assessments of HCC share important limitations, including relatively small input datasets, lack of sufficiently large or diverse cohorts for robust external validation and lack of standardisation of methods or analytical tools. It will be important to define the utility of AI-based prediction tools in prospective cohorts, and in pooled, large-scale and diverse populations.

Histopathology

Histopathology is a cornerstone in the management of many liver diseases, including autoimmune hepatitis and non-alcoholic steatohepatitis (for grading and staging). Although non-invasive criteria allow for the diagnosis of HCC in particular clinical settings, the histological examination of tumour samples is often required for masses with atypical features on imaging or to rule out a diagnosis of benign primary liver tumour, cholangiocarcinoma or even metastasis. However, precise histopathological characterisation of liver tumours can often prove challenging for hepatopathologists, and significant inter-observer disagreement may be observed. To address this, several recent studies have applied AI to assist with the diagnosis of liver tumours. Using 2 large data sets of H&E-stained digital slides, Liao *et al.* used a CNN to distinguish HCC from adjacent normal tissues, with AUCs above 0.90.[27] Kiana *et al.* developed a tool able to classify image patches as HCC or cholangiocarcinoma. The model reached an accuracy of 0.88 on the validation set and, interestingly, the authors observed that the combination of the model and the pathologist outperformed both the model alone and the pathologist alone, suggesting that AI tools should be used to augment, rather than replace, the conventional histological diagnosis. They also showed how an incorrect prediction may negatively impact the final diagnosis made by pathologists, underscoring the need to be cautious with AI models aimed at automating diagnosis.[28]

It has been widely demonstrated that the histological appearance of human cancers, including HCC, contain a massive amount of information related to their underlying molecular alterations and/or to patient prognosis.[29–31] In this line, Wang *et al.* trained a multitask DL NN for automated single-cell segmentation and classification on digital slides. This approach allowed the authors to extract quantitative image features related to individual cells as well as spatial relationships between neoplastic cells and infiltrating lymphocytes. Unsupervised consensus clustering of these features led to the identification of 3 subtypes associated with particular somatic genomic alterations and molecular pathways.[32] Another study showed that DL could predict a subset of recurrent HCC genetic defects with AUCs ranging from 0.71 to 0.89.[33]

Recent pioneering studies have thus aimed to predict molecular signatures/alterations predictive of response to systemic therapies, by processing digital slides through NNs. In gastrointestinal cancers, for example, high performance is achieved for the prediction of microsatellite instability, a feature strongly associated with sensitivity to immunomodulating therapies.[34] Two other pan-cancer studies also demonstrated that NN models were able to predict a wide range of molecular alterations or signatures, some of which are related to response to particular systemic therapies.[35,36] For HCC, no molecular feature is currently used to predict response to the systemic therapies available for patients with advanced disease. However, Sangro *et al.* recently reported that responses to the anti-programmed death 1 receptor (PD1) antibody nivolumab were more frequently observed in patients with tumours showing overexpression of particular immune gene signatures.[37] This was further confirmed by Haber *et al.*, who also observed increased sensitivity to immunotherapy in HCCs in which interferon gamma and gene sets associated with antigen presentation were

upregulated.[38] Immune cells are easily identified by DCNNs, and it is likely that DL will be able to predict this type of gene expression profile.

Most of these different studies share the same limitations, including the limited number of patients, sensitivity to staining protocols and lack of prospective validation. The standardisation of slide encoding and processing will also be key to enable comparisons of model performance. Finally, it will be critical to determine how predictions are impacted by artifacts such as tissue folds or stains. Automated quality control of slides may help to overcome these issues.

Molecular biology and biomarkers

The past 20 years have witnessed an explosion in the availability of large, complex data sets with genomic and molecular data from bulk tissues and from single cells. Consequently, AI algorithms leveraging integrative multiomics approaches have also been designed to improve the detection and characterisation of HCC tumours. Such integrated algorithms have shown promise for informing disease diagnosis and staging, and for the prediction of disease recurrence and therapeutic response.[39,40]

As an example, integrated multiomics analyses are increasingly used to assess individual variation in key patterns of hepatic gene expression, and to define intratumoural heterogeneity.[41] Zeng and colleagues constructed a DL model based on RNA-sequencing (RNA-seq)-defined samples, and used those classified features to construct gene expression signatures for cancer.[42] The DL-defined autoencoder was found to outperform numerous traditional analytical approaches based on principal component analysis or top varying genes.

In another study of HCC samples, Chaudhary and colleagues applied supervised and unsupervised DL approaches to RNA-seq, miRNA-seq and DNA methylation data, and identified 2 distinct HCC subpopulations with significant survival differences, with a C-statistic of 0.68 in the training dataset and 0.67-0.82 in 5 external validation sets.[43] This algorithm has subsequently been applied to external HCC cohorts (n = 1,494), revealing consensus driver genes linked to HCC survival.[44] Future work will need to demonstrate the utility of those signatures for informing therapeutic decision making.

Finally, single-cell RNA-seq technologies now permit thousands of single cells to be profiled simultaneously and in an unbiased fashion, which holds great promise for powerful DL approaches. Single-cell RNA-seq permits the identification of unique cellular subpopulations and their transcriptomic profiles, as well as complex gene regulatory networks.[45] Within the liver, single-cell RNA-seq has been used to more comprehensively elucidate the cellular transcriptomes of non-alcoholic steatohepatitis and cirrhosis, and to identify novel cell types and cell-cell interactions.[46–48] In HCC, it has permitted identification of new subsets of tumour-infiltrating lymphocytes, including clonally expanded exhausted CD8+ T cells and regulatory T cells, and tumour-associated macrophages.[49,50] Collectively, these findings are helping to uncover the immunological landscape of chronic liver disease and HCC, with unprecedented resolution.

