



Artificial intelligence in the detection, characterisation and prediction of hepatocellular carcinoma: a narrative review

Michal Kawka^{1^}, Aleksander Dawidziuk^{1^}, Long R. Jiao^{2,3}, Tamara M. H. Gall²

¹Department of Medicine, Imperial College London, South Kensington Campus, London, UK; ²HPB Surgical Unit, Department of Surgery & Cancer, Imperial College London, Hammersmith Hospital Campus, London, UK; ³Royal Marsden Hospital, London, UK

Contributions: (I) Conception and design: All authors; (II) Administrative support: TMH Gall, LR Jiao; (III) Provision of study materials or patients: TMH Gall, M Kawka, A Dawidziuk; (IV) Collection and assembly of data: A Dawidziuk, M Kawka; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Tamara M. H. Gall. Department of Surgery & Cancer, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London, W12 0HS, UK. Email: t.gall12@imperial.ac.uk.

Abstract: Hepatocellular carcinoma (HCC) is a significant cause of morbidity and mortality worldwide. Despite significant advancements in detection and treatment of HCC, its management remains a challenge. Artificial intelligence (AI) has played a role in medicine for several decades, however, clinically applicable AI-driven solutions have only started to emerge, due to gradual improvement in sensitivity and specificity of AI, and implementation of convoluted neural networks. A review of the existing literature has been conducted to determine the role of AI in HCC, and three main domains were identified in the search: detection, characterisation and prediction. Implementation of AI models into detection of HCC has immense potential, as AI excels at analysis and integration of large datasets. The use of biomarkers, with the rise of ‘-omics’, can revolutionise the detection of HCC. Tumour characterisation (differentiation between benign masses, HCC, and other malignant tumours, as well as staging and grading) using AI was shown to be superior to classical statistical methods, based on radiological and pathological images. Finally, AI solutions for predicting treatment outcomes and survival emerged in recent years with the potential to shape future HCC guidelines. These AI algorithms based on a combination of clinical data and imaging-extracted features can also support clinical decision making, especially treatment choice. However, AI research on HCC has several limitations, hindering its clinical adoption; small sample size, single-centre data collection, lack of collaboration and transparency, lack of external validation, and model overfitting all results in low generalisability of the results that currently exist. AI has potential to revolutionise detection, characterisation and prediction of HCC, however, for AI solutions to reach widespread clinical adoption, interdisciplinary collaboration is needed, to foster an environment in which AI solutions can be further improved, validated and included in treatment algorithms. In conclusion, AI has a multifaceted role in HCC across all aspects of the disease and its importance can increase in the near future, as more sophisticated technologies emerge.

Keywords: Hepatocellular carcinoma (HCC); artificial intelligence (AI); machine learning (ML); computer-aided diagnosis; neural networks

Received: 23 June 2020; Accepted: 13 September 2020; Published: 25 October 2022.

doi: 10.21037/tgh-20-242

View this article at: <http://dx.doi.org/10.21037/tgh-20-242>

[^] ORCID: Michal Kawka: 0000-0002-5826-8355; Aleksander Dawidziuk: 0000-0001-5702-049X.

Introduction

Hepatocellular carcinoma (HCC) is the most common adult primary liver malignancy causing the fourth most cancer-related deaths worldwide (1). The highest burden of HCC is found in East-Asia and Sub-Sahara Africa, however, it is also an emerging major cause of morbidity and mortality globally (2). There were significant improvements in our understanding of the aetiology and pathogenesis of HCC throughout the years, which have contributed to novel surveillance, diagnosis and management strategies. However, despite these advancements, HCC still poses challenges for clinicians worldwide. HCC is characterised by presenting at advanced stages, often on a background of pre-existing liver disease, which poses diagnostic challenges. These challenges are both histological, due to heterogeneity of HCC and continuum of its malignant progression and radiological due to HCC mimickers such as i-CCA or c-CCA and prevalence of atypical radiological features (approx. 40%) (3,4). Moreover, extensive background liver disease can impede potentially curative therapies (5). Due to heterogeneity in pathogenesis and multiple recognised risk factors, the biological behaviour of HCC is mostly unknown, which translates into discrepancies in staging systems and prognosis (6). Moreover, HCC has a high recurrence rate following surgical resection, and even orthotopic liver transplantation has a five year survival of only 65–81% despite using specific criteria (Milan, USFC, Kyoto) aimed at selecting patients thought to have better long-term outcomes (6,7).

Development of artificial intelligence (AI) provides a unique set of novel tools to aimed at solving the aforementioned issues and improve HCC detection, characterisation, prediction of survival and treatment outcomes. Although AI has been applied in medicine for over 60 years (8), only the recent research has provided a plethora of successful studies with potential for clinical impact, including classification of skin cancers (9), breast cancer screening (10) or retinal examination (11). AI-driven solutions can help in early detection of HCC, more accurate diagnosis and classification of the tumour, as well as predicting disease course and outcomes. In order to understand the potential future role of AI in various aspects of HCC management, and in-depth review of currently available evidence of the role of AI in HCC was conducted. By providing an overview of the strengths and limitations of AI in HCC, the authors aimed to understand the factors limiting widespread clinical adoption of AI-driven solutions

and provide recommendations on future AI research.

This review explores AI solutions applied to HCC that can be classified into three main domains: detection, characterisation and prediction. In the context of this review, detection encompasses technologies that highlight an abnormality, both in imaging and clinical data, which allow for further follow-up. Characterisation includes techniques that differentiate between various hepatic abnormalities, as well as, stratify and stage previously diagnosed HCC. Prediction comprises methods, which use AI for evaluating long-term outcomes including overall survival, tumour progression or assessment of response to treatment. We present the following article in accordance with the Narrative Review reporting checklist (available at <https://tgh.amegroups.com/article/view/10.21037/tgh-20-242/rc>).

Definitions

In order to discuss AI in context of HCC, its definition needs to be established. AI describes the use of computers and related technology to emulate intelligent behaviour and critical thinking of human beings (12). Within AI, one of the most commonly used terms is machine learning (ML), which can be defined as a discipline in which machines (computers) learn from data, with emphasis on computational algorithms, which are able to analyse billions of data points (13). Of the most commonly used AI techniques are artificial neural networks (ANN); statistical systems, which mimic the complex architecture of biological networks of neurons, to derive outputs, based on interactions of weighted inputs and outputs (14). Convolutional neural networks (CNN) can be especially useful in the medical context, as they have the ability to process data with a grid pattern (e.g., radiological images) using multiple layers, including convolution and pooling layers performing feature extraction to produce final output (15). Additional terms related to AI that were used throughout this review are defined in *Table 1*.