The field of single-cell RNA-seq is still in its infancy and key challenges remain, including the variation between methods in terms of data quality and sensitivity, as well as the noisiness and incompleteness of generated data.[51–53] Specifically, low-abundance data is frequently lost, rendering an expressed transcript undetectable (a phenomenon called, “dropout”).[54] On the other hand, unnecessary amplification of noise risks artificially accentuating the significance of less relevant pathways.[45] Several DL-based tools are currently available to address these issues in single-cell RNA-seq datasets, including DeepImpute and SAUCIE, which apply node/gene interaction structures, as well as adaptations of generative adversarial networks, which can generate single-cell RNA-seq data and ascertain individual cell types using NNs.[55–57] It is hoped that further improvements in DL algorithms will help to improve the validity of single-cell RNA-seq datasets through imputation, by “denoising” with an autoencoder that predicts genes’ mean, standard deviation and likelihood of dropout, or by streamlining downstream data analyses.[55,58]

New technologies incorporating DL have recently been developed to integrate single-cell RNA-seq profiling with epigenetic and proteomic assays, in order to more comprehensively profile individual cells.[59–61] Such multi-omics approaches have tremendous potential utility for uncovering novel biomarkers and therapeutic targets in HCC. However, universal, standardised methods and protocols must first be established, and much larger datasets will be needed, given that the accuracy of DL algorithms depends upon the size and quality of input data. This, in turn, will require collaboration between investigators and the sharing of algorithms, approaches and raw datasets.

AI for prognostication in established HCC

The development of robust prognostic scoring systems is key to improve patient risk stratification and to plan clinical trials testing neoadjuvant or adjuvant therapies (see Table 2). A DL algorithm based on a residual NN architecture was recently developed in a Korean multicentre study to predict HCC recurrence after transplantation. The features included age, tumour size, and serum levels of alpha-fetoprotein and PIVKA-II (prothrombin induced by vitamin K absence or antagonist-II); the authors showed the advantages of their model (MoRAL-AI, assessed by C-indices) in their external

Key point

A growing body of research has applied AI approaches to improve HCC risk prediction, and to more accurately detect and risk stratify existing HCC tumours, based on EHR data, radiomics approaches, and molecular or histopathological biomarkers.

Table 2. Selected prior studies utilising AI for HCC prognostication.

Author, Year	HCC cases (n)	AI algorithm	Validation method	Input data	Test statistics	Highlight
Abajian A, 2018	36	Logistic regression, random forest	Internal leave-one-out cross validation	MR images and clinical data	Accuracy: 78% Sensitivity: 62.5% Specificity: 82.1%	Prediction of TACE response Successful implementa-tion of AI methods for the combination of clinical and imaging data
Ji GW, 2019	Training: 210 Validation: 107 internal 153 external	RSF/MRMR	External Validation	CT images and clinical data	C-statistic: 0.73	Prediction of HCC recur-rence after resection; outperformed conven-tional outcome prediction scores, *e.g.* BCLC stage
Nam JY, 2020	Training: 349 Validation: 214	Residual neural network	External validation	Clinical data	C-statistic: 0.75 Sensitivity: 76% Specificity: 46%	Prediction of HCC Recur-rence after LT; out-perfor-med conventional recurr-ence prediction scores, *e.g.* Milan criteria
Saillard C, 2020	Training: 194 Validation: 328	Artificial neural network	External validation	Digitised histopathology slides	C-statistic: 0.78	Survival prediction after HCC resection; Out-perfor-med conventional clinical, biological or pathological parameters
Peng J, 2020	Training: 562 Validation: 227	Residual convolutional neural network	External validation	CT images	AUC: >0.95	Prediction of TACE response First study to predict complete/partial response and stable/progressive disease showing good accuracy
Oezdemir I, 2020	36	Distance weighted discrimination method	Internal leave-one-out cross validation	Contrast-enhanced ultra-sound images	Accuracy: 86% Sensitivity: 89% Specificity: 82%	Prediction of TACE response First study providing proof of concept using AI methods with ultrasono-graphy images

AI, artificial intelligence; AUC, area under the curve; HCC, hepatocellular carcinoma; LT, liver transplantation; MRMR, maximum relevance minimum redundancy; RSF, random survival forest; TACE, trans-arterial chemoembolisation.

validation cohort, compared to other state-of-the-art predictive models, like the Milan criteria.[62]

The morphological features of HCC have a major impact on patient prognosis, and several DL algorithms have thus been developed to improve the prediction of HCC recurrence/survival using CT scans, MRI or histopathological images. Saillard *et al.* built a model based on the processing of HCC digital slides that was able to predict the survival of patients with HCC treated by surgical resection with a higher accuracy than scores including all relevant clinical, biological and pathological features. Notably, they were validated in a series of cases for which slides were stained with different protocols, suggesting that such models may generalise well when tested in different clinical centres.[63] A recent study from Yamashita *et al.* confirmed the capability of AI algorithms to predict outcomes based on digital histologic slides.[64] Lu and Daigle used 3 state-of-the-art CNNs (VGG 16, Inception v3, ResNet50), pretrained on ImageNet for feature extraction using HCC histopathology slides from the TCGA-LIHC cohort, and selected features significantly associated with survival using multivariable Cox regression analysis. While this again highlights the possibility of performing outcome prediction using histopathology slides, the conclusions are limited by the missing adjustment for other prognostic factors, as well as the lack of an external validation cohort.[65] Saito *et al.* applied classical ML methods to handcrafted whole slide image features from a relatively small cohort of 158 patients with HCC to develop a combined model, predicting HCC recurrence after resection with an accuracy of 89%. The next step will be to validate these promising results in a larger cohort.[66]

An exponentially growing number of studies also investigate the predictive performance of images from MRI or CT scans. Ji *et al.* combined several clinical and biological features (including serum alpha-fetoprotein, albumin-bilirubin [ALBI] grade and tumour margin status), and radiomics signatures to assess the risk of HCC recurrence after surgical resection.[67] Other authors also aimed to process CT scan or MR images to predict microvascular invasion, cytokeratin 19 expression (progenitor phenotype) or early tumour recurrence.[68–71] Several studies investigated the ability of AI methods to predict responses to transarterial chemoembolisation (TACE) in patients with advanced HCC. Abajian *et al.* used handcrafted radiomics features from MR images to train logistic regression and random forest models to classify patients treated with TACE as responders or non-responders. The models achieved a maximal overall accuracy of 78% but revealed the potential of ML algorithms in TACE response prediction.[72] Classical ML, as well as DL techniques were used on CT image radiomics features by Liu *et al.* to develop AI-based prognostic risk factors for overall survival. Interestingly, these factors were shown to be independently associated with survival; yet, it is important to highlight that the study lacked external validation and a simple train-validate-test split approach was used, which may limit generalisability.[73] Similarly, Zhang *et al.*'s DL score, based on a DenseNet-121 feature extraction architecture was also derived from CT images of patients with HCC treated with TACE plus sorafenib. The DL score was independently associated with overall survival, after controlling for known prognostic factors.[74] Using residual CNNs, Peng *et al.* trained (562 patients) and externally validated (89 and 138 patients) an algorithm yielding AUCs of at least 0.94 for prediction of complete or partial response and stable or progressive disease after TACE therapy.[75] A single study involving ultrasound was conducted by Oezdemir *et al.*, who extracted handcrafted HCC microvascular features from CEUS images to predict response to TACE. The model achieved an accuracy of 86%, yet the results require further evaluation due to the small sample size (n = 36).[76]

Current challenges limiting the use of AI for HCC risk prediction and prognostication

Need for standardisation of algorithms and software

Although AI holds many promises for the improvement of HCC detection and patient stratification, deployment of ML algorithms in clinical settings remains very rare. The safe translation of DL models will indeed require standardisation and robust evaluation using metrics that would ideally include patient outcomes and quality of care, as well as appropriate stakeholder engagement and oversight. To date, there are no standardised methods for AI-based data analysis or interpretation, and no universal approaches to address missing data, which is a fundamental concern in large-scale datasets. A significant number of published studies have investigated large series of patients with extensive benchmarking against expert performance, but, in the vast majority of cases, these studies were retrospective. Further, the performance of these models is likely to decrease when assessed prospectively using "real-world" data.

The establishment of consensus guidelines in reporting data from ML studies is also critical. A group is currently working on the definition of an AI-specific version of the STARD checklist (STARD-AI-Standards for Reporting of Diagnostic Accuracy Study-AI). These guidelines will aim to improve the completeness and transparency of studies investigating diagnostic test accuracy. Other recommendations will be needed for prognostic or theranostic biomarkers. Their performance should finally be compared to existing diagnostic, staging and predictive systems.

Key point

Key limitations of existing AI algorithms include overfitting of data, limited 'explainability' of results, and the possibility of poor generalisability, due to the inherent reliance of ML and DL models on the size and diversity of their training datasets.

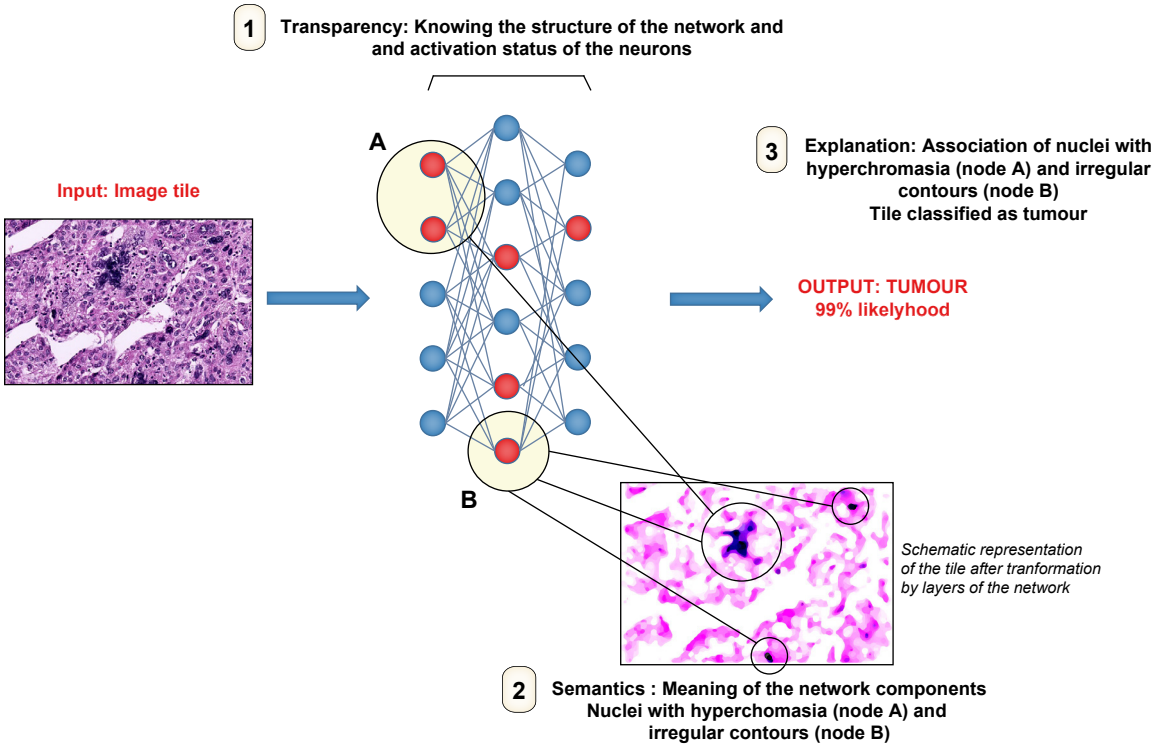

Fig. 3. Explainable artificial intelligence: example of pathology. This virtual model is dedicated to the prediction of the tumour or non-tumour nature of images from digital slides. The aim of explainable artificial intelligence is to better understand, through transparency, semantics and explanation, how the model makes its predictions. Transparency (1) consists of having an in-depth knowledge of the structure of the neural network and the activation status of its different neurons/nodes. Semantics will provide insights on the type of objects that result in the activation of particular parts of the network). Finally, explanation will enable clinicians to understand how the association of different features impact the final prediction.