Methods

A comprehensive search of Embase, MEDLINE and Cochrane Library databases was conducted. The databases were searched from their conception to 14th March 2020. The search was conducted using MeSH terms and keywords for hepatocellular carcinoma (hepatocellular carcinoma, HCC) artificial intelligence (AI, artificial intelligence, machine intelligence), machine learning (machine learning,

Table 1 Glossary of terms related to the use of artificial intelligence in medicine and performance of AI-driven solutions

Term	Definition
Artificial intelligence (AI)	The use of computers and related technology to emulate intelligent behaviour and critical thinking of human beings (12)
Machine learning (ML)	A discipline in which machines (computers) learn from data, with emphasis on computational algorithms, which are able to analyse billions of data points (13)
Supervised learning	Type of ML, which deals with predicting a known outcome, based on inputs, in the presence of an expert 'supervisor' (16)
Unsupervised learning	Type of ML, which deals with finding naturally occurring patterns without a pre-defined outcome, without the presence of an expert 'supervisor' (16)
Artificial neural networks (ANN)	Statistical systems, which mimic the complex architecture of biological networks of neurons, to derive outputs, based on interactions of weighted inputs and outputs (14)
Convolutional neural network (CNN)	Type of deep learning ANN, for processing data with grid pattern (e.g., radiological images) using multiple layers, including convolution and pooling layers performing feature extraction to produce final output (15)
Deep learning	A subset of ML, that uses representation learning (automatic discovery of representations from raw data for classification or detection) (17)
Area under the curve (AUC)	Algorithm performance measure, which can be established based on receiver operator characteristics (ROC) curve. AUC takes values between 0 and 1, depending on average sensitivity and specificity for all analysed values of the, with values approaching 1 indicating higher performance (18)
Accuracy	Algorithm performance measure, taking values between 0% and 100%, based on the number of true positive and true negative results, compared to the overall size of the population (18)
C-index (c-statistic)	Algorithm performance measure, describing the goodness of fit of the model, taking values between 0 and 1 (19)

ML, deep learning, supervised learning, unsupervised learning) artificial neural network (artificial neural networks, neural network, ANN, NN) and convolutional neural network (convolutional neural network, CNN) combined with Boolean operators. Relevant keywords were identified using recent related publications on AI in HCC (20,21).

In total, 702 records were identified, and after removal of duplicates, 470 records have remained. The titles and abstracts were screened by two authors independently (MK and AD), and a third author resolved any conflicts (TMHG). Inclusion criteria included: original research studies, English language, use of AI-driven solutions and HCC as primary pathology of interest. Exclusion criteria included: malignancies other than HCC as primary pathology of interest, age <18, non-original research studies (e.g., commentaries, letters to editors, reviews). Relevant original research studies were included in the analysis, however, conference abstracts and review articles were also screened. Manual screening of reference lists of included full texts was also performed by two authors (MK and AD) independently to look for any missing studies. After the

list of accepted texts was finalised, full texts were classified into one of three domains that have been described in the introduction (detection, characterisation and prediction). Following classification into domains, data extraction was conducted; variables of interest were the number of participants/pathological slides/radiological images, AI algorithm used, type of validation, AUC (or if not available, alternative diagnostic accuracy measures such as sensitivity and specificity, c-index or F-score).

Detection

The summary of literature on HCC detection, based on pre-HCC disease models, imaging and biomarkers is presented in *Table 2*.

Pre-HCC disease models

HCC pathogenesis is strongly linked to chronic inflammatory disease of the liver, which allows for HCC detection based on the range of pre-malignant changes.

Table 2 Summary of literature on HCC detection, based on pre-HCC diseases models, imaging and biomarkers

Domain	Sub-category	Notes	N	AI algorithm	Type of validation	AUC (95% CI)	Limitations	Reference
Pre-HCC disease models	NAFLD/NASH	Serum and liver lipids in murine models	15 (5 intervention, 10 control)	Random forest	Development only (no validation)	N/A	Animal (murine) model not replicated on humans	Chiappini <i>et al.</i> 2016 (22)
		Population screening for NAFLD	500 (146 cases, 354 controls)	Logistic regression	Internal validation (cross-validation)	0.87 (0.83–0.90)	Retrospective study design and lack of external validation	Yip <i>et al.</i> 2017 (23)
	Cirrhosis/fibrosis/hepatitis B/hepatitis C	Progression of cirrhosis into HCC	442 patients	Random forest	Internal validation	c-index 0.64 (0.60–0.90)	Study performed in a tertiary centre. Attrition bias. Low clinical adaptability potential	Singal <i>et al.</i> 2012 (24)
		Progression of chronic HepC infection into fibrosis	533 (349 normal, 184 fibrosis)	Random forest	Internal validation	0.84 (0.82–0.86)	Narrow enrolment criteria, reducing generalisability of conclusions	Konerman <i>et al.</i> 2015 (25)
		Progression of HepB/C into HCC	6,561 (Reddy <i>et al.</i>); 146,218 (Ioannou <i>et al.</i>)	ANN	Internal validation (random sample split)	0.911–0.962 (range, Reddy <i>et al.</i>); 0.89 (Ioannou <i>et al.</i>)	Retrospective and cross-sectional character of the studies and lack of external validation	Reddy <i>et al.</i> 2017 (26); Ioannou <i>et al.</i> 2019 (27)
Miscellaneous	HCC development based on viral status and clinical data	165 patients	Support vector machine	Internal validation (cross-validation)	0.88	Small sample size	Książek <i>et al.</i> 2019 (28)	
Imaging	CT	HCC detection on multi-phase CT scans	25 (Lee <i>et al.</i>); 21 (Okumura <i>et al.</i>)	Temporal subtraction, 3D global matching	Development only (no validation)	N/A	Small sample size, proof-of-concept character of both studies	Lee <i>et al.</i> 2015 (29); Okumura <i>et al.</i> 2011(30)
	MRI	HCC nodule detection using SPIO-MRI in rat models	40 images	ANN	Development only (no validation)	Classification accuracy 91.76%	Animal (rat) models used. No test-retest reproducibility	Guo <i>et al.</i> 2009 (31)
Biomarkers	N/A	Serum	167 (Poon <i>et al.</i>); 1,582 (Sato <i>et al.</i>)	Multiple techniques combined	Development only (no validation)	Accuracy 75.9% (Poon <i>et al.</i>); 0.844–0.940 (range, Sato <i>et al.</i>)	Single centre, internally validated studies with small sample size	Poon <i>et al.</i> 2001 (32); Sato <i>et al.</i> 2019 (33)
		Transcriptome	3,981 (2,316 HCC, 1,665 non-tumorous tissue)	Multiple techniques combined	External validation	0.91–0.96 (range)	Heterogeneity of data due to pooled analysis of 30 studies	Kaur <i>et al.</i> 2020 (34)
		Gene co-expression	57 (38 HCC, 19 normal samples)	PCA	Development only (no validation)	N/A	Use of retrospective databases and no clinical applicability	Zhang <i>et al.</i> 2017 (35)
		miRNA	N/A	Deep belief nets	Internal validation (cross-validation)	F1-score 75.48%	Lack of external validation	Ibrahim <i>et al.</i> 2014 (36)
		Genes	95 (43 tumour, 52 non-tumour)	ANN	Internal validation (cross-validation)	N/A	Proof-of-concept study, use of retrospective gene databases	Gui <i>et al.</i> 2015 (37)
		Urine	15 samples	PCA, random forest	Internal validation	0.903	Small sample size	Liang <i>et al.</i> 2016 (38)
		Biomarker identification from literature	N/A	Data mining	Internal validation (cross-validation)	F-score 0.89	Use of impact factor as scoring tool, introducing systematic bias into the results	Chang <i>et al.</i> 2017 (39)

HCC, hepatocellular carcinoma; AI, artificial intelligence; AUC, area under the curve; 95% CI, 95% confidence intervals; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; HepB, viral hepatitis type B; HepC, viral hepatitis type C; ANN, artificial neural network; PCA, principal component analysis; miRNA, microRNA; N/A, non-applicable; CT, computed tomography; SPIO-MRI, superparamagnetic iron oxide magnetic resonance imaging.