Need for data sharing/open-source algorithms

As the performance of AI models is highly dependent on the amount of data used for training, the availability of large data sets is key to fostering the development of research and its future impact on clinical care. To this end, the deposition and sharing of large datasets should be encouraged. This includes utilisation and sharing of large-scale data from EHRs across and between health systems. Moreover, sharing of individual-participant data (IPD) from clinical trials or purely academic research studies, a clear "ethical and scientific imperative", has gained increasing traction and is now advocated by many scientists and organisations, and would assist in constructing datasets of sufficient size and detail to appropriately train and validate AI models.[77] Moreover, a universal, standardised method for addressing and analysing missing data in AI models is necessary, and this is particularly important when considering shared datasets. The International Committee of Medical Journal Editors has thus implemented a clinical trial data policy that requires an IPD sharing statement for manuscripts reporting clinical trials. Although several repositories are now able to store IPD and make it available to third parties, the rate of sharing remains very low. The main obstacle is likely to be cultural, however other issues remain, such as patients' anonymity and the residual risk of re-identification, cost of data storage/provision, and need for specific consent regarding sharing. However, the availability of IPD from clinical trials (including imaging and digital slides) testing systemic therapies will be key for the development of AI models able to predict response/survival.

Need for sufficiently diverse populations

To date, cohorts used to develop and train AI models focused on HCC risk prediction, diagnosis and prognostication have lacked sufficient racial, ethnic and socioeconomic diversity. This is a critical issue, given that the accuracy of AI-based algorithms depends upon the validity and size of their input data. Consequently, future studies will need to ensure that promising AI-based tools are thoughtfully validated in diverse cohorts that include racial and ethnic minorities as well as patients across the complete socioeconomic spectrum. This once again underscores the need for data sharing between investigators and across institutions, so that representative cohorts can be constructed.

Examples from other disciplines

Currently, approximately 150 AI-based medical devices have been approved by the FDA. Most of these

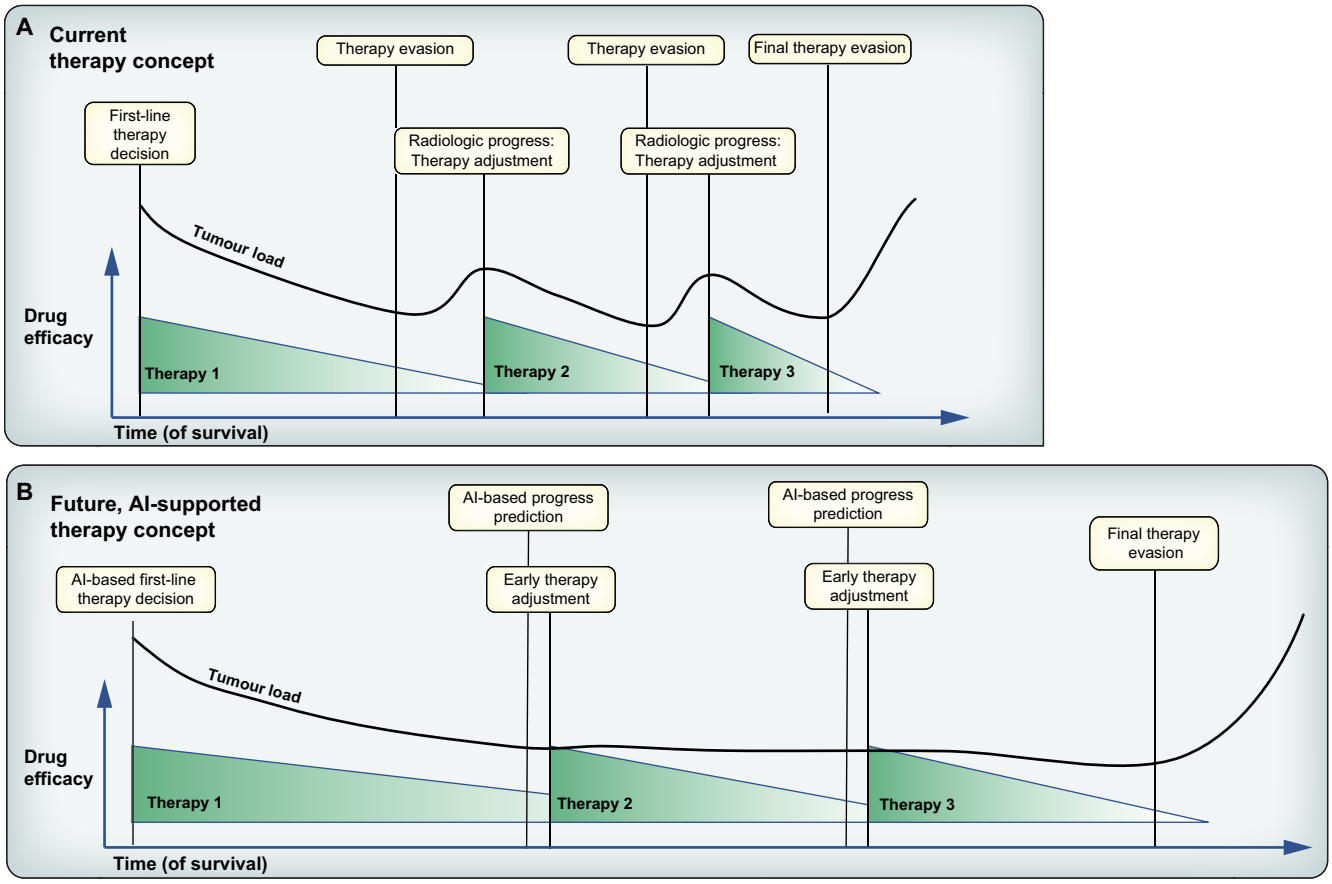

Fig. 4. Artificial intelligence could support doctors in decision making in tumour therapy in the future. (A) Current oncologic therapy pattern. After an initial first-line therapy, the tumour is evading therapy through resistance mechanisms. The following tumour growth is recognised during radiologic follow-up leading to therapy adjustment. (B) Hypothetical, future, AI-supported therapy pattern. Initial, individualized first-line therapy decision, accounting for an AI-based recommendation. After an AI algorithm predicts progression of a tumour, doctors decide to adjust therapy before the tumour can develop resistance to therapy and grow again.

models were developed for the fields of radiology (*e.g.* CT scan image reconstruction or brain MRI interpretation), cardiology (*e.g.* electrocardiogram analysis, cardiac monitoring) and ophthalmology (detection of diabetic retinopathy). Interestingly the FDA has also very recently granted its first clearance for an AI-based pathology software application. The product analyses digital slides of prostatic biopsies, highlights areas that are most likely to contain cancer and flags them for further review by a pathologist (https://www.paige.ai/). This landmark approval marks the beginning of a new era in the use of AI-assisted diagnostics for pathology, and it is very likely that models aiming to assist HCC histological diagnosis/prognosis assessment will also be available soon. They are particularly needed to assist with the differentiation of benign *vs.* malignant hepatocellular tumours, and also for a more robust and standardised diagnosis of rare pathological entities, such as combined hepatocellular-cholangiocarcinoma or fibrolamellar carcinoma.

Explaining "the black box" of AI

A common issue for all existing and future AI applications is to make their decisions comprehensible to the user. The term "explainable AI" refers to a particular set of methods that allows users to comprehend how the AI models work and make their decisions. It thus provides feedback on the most important features involved in the predictions and helps to understand the potential biases. This transparency is critical to build up the trust needed to convince doctors to rely on these computer-aided devices they might be using in the future. The approaches most commonly used in DL consist of extremely complex layers of mathematical computation, and it is thus very difficult to gain insights into how the data are transformed throughout the whole network.

Explainable AI is however an active field of research and many aim to open the black boxes of NNs. The main strands of work are making the networks "transparent", learning the semantics of

its different components and finally generating *post hoc* explanations. Transparency mainly consists of understanding the model structure and its function. Semantics of the different network components will provide insights on the meaning of particular neurons and the *post hoc* explanation finally analyses why a result is inferred (Fig. 3).[78]

For example, *post hoc* explanations of models processing digital histology slides can be established by getting a human expert to review the image areas associated with the highest predictive value.. This type of approach was used in the study by Saillard *et al.*, who built a model able to predict the survival of patients after resection of HCC. Interestingly, reviewing the tumoural tiles associated with a high risk of death showed an enrichment in several features (including macrotrabecular-massive subtype, cellular atypia) previously shown to be predictive of dismal clinical outcome.[63] These results show that the models, at least in part, rely on known histological parameters. The authors also identified a new prognostic feature, *i.e.* the presence of vascular spaces. Together, these results underscore the importance of human/machine interactions and show that novel hypotheses can be generated with this type of approache. Altogether, addressing 'explainability' is a critical issue, and will be necessary to: i) gain the required confidence in AI models' outputs, and ii) exploit NNs to discover key features that may have been overlooked.

Key point

There remains a great need to standardise and robustly evaluate AI algorithms in prospective studies and using large-scale "real-world" datasets, as well as to establish consensus guidelines to ensure accurate and comprehensive reporting of data from ML and DL studies.

Future applications of AI: towards tailored clinical trials

Prospective studies are needed to fully demonstrate the potential of AI to improve the clinical care of patients with HCC. In other medical areas, several AI-based randomised clinical trials have already been conducted. As such, in endoscopy, numerous randomised clinical trials have evaluated the impact of computer-aided systems on physicians' performance in diagnosing intestinal adenoma or indicating blind spots of colonoscopy.[79,80] The need to incorporate these new developments prompted the research community to extend the widely used SPIRIT and CONSORT guidelines for the use of AI methods in 2020.[81,82] According to ClinicalTrials.gov (https://clinicaltrials.gov/), there are currently 6 ongoing trials involving AI for the management of HCC. A research group at the University of Hong Kong is comparing an algorithm designed to diagnose HCC from CT images against the standard diagnostic procedure that relies on the LI-RADS criteria (NCT04843176).[83] A multicentre study from France is prospectively developing an AI algorithm in a non-randomised clinical trial. The research group uses clinical, biological and ultrasound data to stratify the risk of HCC emergence in high- and low-risk patients.[84]

Treatment with immune checkpoint inhibitors (ICIs) has represented a fundamental breakthrough in many cancers.[85–87] In palliative treatment of HCC patients, the IMBRAVE-150 trial showed that the combination of atezolizumab and bevacizumab conferred a significant survival benefit compared to sorafenib in patients with HCC.[3] However, like in many previous trials in distinct entities, it became apparent that not all patients with HCC benefit from ICIs to a similar extent. While there are signals for HCC subgroups with a potentially higher benefit (*e.g.* viral hepatitis *vs.* non-viral liver disease[88]), there is still no biomarker that reliably predicts therapeutic response before or very early after starting ICI therapy in patients with HCC. Therefore, a significant fraction of patients will be subjected to the (low) risk of severe ICI-related toxicity without benefit, thereby being at an increased risk of tumour progression and worsened liver function, while the cost of ICI therapy is remarkably high. In this setting, AI-based response prediction could play a key role in improving patient outcomes and reducing healthcare expenditure.