Chronic viral hepatitis B and C infections are both associated with the development of HCC (40,41). Similarly, non-viral causes, including non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH), cirrhosis and fibrosis of the liver are also risk factors for developing HCC (42). The stepwise progression of these pathologies creates the potential for a screening window, during which high-risk individuals can be identified.

The use of AI-driven solutions in detection of NAFLD and NASH has not been comprehensively researched. In a 2016 study, Chiappini and colleagues investigated serum and liver lipids in NAFLD and NASH murine models, using supervised ML (random forest analysis). They identified unique signatures of NASH, opening new possibilities of pre-HCC change detection (22). Population screening for NAFLD has also been suggested, with ML-algorithm, based on 23 clinical parameters, achieving an AUC of 0.88 (95% CI, 0.84–0.91) in detecting NAFLD (23).

Substantially more research has been made into the association of chronic HepB and HepC infections and development of cirrhosis, fibrosis and eventually HCC. Progression of cirrhosis into HCC was studied as early as 2012; it was found that a supervised ML (random forest) model outperformed conventional regression analysis, achieving a c-index of 0.64 (95% CI, 0.6–0.69) (24). Konerman and colleagues used a very similar model to predict progression of chronic hepatitis C infection into fibrosis, with AUC of 0.86 (95% CI, 0.85–0.87) (25).

Throughout the years, as ML models have become more sophisticated, their diagnostic performance has improved. An ANN model by Reddy and Imler, achieved an AUC as high as 0.962 in the prediction of malignant transformation from hepatitis B or C chronic infection (26). Similarly, another recurrent neural network model achieved an AUC of 0.89 (27). It is worth highlighting that in all cases, ML models have proven to be superior to the classical statistical regression model when analysing big data sets.

Moreover, most recent model by Książek *et al.*, used patient characteristics, such as viral status, presence of comorbidities and laboratory results to predict the development of HCC based on 23 quantitative and 26 qualitative features, has achieved 88.5% accuracy using this approach (28).

Imaging

Detection is the most basic way of utilising AI in imaging, while more novel approaches focus on characterisation and

stratification of suspicious imaging regions. While more advanced applications of AI in imaging will be discussed later in this review, it is important to consider the detection of HCC lesions on imaging, which were the precursors to current AI applications.

Lee *et al.* used computer-aided diagnosis (CAD) on multi-phase CT scans to detect HCC in a set of 15 moderate HCC (mean \varnothing 3.1 cm) and 10 small HCC (mean \varnothing 1.04 cm). Using a non-rigid registration model, which accounted for deformation between phases due to respiratory movements and heartbeats, they have achieved 100% detection accuracy, when compared with radiological diagnosis (29). Similar results were obtained using 3D non-linear image wrapping (30). Moreover, in 2009 Guo *et al.* have looked into HCC nodule detection on rat models, using SPIO-enhanced MRI, achieving 91.67% classification accuracy (31).

Biomarkers

A variety of biomarkers have been researched, to equip clinicians with a reliable tool for HCC detection. Recently advances in bioinformatics and technology, resulting in have revolutionised the way large biological datasets ('-omics' datasets) can be generated and analysed, allowing for the integration of multiple datasets (genome, proteome, transcriptome, etc.) (43). Combination of 'omics' with AI algorithms has led to the identification of suitable biomarkers with the potential to translate data into therapeutics (44). The use of biomarkers, identify with the aid of neural networks, combined with the classically used serological marker for HCC (alpha-fetoprotein), was shown to increase diagnostic sensitivity from 60% to 73.8% while maintaining the sensitivity of 88.2% (32).

Harnessing the power of ML has also allowed for new types of clinical data to be analysed in the hope of detecting HCC. Large-scale (2,316 HCC tumour samples and 1,665 non-tumorous tissue samples) transcriptomic profiling by Kaur *et al.* derived a 3-gene signature (*FCN3*, *CLEC1B*, and *PRCI*), which has achieved an AUC of 0.91–0.96 on its validation set, which used peripheral blood samples containing mononuclear cells of both HCC and healthy patients (34). Further, each one of the three genes, showed prognostic character, as their expression levels, when stratified to greater than mean and lesser than mean groups, correlated with overall survival, progression-free survival (PFS) and disease-free survival (DFS) on univariate analysis.

A wide array of HCC biomarkers have been investigated

using AI-driven techniques: gene co-expression patterns (35), miRNA (36), HCC-related genes (37) and serum biomarkers (33) have all been suggested as potential diagnostic signatures. Urine biomarkers have also been identified using ML approaches—Liang and colleagues identified five differential metabolites, from 37 urine samples (25 early-HCC and 12 controls), showing a sensitivity of 96.5% and specificity of 83% in differentiating between healthy and HCC patients, on an independent validation set (n=25) (38).

Another application of AI is data mining, which can be used to automatically screen existing literature to identify biomarker candidates, making it a useful tool for bioinformaticians (39).

Characterisation

The summary of literature on HCC characterisation, based on biomarker, imaging and pathology is presented in *Table 3*.

Biomarkers

Although the primary use of biomarkers is the detection of HCC, ML approaches also allow for stratification of HCC patients based on their biomarker profile, which can have therapeutic implications, as the response to treatment can be dependent on HCC subtype.

Genomics and epigenomics (DNA methylation patterns) analysis using ML has allowed for accurate differentiation between early-stage (stage I) and late-stage (stages II-IV) HCC (45). Moreover, Estevez *et al.* have established a biomarker-based classification into HepB-HCC, HepC-HCC and non-viral HCC, using cytokine profile from serum samples (46). Such stratification can have clinical implications for management, as well as understanding differences in disease pathogenesis.

Imaging

AI can help analyse radiological features from ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI), all of which are routinely used in the diagnosis and differentiation of liver pathology.

The first use of AI in HCC characterisation based on imaging focused on the region of interest (ROI) analysis and computer-aided diagnosis systems using US and (47), CT (51) or MRI (55). These simple models employed analysis of features, such as lesion border or texture, to

differentiate between benign and malignant liver tumour and thus identify HCC. Nowadays, more sophisticated AI-driven solutions are used in lesion differentiation, with greater accuracy and better differentiation, staging and stratification abilities.

For ultrasound scan, differentiation between cirrhotic liver and HCC, using neural networks can be made with 94.5% classification accuracy (48). Differentiation between atypical HCC and focal nodular hyperplasia (FNH) on the contrast-enhanced US, have also been reported, achieving 94.4% classification accuracy, when compared against pathology report analysis (biopsy or resection) and subsequent clinical follow-up (49). High differentiation accuracy has also been found for CT and MRI scans (52,64). What is more, Yamashita *et al.* have proven the feasibility of CNN assigning Li-RADS grades [an HCC CT/MRI scan probability classification used by American College of Radiologists (65)], to guide treatment decision-making (66).

AI models can also assist in imaging-based grading of HCC. Using contrast-enhanced ultrasound (CEUS), Sugimoto *et al.* established classification into well-differentiated, moderately differentiated and poorly differentiated HCC, with an AUC of 0.863–0.872 (50). Similar differentiation based on MRI scans was shown by Zhou *et al.* (56).

Furthermore, tumour segmentation algorithms have been employed to aid in management planning. Visualisation of the tumour can dictate decisions regarding tumour extent and resection. These algorithms, based on contrast-enhanced CT scans, can provide clinically useful 3D projections with a high degree of accuracy, as shown by Li *et al.* (53).