Generating, training and applying an algorithm could involve a deep net trained on histologic data, *e.g.* from randomised clinical trials in immunotherapy, and/or the combination of different deep nets including histology, radiology, genomic and clinical information. Importantly, a DL-based algorithm could either be trained on data available before the start of therapy or on data extracted immediately after the initiation of therapy. Thus, it may, before the first radiological response evaluation, provide early predictions of whether a patient will benefit or should be switched to another therapeutic strategy. Beyond determining the ideal first-line therapy per patient, AI-based decision making could also provide a basis for a fundamental switch in the way that treatment changes are implemented into long term palliative treatment of oncologic patients. Currently, a successful line of therapy is provided to a patient until radiological progression is evident (Fig. 4). However, it could be beneficial to establish a tool for the early prediction of treatment failure, recommending a switch to another therapy, even before full progression is documented on imaging. This tool could enable preemptive therapy adjustment in the interval between molecular resistance and imaging (Fig. 3). AI could represent the ideal toolbox to facilitate such a concept. Similar to a first-line decision, an algorithm would need to be trained within clinical trials, first proving that radiological progression can be reliably predicted, *e.g.* on an algorithm trained on radiology, but also on laboratory values and clinical parameters. Once a proof of concept for an AI algorithm is achieved, future clinical trials could compare a possible benefit from

early AI-based regimen switches to a conventional approach based on pure radiological progression within the standard clinical imaging intervals (*e.g.* 6, 8, or 12 weeks).

While these concepts are still hypothetical, it will be important to integrate AI-based algorithms into current and future clinical trials, in order to prove that they are valuable tools to predict responses to first-line therapy and to predict early progression. Implementing these steps will depend on access to biological samples and clinical data within large clinical trials, and will require acceptance of these concepts and further that these data are made accessible to the clinician scientists who are contributing patients to these trials. To that end, collaborative networks based on trust and united in the collective aim of improving patient outcomes need to be implemented not only between clinicians but also with industry. Nevertheless, it is paramount for any model developed and trained within the framework of a clinical trial to be thoroughly validated in diverse, real-world patient populations before clinical implementation, to address possible biases introduced by the trial's inclusion criteria. Moreover, AI-based algorithms and any resultant clinical tools must also be constructed with appropriate stakeholder engagement and oversight, to ensure that validated algorithms are standardised according to protocol and that they are used in the correct clinical contexts, and further that data output is interpreted properly to maximise clinical benefit. Correctly interpretating data output from an AI-based clinical tool will in turn require appropriate training and awareness, both amongst the public and clinical providers.

Conclusion

It is hoped that AI will profoundly change the way we care for patients with HCC. Although significant progress has been made during the last decade, improvements in HCC risk prediction, diagnosis and response prediction are still critically needed. Several challenges remain to fully implement such technologies in clinical practice, including the need to develop robust approaches for structured data collection, sharing and storage, and the need to demonstrate the reliability and robustness of models. We know that AI can predict a very large set of clinically relevant features, and we must also now demonstrate that these approaches work in a clinical setting, by comparing model performance to that of conventional staging systems, and further through the careful design of large prospective trials.

Abbreviations

AI, artificial intelligence; CEUS, contrast-enhanced ultrasound; DCNN, deep convolutional neural network; DL, deep learning; EHR, electronic health record; HCC, hepatocellular carcinoma; ICI, immune checkpoint inhibitors; IPD, individual-participant data; ML, machine learning; NAFLD, non-alcoholic fatty liver disease; NN(s), neural network(s); TACE, transarterial chemoembolisation.

Financial support

NIH K23 DK122104 (TGS). Dana-Farber/Harvard Cancer Center GI SPORE Career Enhancement Award (TGS).

Conflict of interest

Dr. Simon has served as a consultant to Aetion and has received grants to the institution from Amgen, for work unrelated to this manuscript. Pr Calderaro serves as a consultant for Keen Eye, Crosscope and Owkin.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Concept and literature review: all co-authors. Drafting of manuscript: all co-authors. Critical revision of the manuscript: all co-authors. Guarantor of the manuscript: Simon. All authors contributed to the critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. The corresponding author (TGS) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2022.01.014.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–249.
[2] Baecker A, Liu X, La Vecchia C, Zhang Z-F. Worldwide incidence of hepatocellular carcinoma cases attributable to major risk factors. Eur J Cancer Prev 2018;27:205–212.
[3] Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim T-Y, et al. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. N Engl J Med 2020;382:1894–1905.
[4] El-Serag HB, Kanwal F. Epidemiology of hepatocellular carcinoma in the United States: where are we? Where do we go? Hepatology 2014;60:1767–1775.
[5] Singal AG, Mukherjee A, Elmunzer BJ, Higgins PDR, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol 2013;108:1723–1730.
[6] Ioannou GN, Tang W, Beste LA, Tincopa MA, Su GL, Van T, et al. Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis C cirrhosis. JAMA Netw Open 2020;3:e2015626.
[7] Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 2018;67:358–380.
[8] Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 2018;68:723–750.
[9] European 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.

[10] Yu NC, Chaudhari V, Raman SS, Lassman C, Tong MJ, Busuttil RW, et al. CT and MRI improve detection of hepatocellular carcinoma, compared with ultrasound alone, in patients with cirrhosis. Clin Gastroenterol Hepatol 2011;9:161–167.
[11] Vecchiato F, D'Onofrio M, Malagò R, Martone E, Gallotti A, Faccioli N, et al. Detection of focal liver lesions: from the subjectivity of conventional ultrasound to the objectivity of volume ultrasound. Radiol Med 2009;114:792–801.
[12] Schmauch B, Herent P, Jehanno P, Dehaene O, Saillard C, Aubé C, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. Diagn Interv Imaging 2019;100:227–233.
[13] Yang Q, Wei J, Hao X, Kong D, Yu X, Jiang T, et al. Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: a multicentre study. EBioMedicine 2020;56:102777.
[14] Guo L-H, Wang D, Qian Y-Y, Zheng X, Zhao C-K, Li X-L, 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;69:343–354.
[15] Ta CN, Kono Y, Eghtedari M, Oh YT, Robbin ML, Barr RG, et al. Focal liver lesions: computer-aided diagnosis by using contrast-enhanced US cine recordings. Radiology 2018;286:1062–1071.
[16] Preis O, Blake MA, Scott JA. Neural network evaluation of PET scans of the liver: a potentially useful adjunct in clinical interpretation. Radiology 2011;258:714–721.
[17] Mokrane F-Z, Lu L, Vavasseur A, Otal P, Peron J-M, Luk L, et al. Radiomics machine-learning signature for diagnosis of hepatocellular carcinoma in cirrhotic patients with indeterminate liver nodules. Eur Radiol 2020;30:558–570.
[18] 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;286:887–896.
[19] Shi W, Kuang S, Cao S, Hu B, Xie S, Chen 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;45:2688–2697.
[20] Christ P, Ettlinger F, Grün F, Lipkova J, Kaissis G. Lits - liver tumor segmentation challenge n.d. http://www.lits-challenge.com (accessed December 12, 2021).
[21] Chlebus G, Schenk A, Moltz JH, van Ginneken B, Hahn HK, Meine H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. Sci Rep 2018;8:15497.
[22] 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.
[23] Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol 2019;29:3338–3347.
[24] Zhang F, Yang J, Nezami N, Laage-Gaupp F, Chapiro J, De Lin M, et al. Liver tissue classification using an auto-context-based deep neural network with a multi-phase training framework. Patch Based Tech Med Imaging 2018;11075:59–66. 2018.
[25] Zhen S-H, Cheng M, Tao Y-B, Wang Y-F, Juengpanich S, Jiang Z-Y, et al. Deep learning for accurate diagnosis of liver tumor based on magnetic resonance imaging and clinical data. Front Oncol 2020;10:680.
[26] Wang CJ, Hamm CA, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. Eur Radiol 2019;29:3348–3357.
[27] Liao H, Long Y, Han R, Wang W, Xu L, Liao M, et al. Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma. Clin Transl Med 2020;10:e102.
[28] Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. Npj Digital Med 2020;3. https://doi.org/10.1038/s41746-020-0232-8.
[29] Calderaro J, Couchy G, Imbeaud S, Amaddeo G, Letouzé E, Blanc J-F, et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. J Hepatol 2017;67:727–738.
[30] Calderaro J, Ziol M, Paradis V, Zucman-Rossi J. Molecular and histological correlations in liver cancer. J Hepatol 2019;71:616–630.
[31] Ziol M, Poté N, Amaddeo G, Laurent A, Nault J-C, Oberti F, et al. Macrotrabecular-massive hepatocellular carcinoma: a distinctive histological subtype with clinical relevance. Hepatology 2018;68:103–112.
[32] Wang H, Jiang Y, Li B, Cui Y, Li D, Li R. Single-cell spatial analysis of tumor and immune microenvironment on whole-slide image reveals hepatocellular carcinoma subtypes. Cancers (Basel) 2020;12:3562.
[33] Chen M, Zhang B, Topatana W, Cao J, Zhu H, Juengpanich S, et al. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. NPJ Precis Oncol 2020;4:14.
[34] Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019;25:1054–1056.
[35] Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer 2020;1:789–799. https://doi.org/10.1038/s43018-020-0087-6.
[36] Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer 2020;1:800–810.
[37] Sangro B, Melero I, Wadhawan S, Finn RS, Abou-Alfa GK, Cheng A-L, et al. Association of inflammatory biomarkers with clinical outcomes in nivolumab-treated patients with advanced hepatocellular carcinoma. J Hepatol 2020;73:1460–1469.
[38] Haber PK, Torres-Martin M, Dufour J-F, Verslype C, Marquardt J, Galle PR, et al. Molecular markers of response to anti-PD1 therapy in advanced hepatocellular carcinoma. J Clin Oncol 2021;39. 4100–4100.
[39] Johannet P, Coudray N, Donnelly DM, Jour G, Illa-Bochaca I, Xia Y, et al. Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma. Clin Cancer Res 2021;27:131–140.
[40] Patel SK, George B, Rai V. Artificial intelligence to decode cancer mechanism: beyond patient stratification for precision oncology. Front Pharmacol 2020. https://doi.org/10.3389/fphar.2020.01177. 0.
[41] Liu S, Yang Z, Li G, Li C, Luo Y, Gong Q, et al. Multi-omics analysis of primary cell culture models reveals genetic and epigenetic basis of intratumoral phenotypic diversity. Genomics Proteomics Bioinformatics 2019;17:576–589.
[42] Zeng WZD, Glicksberg BS, Li Y, Chen B. Selecting precise reference normal tissue samples for cancer research using a deep learning approach. BMC Med Genomics 2019;12:21.
[43] Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res 2018;24:1248–1259.
[44] Chaudhary K, Poirion OB, Lu L, Huang S, Ching T, Garmire LX. Multimodal meta-analysis of 1,494 hepatocellular carcinoma samples reveals significant impact of consensus driver genes on phenotypes. Clin Cancer Res 2019;25:463–472.
[45] Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 2018;50:1–14.
[46] Xiong X, Kuang H, Ansari S, Liu T, Gong J, Wang S, et al. Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. Mol Cell 2019;75:644–660.e5.
[47] Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature 2019;575:512–518.
[48] Aizarani N, Saviano A, Sagar, Mailly L, Durand S, Herman JS, et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. Nature 2019;572:199–204.
[49] Zheng C, Zheng L, Yoo J-K, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell 2017;169:1342–1356.e16.
[50] Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 2019;179:829–845.e20.
[51] Kim JK, Kolodziejczyk AA, Ilicic T, Teichmann SA, Marioni JC. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. Nat Commun 2015;6:8687.
[52] Jia C, Hu Y, Kelly D, Kim J, Li M, Zhang NR. Accounting for technical noise in differential expression analysis of single-cell RNA sequencing data. Nucleic Acids Res 2017;45:10978–10988.
[53] Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol 2018;18:35–45.
[54] Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nat Methods 2014;11:740–742.
[55] Arisdakessian C, Poirion O, Yunits B, Zhu X, Garmire LX. DeepImpute: an accurate, fast, and scalable deep neural network method to impute single-cell RNA-seq data. Genome Biol 2019;20:211.
[56] Amodio M, van Dijk D, Srinivasan K, Chen WS, Mohsen H, Moon KR, et al. Exploring single-cell data with deep multitasking neural networks. Nat Methods 2019;16:1139–1145.