Most recently, radiomics approach to imaging analysis has been proposed, involving a multi-step process to derive large datasets of radiological features, via image acquisition, segmentation, feature extraction and automated analysis of patterns by using high throughput computing (67). Studies utilising this methodology shed light on the future directions of AI in imaging for HCC, highlighting its potential for high accuracy tumour characterisation and classification on MRI (57) and multi-phase contrast-enhanced CT (54) by analysing of textural features.

Pathology

AI has been used in pathology in order to precisely analyse the results of biopsies and resections to help with lesion characterisation and differentiation, using image-analysis

Table 3 Summary of literature on HCC characterisation, based on biomarkers, imaging and pathology

Domain	Sub-category	Notes	N	AI algorithm	Type of validation	AUC (95% CI)	Limitations	Reference
Biomarkers	N/A	Genomics and epigenomics for HCC staging	400 (173 early stage, 177 late stage, 50 normal)	Support machine vector	Internal validation (cross-validation)	0.99 (0.98–0.99)	Biomarkers derived from tissue, which requires invasive approaches for sample isolation	Kaur <i>et al.</i> 2019 (45)
		Cytokine profile for HepC-HCC, HepB HCC and non-viral HCC differentiation	411 (102 HCC, 309 normal)	Random forest	Development only (no validation)	0.90	Small sample size and comparison groups not adjusted for ethnicity	Estavez <i>et al.</i> 2017 (46)
Imaging	US	Differentiation between focal hepatic lesions (benign and malignant)	51 images	ROI analysis	Development only (no validation)	N/A	Only two data validators (radiologist); data possibly not generalisable	Kim <i>et al.</i> 2009 (47)
		Differentiation between cirrhosis and HCC	189 images	ANN	Internal validation (cross-validation)	Accuracy 94.5%	Small sample size	Bharti <i>et al.</i> 2018 (48)
		Differentiation between atypical HCC and focal nodular hyperplasia	257 images	ANN	Internal validation (cross-validation)	F1-score 94.62%	Small sample size leading to lack of generalisability and network which is difficult to interpret	Huang <i>et al.</i> 2020 (49)
		Grading based on tumour differentiation	232 (76 well-differentiated HCC, 133 moderately differentiated HCC, 23 poorly differentiated HCC)	ANN	Development only (no validation)	Accuracy 87.5%	Use of 2D ultrasound and fine-needle biopsy specimen for establishing differentiation instead of surgical specimen	Sugimoto <i>et al.</i> 2016 (50)
	CT	Differentiation between focal hepatic lesions (benign and malignant)	147 images	Region of interest (ROI) analysis	Internal validation (cross-validation)	Accuracy 84.96%	Small sample size	Mougiakakou <i>et al.</i> 2007 (51)
		HCC diagnosis from nodular, diffuse and massive tumours	165 (46 diffuse tumour, 43 nodular tumours, 76 massive tumours)	Convolutional neural networks (CNN)	Internal validation (random sample split)	Accuracy 98.4–99.7% (range)	Segmentation performance for diffuse tumour is not as good as other types, creating noise in the data	Li <i>et al.</i> 2020 (52)
		Tumour segmentation on contrast-enhanced enhanced CT	201 images	Fully convolutional neural networks	External validation	Accuracy 93.7%	Network which is difficult to interpret and restricted by GPU memory	Li <i>et al.</i> 2018 (53)
		Differentiation between five phases of CT	502 images	Random forest	External validation	Accuracy 84–98% (range)	Overlap between five phases on CT scan (no clear guidelines on start and end of each phase); decision based on expertise of principal investigators	Dercle <i>et al.</i> 2020 (54)
	MRI	Differentiation between focal hepatic lesions (benign and malignant)	320 images	ANN, CNN	Development only (no validation)	Accuracy 93%	Single centre character of the studies and sample size insufficient for neural network training	Zhang <i>et al.</i> 2009 (55)
		Grading based on tumour differentiation	100 (47 low grade HCC, 53 high grade HCC)	CNN	Internal validation (random sample split)	0.73–0.83 (range)	Single-centre character of the study, small sample size and lack of external validation	Zhou <i>et al.</i> 2019 (56)
Focal lesion differentiation using texture and topological analysis		150 (50 HCC, 50 metastatic tumours, 50 hepatic haemangioma)	ROI analysis	Internal validation (cross-validation)	0.75–0.95 (range)	Large variation in tumour size, affecting classification accuracy for the outliers	Oyama <i>et al.</i> 2019 (57)	
Pathology	N/A	Differentiation between HCC and cholangiocarcinoma	106 whole-slide images	Deep learning	Internal validation (random sample split)	0.842	Methodology not reflective of clinical practice, reducing general applicability of the results	Kiani <i>et al.</i> 2020 (58)
		Differentiation between healthy tissue and HCC	1,773 image features	Random forest	External validation	0.886	Different in ethological factors between two datasets used	Liao <i>et al.</i> 2020 (59)
		Grading of HCC	109 patients	ROI analysis, fractal dimensions	Internal validation (cross-validation)	Accuracy 95.97% (Atupelage <i>et al.</i> 2014); accuracy 90.51% (Atupelage <i>et al.</i> 2013)	Small sample size and lack of external validation. Inclusion of non-informative texture features into the classifiers	Atupelage <i>et al.</i> 2014 (60); Atupelage <i>et al.</i> 2013 (61)
		HCC diagnosis using hyperspectral imaging analysis	14 hyperspectral images	Deep learning	Internal validation (cross-validation)	0.950	Small sample size and single-centre character of the study	Wang <i>et al.</i> 2020 (62)
		HCC grading using multiphoton microscopy	217 images	CNN	Internal validation (cross-validation)	0.941(0.913–0.968)	Small sample size, insufficient for deep learning purpose	Lin <i>et al.</i> 2019 (63)

HCC, hepatocellular carcinoma; AI, artificial intelligence; AUC, area under the curve; 95% CI, 95% confidence intervals; HepB, viral hepatitis type B; HepC, viral hepatitis type C; ANN, artificial neural network; ROI, region of interest; CNN, convolutional neural network; N/A, non-applicable; CT, computed tomography; MRI, magnetic resonance imaging, US, ultrasound scan.

techniques.

Whole-slide images using H&E stain were analysed by Kiani *et al.* to differentiate between cholangiocarcinoma and HCC, achieving 84.2% differentiation accuracy on an independent validation set (58). When this model was offered as assistance to experienced histopathologists, it increased accuracy when correct, but decreased accuracy when incorrect, bringing attention to possible negative consequences of introducing AI into clinical practice (58).

Such models, on top of having high diagnostic accuracy (88.6% in differentiation accuracy between healthy liver tissue and HCC), can also have prognostic value, predicting response to treatment, a crucial characteristic for an AI-driven model to have in the future, to facilitate its clinical applicability (59).

Pathological slides can also be used for HCC grading (low-grade HCC *vs.* high-grade HCC) based on the number of nuclei (60) or whole-slide texture analysis (61).

Finally, one of the future applications of AI-driven solutions is lab-free, real-time pathological diagnosis, suggested by Wang *et al.* (62). The deep learning model yielded 88.1% accuracy, using hyperspectral imaging instead of the gold standard frozen sections; a method which is both time and labour-intensive. Lin *et al.* investigated another alternative to lab-based tissue processing, multiphoton microscopy (MPM), which also shows potential, as the use of CNN resulted in 90% differentiation accuracy into low-, moderate- and high-grade HCC (63).

Prediction

The summary of literature on HCC prediction, based on treatment outcomes and overall survival is presented in *Table 4*.

Prediction of treatment outcomes

AI algorithms were successfully developed to predict response to and survival after transarterial chemoembolization (TACE). To forecast survival after TACE, an ANN model was produced based on all parameters used by ART, ABCR and SNACOR risk scoring systems, as well as, age, gender, type of TACE and type of imaging before the second TACE. When compared with abovementioned scores, risk prediction made by the ANN (AUC 0.83±0.06) was found to be significantly superior ($P<0.001$) to that of ART (AUC 0.54±0.08) (68). ANN model has also outperformed SNACOR and ABCR, however, the difference was not

significant ($P=0.201$ and $P=0.143$ respectively). All of the other studies predicting response to TACE utilised analysis of imaging (69-71,87,88). When a combination of extracted MRI features, clinical information and therapeutic features were used to train logistic regression and random forest models, to classify patients as responders or non-responders, overall accuracy of 78% was reached (69). Moreover, two research groups developed computer tomography based convolutional neural network models (70,88). They predicted time to progression (TTP) based on follow-up CT radiological criteria (mRECIST), to divide patients as TACE-susceptible (TTP >14 weeks) or TACE-refractory (TTP <14 weeks), and classified patients into one of four groups: complete TACE response, partial response, stable disease or progressive disease respectively. The models achieved 74.2%, followed by 84.3% accuracy with AUC scores of 0.95–0.97 for individual prediction categories.

Finally, Liu and colleagues validated three (one deep and two ML), predictive models, based on radiomic features of CEUS scans (71). Deep learning model outperformed the other two in assigning patients in the validation cohort to either objective-response to TACE or non-response, reaching 0.93 AUC score.

AI solutions were also used to predict stereotactic body radiotherapy (SBRT) and radiofrequency ablation (RFA) outcomes. To challenge currently used dose-volume histogram (DVH) based metric and forecast hepatobiliary toxicity of SBRT more accurately, Ibragimov *et al.* trained a CNN model on healthy organ CT images, liver SBRT cases and nanodosimetric pre-treatment patient features (72). The approach used was general with only 36 HCC patients out of 125 cases in total. CNN model was moderately successful with AUC of 0.76 and overall accuracy similar to that of current DVH metric. Enriching the liver SBRT database and checking the performance of the toxicity prediction framework may improve the future performance of the model.

To predict 1- and 2-year DFS of patients who underwent CT-guided percutaneous RFA used in early HCC stages, Wu and colleagues developed an ANN-based on 15 clinical features (73). The model was more accurate when anticipating 1-year DFS than 2-year DFS, with 85.0% and 67.9% accuracy, respectively.

Several AI predictive models were also validated to explore post-resection survival and HCC recurrence. Currently, the risk is calculated based on the histological analysis of resected specimens (74). Deep CNN models were shown to optimise this process by implementing the

Table 4 Summary of literature on HCC prediction based on treatment outcomes and overall survival

Domain	Sub-category	Notes	N	AI algorithm	Type of validation	AUC (95% CI)	Limitations	Reference
Treatment outcomes	TACE	Response to treatment based on clinical data and risk scoring systems	282	ANN	Internal validation (random sample split)	0.83±0.06	No independent external validation cohort	Mähringer-Kunz 2020 (68)
		Response to treatment based on MRI and clinical data	36 HCC patients	Random forest	Internal validation (cross-validation)	Accuracy 78%	Small patient cohort	Abajian <i>et al.</i> 2018 (69)
		Response to treatment based on CT images	105 patients	CNN	Internal validation (cross-validation)	Accuracy 74.2% (64–82%)	General model applied to multiple TACE chemotherapy regimens	Morshid <i>et al.</i> 2019 (70)
		Response to treatment based on the contrast-enhanced US	130 patients	Deep learning	Internal validation (random sample split)	0.93 (0.80–0.98)	Limited sample size; single-centre retrospective data	Liu <i>et al.</i> 2020 (71)
	SBRT	Hepatobiliary toxicity prediction based on CT images	125 patients and 2,644 images of human organs	CNN	Internal validation (cross-validation)	0.85	Limited liver SBRT database	Ibragimov <i>et al.</i> 2018 (72)
	RFA	Disease-free survival prediction	252 1-year and 179 2-year DFS	ANN	Internal and external validation	0.84 for 1-year, 0.75 for 2-year DFS prediction	Uneven 1-year and 2-year DFS group sizes	Wu <i>et al.</i> 2017 (73)
	Resection	Recurrence risk based on whole-slide histological analysis	522 patients	CNN	Internal and external validation	c-index 0.70	Overfitting (inferior performance on external validation set)	Salliard <i>et al.</i> 2020 (74)
		Recurrence and progression-free survival based on immunological biomarkers	221 patients	Random forest	Internal validation (cross-validation)	0.80	No external dataset validation	Zhou <i>et al.</i> 2019 (75)
		Survival based on CT images	470 patients	ANN	Internal and external validation	0.803	Most patients had hepatitis B-related HCC	Ji <i>et al.</i> 2019 (76)
			167 patients		Internal validation (cross-validation)	0.825	No external dataset validation	Wang <i>et al.</i> 2019 (77)
			995 patients	Bayesian network	Internal validation (cross-validation)	Accuracy 57%	Lack of temporal information in the patient data	Xu <i>et al.</i> 2019 (78)
	Microinvasion based on biomarkers and MRI	160 patients	Logistic regression	Internal validation (random sample split)	0.83 (0.71–0.95)	Normalisation of the signal intensities on MR images not performed	Feng <i>et al.</i> 2019 (79)	
	Survival based on BCLC criteria	976 patients	Classification and Regression Tree	Internal validation (random sample split)	c-index 0.604	Majority of patients had favourable liver function	Tsilmigras <i>et al.</i> 2020 (80)	
Transplantation	Recurrence based on clinical data and CT images	133 patients	Classification and Regression Tree	Internal validation (random sample split)	c-index 0.789 (0.620–0.957)	Retrospective design	Guo <i>et al.</i> 2019 (81)	
Overall survival	N/A	Survival based on biomarkers in HepB-HCC	67 samples (40 patients)	Supervised ML	Internal validation (cross-validation)	Accuracy 78%	Small dataset	Ye <i>et al.</i> 2003 (82)
		Survival based on DNA methylation patens	488 samples	Multiple techniques combined	External validation	Accuracy 63%	Reason for 40% of all patients being hard to predict remained unclear	Itzel <i>et al.</i> 2019 (83)
			377 HCC, 50 control samples		Internal validation (cross-validation)	0.95 (mean 10-fold cross-validation score)	Limited validation outcomes reported	Dong <i>et al.</i> 2019 (84)
		Survival based on gene-expression pathways	355 patients	Support vector machine	Internal and external validation	c-index 0.83	Class label of the TCGA HCC samples obtain using whole TCGA dataset	Fa <i>et al.</i> 2019 (85)
		Survival based on clinical features	165 patients	ANN	Internal validation (cross-validation)	0.700	No external dataset validation	Santos <i>et al.</i> 2015 (86)

HCC, hepatocellular carcinoma; AI, artificial intelligence; AUC, area under the curve; 95% CI, 95% confidence intervals; HepB, viral hepatitis type B; ANN, artificial neural network; CNN, convolutional neural network; TACE, transarterial chemoembolization; SBRT, stereotactic body radiation therapy; RFA, radiofrequency ablation; BCLC criteria, Barcelona Clinic Liver Cancer criteria; DFS, disease-free survival; TCGA, The Cancer Genome Atlas; CT, computed tomography; MRI, magnetic resonance imaging.

analysis of whole-slide digitised histological slides (74). Two CNN algorithms reached similar efficiency, however, the combination of CNN with human input (tumour areas annotated by the pathologist) slightly outperformed the one without (AUC 0.78 *vs.* 0.75).