[57] Marouf M, Machart P, Bansal V, Kilian C, Magruder DS, Krebs CF, et al. Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. Nat Commun 2020;11:166.
[58] Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun 2019;10:390.
[59] Genshaft AS, Li S, Gallant CJ, Darmanis S, Prakadan SM, Ziegler CGK, et al. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. Genome Biol 2016;17. https://doi.org/10.1186/s13059-016-1045-6.
[60] Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol 2015;33:285–289.
[61] Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods 2015;12:519–522.
[62] Nam JY, Lee J-H, Bae J, Chang Y, Cho Y, Sinn DH, et al. Novel model to predict HCC recurrence after liver transplantation obtained using deep learning: a multicenter study. Cancers 2020;12:2791. https://doi.org/10.3390/cancers12102791.
[63] Saillard C, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. Hepatology 2020;72:2000–2013.
[64] Yamashita R, Long J, Saleem A, Rubin DL, Shen J. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. Sci Rep 2021;11:2047.
[65] Lu L, Daigle Jr BJ. Prognostic analysis of histopathological images using pre-trained convolutional neural networks: application to hepatocellular carcinoma. PeerJ 2020;8:e8668.
[66] Saito A, Toyoda H, Kobayashi M, Koiwa Y, Fujii H, Fujita K, et al. Prediction of early recurrence of hepatocellular carcinoma after resection using digital pathology images assessed by machine learning. Mod Pathol 2021;34:417–425.
[67] Ji G-W, Zhu F-P, Xu Q, Wang K, Wu M-Y, Tang W-W, et al. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: a multi-institutional study. EBioMedicine 2019;50:156–165.
[68] Song D, Wang Y, Wang W, Wang Y, Cai J, Zhu K, et al. Using deep learning to predict microvascular invasion in hepatocellular carcinoma based on dynamic contrast-enhanced MRI combined with clinical parameters. J Cancer Res Clin Oncol 2021. https://doi.org/10.1007/s00432-021-03617-3.
[69] Zhang Y, Lv X, Qiu J, Zhang B, Zhang L, Fang J, et al. Deep learning with 3D convolutional neural network for noninvasive prediction of microvascular invasion in hepatocellular carcinoma. J Magn Reson Imaging 2021;54:134–143.
[70] Jiang Y-Q, Cao S-E, Cao S, Chen J-N, Wang G-Y, Shi W-Q, et al. Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning. J Cancer Res Clin Oncol 2021;147:821–833.
[71] Wang W, Chen Q, Iwamoto Y, Han X, Zhang Q, Hu H, et al. Deep learning-based radiomics models for early recurrence prediction of hepatocellular carcinoma with multi-phase CT images and clinical data. Conf Proc IEEE Eng Med Biol Soc 2019;2019:4881–4884.
[72] Abajian A, Murali N, Savic LJ, Laage-Gaupp FM, Nezami N, Duncan JS, et al. Predicting treatment response to intra-arterial therapies for hepatocellular carcinoma with the use of supervised machine learning-an artificial intelligence concept. J Vasc Interv Radiol 2018;29:850–857.e1.
[73] Liu Q-P, Xu X, Zhu F-P, Zhang Y-D, Liu X-S. Prediction of prognostic risk factors in hepatocellular carcinoma with transarterial chemoembolization using multi-modal multi-task deep learning. EClinicalMedicine 2020;23:100379.
[74] Zhang L, Xia W, Yan Z-P, Sun J-H, Zhong B-Y, Hou Z-H, et al. Deep learning predicts overall survival of patients with unresectable hepatocellular carcinoma treated by transarterial chemoembolization plus sorafenib. Front Oncol 2020;10:593292.
[75] Peng J, Kang S, Ning Z, Deng H, Shen J, Xu Y, et al. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. Eur Radiol 2020;30:413–424.
[76] Oezdemir I, Wessner CE, Shaw C, Eisenbrey JR, Hoyt K. Tumor vascular networks depicted in contrast-enhanced ultrasound images as a predictor for transarterial chemoembolization treatment response. Ultrasound Med Biol 2020;46:2276–2286.
[77] Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. JAMA 2016;315:1237–1239.
[78] Xu F, Uszkoreit H, Du Y, Fan W, Zhao D, Zhu J. Explainable AI: a brief survey on history, research areas, approaches and challenges. Natural language processing and Chinese computing. Springer International Publishing; 2019. p. 563–574.
[79] Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–1819.
[80] Wu L, Zhang J, Zhou W, An P, Shen L, Liu J, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019;68:2161–2169.
[81] Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, et al., SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med 2020;26:1351–1363.
[82] Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020;26:1364–1374.
[83] (md) CGB. Identifier NCT04843176. A prototype Artificial intelligence algorithm versus liver imaging reporting and data system (LI-RADS) criteria in diagnosing hepatocellular carcinoma on computed tomography: a randomized trial. National Library of Medicine (US); 2021.
[84] Gov C. NCT04802954. Risk stratification of hepatocarcinogenesis using a deep learning based clinical, biological and ultrasound model in high-risk patients (STARHE). 2021.
[85] Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010;363:711–723.
[86] Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. N Engl J Med 2015;373:1627–1639.
[87] Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non–small-cell lung cancer. N Engl J Med 2015;372:2018–2028.
[88] Li X, Ramadori P, Pfister D, Seehawer M, Zender L, Heikenwalder M. The immunological and metabolic landscape in primary and metastatic liver cancer. Nat Rev Cancer 2021;21:541–557.