Immunological tumour biomarkers were also used as a tool for predicting survival, using three indices: Overall Survival (≤ 24 or >24 months), PFS (≤ 6 or >6 months) and recurrence/death producing AUC between 0.76 and 0.8 and accuracy over 85% (75).

Moreover, groups led by Ji and Wang validated CT-based ANN and deep CNN to predict survival (76,77). The first group developed a novel three-feature radiomic signature of contrast-enhanced CT image, where performance was improved by combining it with clinical features (c-index 0.63–0.69 *vs.* 0.73–0.801). Wang and colleagues employed multi-phase CT radiomics features together with clinical models to yield a combined model with AUC of 0.82. A Bayesian network-based approach was also used to predict the probability of post-resection HCC recurrence which considered respective recurrence evolution paths for clinical feature datasets (78).

At the same time, AI techniques were explored to provide information about the predictive power of particular biomarkers, which could guide decisions on liver resection. Feng *et al.* used MRI radiomics to predict microvascular invasion status of the tumours, an important factor for hepatectomy, reaching AUC of 0.83 for validation dataset (79). Furthermore, an AI model was applied to determine the prognostic weight of factors comprising the Barcelona Clinic Liver Cancer (BCLC) guidelines, which selected alpha-fetoprotein and Charlson comorbidity score as the most important preoperative factors of overall survival among BCLC-0/A patients, and radiologic tumour burden score for BCLC-B patients (80). These results have the potential to shape the next iteration of BCLC guidelines.

AI was also used to predict DFS following liver transplantation. CT radiomics and clinical risk factors were combined to train the model, which yielded c-index of 0.79 when tested on the validation cohort (81).

Prediction of overall survival of HCC patients

ML algorithms have also been used for the general prediction of survival in HCC patient population. The earliest studies used supervised ML to differentiate between metastatic and non-metastatic HCC in HepB

positive patients based on gene expression, hence obtaining information about probable survival chance (82). It also identified osteopontin as a biomarker of metastatic HCC.

Recently, groups led by Itzel and Dong explored possibilities of using random gene sets and DNA methylation levels for survival prognostics (83,84), while Fa *et al.* followed disease-specific patterns in dysregulated gene-expression pathways instead of single genes (85). More studies followed with training predictive algorithms of clinical features of HCC patients to forecast survival. Santos *et al.* combined a cluster-based oversampling method with the neural network model to account for small and incomplete datasets, improving the AUC score from 0.69 to 0.75 (86).

Discussion

AI solutions have been applied in all aspects of medicine in recent years and HCC is no exception. AI has led to advances in detection of HCC (based on pre-malignant changes, imaging and biomarkers) due to its ability to analyse large datasets and integrate information efficiently. Biomarkers identified by the integration of multiple ‘-omics’ datasets are especially promising, potentially leading to the identification of a biochemical tumour signature, revolutionising HCC detection in the future.

As AI algorithms have become more sophisticated, research emphasis shifted towards lesion characterisation, differentiation between various types of hepatic malignancies and stratification of patients into groups, based on the tumour stage or grade. Various datasets, such as radiological images or clinical and pathological data, can be used separately or in combination to provide accuracy superior to that of traditional statistical tools. What is more, AI-driven solutions can help in reducing interobserver variability when analysing imaging studies, leading to standardisation. Since most of the hepatocellular cancers develop on the background of chronic liver disease, future efforts in HCC characterisation should focus on accurate differentiation between pre-malignant changes and early malignancy, to provide the most clinical benefit.

AI methods for prediction of overall survival and treatment outcomes in HCC have emerged in the past two years and remain a dynamic area of study. Predictive potential of current models is higher for short-term outcomes rather than long-term survival, however, this approach offers an array of novel predictive tools to shape HCC guidelines and support clinical decision making. In

the future, models combining radiomic data with clinical features to provide characterisation and prognosis for HCC patients are likely to be implemented in the clinical practice to support decisions on treatment.

Limitations

AI has inherent limitations, and this holds true for its medical applications for HCC detection, characterisation and prediction. Most of the studies discussed in this review suffer from a small sample size, which is a major issue for deep learning algorithms, as they require large training dataset to perform well. What is more, a lot of the studies use single-institution data, often from tertiary care centres. As a result, despite achieving high AUC and accuracy, such AI algorithms often cannot be used outside a narrow context, ultimately hampering widespread clinical adoption of AI within HCC. Specific limitations also exist in each of the three domains discussed. Studies on detection, focus on specific subtypes of HCC and narrow populations (e.g., HBV positive patients), rendering the proposed AI algorithms are unable to perform screening for HCC on the level of the general population. Studies on characterisation are limited by lack of standardisation of biomarker assays, imaging techniques and histological specimen preparation, all of which contribute to difficulties in applying the results of the research in settings different than the original. Even though significant advancements in the implementation of HCC outcomes prediction have been made in the recent years, there are many outstanding questions. Interpretation of algorithm outcomes remains a major challenge as it is difficult to explain why the model fails to make accurate predictions for a proportion of cases.

This review is also not devoid of limitations. Firstly, the methodology of the study, being a literature review limits the applicability of its conclusions. Moreover, the heterogeneity of quantitative data and AI methodologies has not allowed for pooled analysis of outcomes, but only a qualitative synthesis of evidence. The three-domain classification that was adopted for the purposes for that review is also imperfect, as more recent studies often discuss potential uses of HCC across more than one domain, by utilising multiple types of data to inform clinical decisions. Moreover, selection bias might exist, and studies as studies with negative findings might not be published. Finally, this review only aimed at assessing studies in English, however, high-quality studies in other languages might exist.

Future research directions

Consistency in reporting and transparency in publishing AI algorithms can significantly improve the clinical value of studies exploring AI models applied to HCC. Discrepancies in data standards and diagnostic devices used across different treatment centres contribute to overfitting of AI models, which needs to be overcome to facilitate the development of AI solutions with general applicability (89). External validation of the AI algorithms should also be favoured over retrospective internal validation, further increasing the applicability of the AI-driven solutions. International and interdisciplinary collaboration is instrumental in approaching this issue, as shown in a recent study that investigated an AI model in breast cancer diagnosis using data from both the UK and US (10). What is more, widespread availability of source code for algorithms can speed the process of AI development and validation, also contributing to larger applicability of these solutions. Finally, effective communication between computer scientists, engineers and clinicians is crucial for generating research which can redefine the current practice to address unmet clinical needs.

Conclusions

AI will revolutionise the way we detect and characterise HCC, as well as predict the course of its development, however, it is still experimental. In recent years, the rise of big data has caused AI-driven solutions utilising clinical data, radiological images, biomarkers and pathology results to emerge and gradually improve in accuracy, however, their widespread introduction into the clinical practice has not occurred yet. Robust validation, large scale studies, multicentre cooperation, advocacy for AI and education on AI amongst clinicians are all necessary for AI models to take the next step, so that in the future, such models using multiple data modalities, have the chance of influencing HCC guidelines and shaping clinical practice.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at <https://>

tgh.amegroups.com/article/view/10.21037/tgh-20-242/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tgh.amegroups.com/article/view/10.21037/tgh-20-242/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70:7-30.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
3. Quaglia A. Hepatocellular carcinoma: a review of diagnostic challenges for the pathologist. *J Hepatocell Carcinoma* 2018;5:99-108.
4. Kim JH, Joo I, Lee JM. Atypical appearance of hepatocellular carcinoma and its mimickers: How to solve challenging cases using gadoteric acid-enhanced liver magnetic resonance imaging. *Korean J Radiol* 2019;20:1019-41.
5. Balogh J, Iii DV, Gordon S, et al. Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma* 2016;3:41-53.
6. Tsoulfas G, Agorastou P, Tooulias A, et al. Current and future challenges in the surgical treatment of hepatocellular carcinoma: a review. *Int Surg* 2014;99:779-86.
7. Santopaolo F, Lenci I, Milana M, et al. Liver transplantation for hepatocellular carcinoma: Where do we stand? *World J Gastroenterol* 2019;25:2591-602.
8. Lindsay RK, Buchanan BG, Feigenbaum EA, et al. DENDRAL: A case study of the first expert system for scientific hypothesis formation. *Artif Intell* 1993;61:209-61.
9. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8.
10. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89-94.
11. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, et al. Artificial intelligence in retina. *Prog Retin Eye Res* 2018;67:1-29.
12. Amisha, Malik P, Pathania M, et al. Overview of artificial intelligence in medicine. *J Family Med Prim Care* 2019;8:2328.
13. Deo RC. Machine Learning in Medicine HHS Public Access. *Circulation* 2015;132:1920-30.
14. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J Pharm Biomed Anal* 2000;22;717-27.
15. Yamashita R, Nishio M, Do RKG, et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;9:611-29.
16. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019;19:64.
17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
18. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 2013;4:627-35.
19. Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: Relation to the variance and odds ratio of a continuous explanatory variable. *BMC Med Res Methodol* 2012;12:82.
20. 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.
21. Nanni L, Brahnam S, Brattin R, et al. An introduction to deep learners and deep learner descriptors for medical applications. In: *Deep Learners and Deep Learner Descriptors for Medical Applications*. Springer; 2020:1-5.
22. Chiappini F, Desterke C, Bertrand-Michel J, et al. Hepatic and serum lipid signatures specific to nonalcoholic steatohepatitis in murine models. *Sci Rep* 2016;6:31587.
23. Abstracts of the 26th Annual Conference of APASL, February 15-19, 2017, Shanghai, China. *Hepatol Int* 2017;11:1-1093.

24. Singal AG, Waljee AK, Mukherjee A, et al. Su2057 Machine Learning Algorithms Outperform Conventional Regression Models in Identifying Risk Factors for Hepatocellular Carcinoma in Patients With Cirrhosis. *Gastroenterology* 2012;142:S984.
25. Konerman MA, Zhang Y, Zhu J, et al. Improvement of predictive models of risk of disease progression in chronic hepatitis C by incorporating longitudinal data. *Hepatology* 2015;61:1832-41.
26. Reddy R, Imler TD. Artificial Neural Networks are Highly Predictive for Hepatocellular Carcinoma in Patients with Cirrhosis. *Gastroenterology* 2017;152:S1193.
27. Ioannou GN, Tang W, Beste L, et al. 498 - Deep Learning Models Accurately Predict Development of Hcc in 146,218 Patients with Chronic Hepatitis C. *Gastroenterology* 2019;156:S1201.
28. Książek W, Abdar M, Acharya UR, et al. A novel machine learning approach for early detection of hepatocellular carcinoma patients. *Cogn Syst Res* 2019;54:116-27.
29. Lee J, Kim KW, Kim SY, et al. Automatic detection method of hepatocellular carcinomas using the non-rigid registration method of multi-phase liver CT images. *J Xray Sci Technol* 2015;23:275-88.
30. Okumura E, Sanada S, Suzuki M, et al. Effectiveness of temporal and dynamic subtraction images of the liver for detection of small HCC on abdominal CT images: Comparison of 3D nonlinear image-warping and 3D global-matching techniques. *Radiol Phys Technol* 2011;4:109-20.
31. Guo D, Qiu T, Bian J, et al. A computer-aided diagnostic system to discriminate SPIO-enhanced magnetic resonance hepatocellular carcinoma by a neural network classifier. *Comput Med Imaging Graph* 2009;33:588-92.
32. Poon TCW, Chan ATC, Zee B, et al. Application of classification tree and neural network algorithms to the identification of serological liver marker profiles for the diagnosis of hepatocellular carcinoma. *Oncology* 2001;61:275-83.
33. Sato M, Morimoto K, Kajihara S, et al. Machine-learning Approach for the Development of a Novel Predictive Model for the Diagnosis of Hepatocellular Carcinoma. *Sci Rep* 2019;9:7704.
34. Kaur H, Dhall A, Kumar R, et al. Identification of Platform-Independent Diagnostic Biomarker Panel for Hepatocellular Carcinoma Using Large-Scale Transcriptomics Data. *Front Genet* 2020;10:1306.
35. Zhang C, Peng L, Zhang Y, et al. The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. *Med Oncol* 2017;34:101.
36. Ibrahim R, Yousri NA, Ismail MA, et al. Multi-level gene/MiRNA feature selection using deep belief nets and active learning. *Annu Int Conf IEEE Eng Med Biol Soc* 2014;2014:3957-60.
37. Gui T, Dong X, Li R, et al. Identification of Hepatocellular Carcinoma-Related Genes with a Machine Learning and Network Analysis. *J Comput Biol* 2015;22:63-71.
38. Liang Q, Liu H, Wang C, et al. Phenotypic Characterization Analysis of Human Hepatocarcinoma by Urine Metabolomics Approach. *Sci Rep* 2016;6:19763.
39. Chang NW, Dai HJ, Shih YY, et al. Biomarker identification of hepatocellular carcinoma using a methodical literature mining strategy. *Database (Oxford)* 2017;2017:bax082.
40. Di Bisceglie AM. Hepatitis C and hepatocellular carcinoma. *Hepatology* 1997;26:34S-38S.
41. Di Bisceglie AM. Hepatitis B and hepatocellular carcinoma. *Hepatology* 2009;49:S56.
42. Anstee QM, Reeves HL, Kotsiliti E, et al. From NASH to HCC: current concepts and future challenges. *Nat Rev Gastroenterol Hepatol* 2019;16:411-28.
43. Manzoni C, Kia DA, Vandrovcova J, et al. Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Brief Bioinform* 2018;19:286-302.
44. Chen B, Garmire L, Calvisi DF, et al. Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 2020;17:238-51.
45. Kaur H, Bhalla S, Raghava GPS. Classification of early and late stage liver hepatocellular carcinoma patients from their genomics and epigenomics profiles. *PLoS One* 2019;14:e0221476.
46. Estevez J, Chen VL, Podlaha O, et al. Differential Serum Cytokine Profiles in Patients with Chronic Hepatitis B, C, and Hepatocellular Carcinoma. *Sci Rep* 2017;7:11867.
47. Kim SH, Lee JM, Kim KG, et al. Computer-aided image analysis of focal hepatic lesions in ultrasonography: Preliminary results. *Abdom Imaging* 2009;34:183-91.
48. Bharti P, Mittal D, Ananthasivan R. Characterization of chronic liver disease based on ultrasound images using the variants of grey-level difference matrix. *Proc Inst Mech Eng H* 2018;232:884-900.
49. Huang Q, Pan F, Li W, et al. Differential Diagnosis of Atypical Hepatocellular Carcinoma in Contrast-Enhanced Ultrasound Using Spatio-Temporal Diagnostic Semantics. *IEEE J Biomed Health* 2020. doi: 10.1109/

- JBHI.2020.2977937.
50. Sugimoto K, Shiraishi J, Tanaka H, et al. Computer-aided diagnosis for estimating the malignancy grade of hepatocellular carcinoma using contrast-enhanced ultrasound: an ROC observer study. *Liver Int* 2016;36:1026-32.
 51. Mouggiakakou SG, Valavanis IK, Nikita A, et al. Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. *Artif Intell Med* 2007;41:25-37.
 52. Li J, Wu Y, Shen N, et al. A fully automatic computer-aided diagnosis system for hepatocellular carcinoma using convolutional neural networks. *Biocybern Biomed Eng* 2020;40:238-48.
 53. Li X, Chen H, Qi X, et al. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Trans Med Imaging* 2018;37:2663-74.
 54. Dercle L, Ma J, Xie C, et al. Using a single abdominal computed tomography image to differentiate five contrast-enhancement phases: A machine-learning algorithm for radiomics-based precision medicine. *Eur J Radiol* 2020;125:108850.
 55. Zhang X, Kanematsu M, Fujita H, et al. Application of an artificial neural network to the computer-aided differentiation of focal liver disease in MR imaging. *Radiol Phys Technol* 2009;2:175-82.
 56. Zhou W, Wang G, Xie G, et al. Grading of hepatocellular carcinoma based on diffusion weighted images with multiple b-values using convolutional neural networks. *Med Phys* 2019;46:3951-60.
 57. Oyama A, Hiraoka Y, Obayashi I, et al. Hepatic tumor classification using texture and topology analysis of non-contrast-enhanced three-dimensional T1-weighted MR images with a radiomics approach. *Sci Rep* 2019;9:8764.
 58. 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.
 59. Liao H, Xiong T, Peng J, et al. Classification and Prognosis Prediction from Histopathological Images of Hepatocellular Carcinoma by a Fully Automated Pipeline Based on Machine Learning. *Ann Surg Oncol* 2020;27:2359-69.
 60. Atupelage C, Nagahashi H, Kimura F, et al. Computational hepatocellular carcinoma tumor grading based on cell nuclei classification. *J Med Imaging (Bellingham)* 2014;1:034501.
 61. Atupelage C, Nagahashi H, Yamaguchi M, et al. Computational grading of hepatocellular carcinoma using multifractal feature description. *Comput Med Imaging Graph* 2013;37:61-71.
 62. Wang R, He Y, Yao C, et al. Classification and Segmentation of Hyperspectral Data of Hepatocellular Carcinoma Samples Using 1-D Convolutional Neural Network. *Cytometry A* 2020;97:31-8.
 63. Lin H, Wei C, Wang G, et al. Automated classification of hepatocellular carcinoma differentiation using multiphoton microscopy and deep learning. *J Biophotonics* 2019;12:e201800435.
 64. Wang CJ, Hamm CA, Savic LJ, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. *Eur Radiol* 2019;29:3348-57.
 65. Chernyak V, Fowler KJ, Kamaya A, et al. Liver Imaging Reporting and Data System (LI-RADS) version 2018: Imaging of hepatocellular carcinoma in at-risk patients. *Radiology* 2018;289:816-30.
 66. Yamashita R, Mittendorf A, Zhu Z, et al. Deep convolutional neural network applied to the liver imaging reporting and data system (LI-RADS) version 2014 category classification: a pilot study. *Abdom Radiol (NY)* 2020;45:24-35.
 67. Kumar V, Gu Y, Basu S, et al. Radiomics: The process and the challenges. *Magn Reson Imaging* 2012;30:1234-48.
 68. 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.
 69. Abajian A, Murali N, Savic LJ, 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.
 70. 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.
 71. 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.
 72. Ibragimov B, Toesca D, Chang D, et al. Development of deep neural network for individualized hepatobiliary toxicity prediction after liver SBRT. *Med Phys* 2018;45:4763-74.
 73. Wu CF, Wu YJ, Liang PC, et al. Disease-free survival assessment by artificial neural networks for hepatocellular

- carcinoma patients after radiofrequency ablation. *J Formos Med Assoc* 2017;116:765-73.
74. Saillard C, Schmauch B, Laifa O, et al. Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides. *Hepatology* 2021;73:2077-8.
75. Zhou W, Chen H, Han W, et al. Prediction of hepatocellular carcinoma patient survival using machine learning classification rules. *J Clin Oncol* 2019;37:e15649-e15649.
76. 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.
77. Wang W, Chen Q, Iwamoto Y, et al. Deep Learning-Based Radiomics Models for Early Recurrence Prediction of Hepatocellular Carcinoma with Multi-phase CT Images and Clinical Data. *Annu Int Conf IEEE Eng Med Biol Soc* 2019;2019:4881-4.
78. Xu D, Sheng JQ, Hu PJH, et al. Predicting hepatocellular carcinoma recurrences: A data-driven multiclass classification method incorporating latent variables. *J Biomed Inform* 2019;96:103237.
79. Feng ST, Jia Y, Liao B, et al. Preoperative prediction of microvascular invasion in hepatocellular cancer: a radiomics model using Gd-EOB-DTPA-enhanced MRI. *Eur Radiol* 2019;29:4648-59.
80. Tsilimigras DI, Mehta R, Moris D, et al. Utilizing Machine Learning for Pre- and Postoperative Assessment of Patients Undergoing Resection for BCLC-0, A and B Hepatocellular Carcinoma: Implications for Resection Beyond the BCLC Guidelines. *Ann Surg Oncol* 2020;27:866-74.
81. Guo D, Gu D, Wang H, et al. Radiomics analysis enables recurrence prediction for hepatocellular carcinoma after liver transplantation. *Eur J Radiol* 2019;117:33-40.
82. Ye QH, Qin LX, Forgues M, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med* 2003;9:416-23.
83. Itzel T, Spang R, Maass T, et al. Random gene sets in predicting survival of patients with hepatocellular carcinoma. *J Mol Med* 2019;97:879-88.
84. Dong RZ, Yang X, Zhang XY, et al. Predicting overall survival of patients with hepatocellular carcinoma using a three-category method based on DNA methylation and machine learning. *J Cell Mol Med* 2019;23:3369-74.
85. Fa B, Luo C, Tang Z, et al. Pathway-based biomarker identification with crosstalk analysis for robust prognosis prediction in hepatocellular carcinoma. *EBioMedicine* 2019;44:250-60.
86. Santos MS, Abreu PH, García-Laencina PJ, et al. A new cluster-based oversampling method for improving survival prediction of hepatocellular carcinoma patients. *J Biomed Inform* 2015;58:49-59.
87. Abajian A, Murali N, Savic LJ, et al. Predicting treatment response to image-guided therapies using machine learning: An example for trans-arterial treatment of hepatocellular carcinoma. *J Vis Exp* 2018;2018:e58382.
88. 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.
89. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15:1929-58.

doi: 10.21037/tgh-20-242

Cite this article as: Kawka M, Dawidziuk A, Jiao LR, Gall TMH. Artificial intelligence in the detection, characterisation and prediction of hepatocellular carcinoma: a narrative review. *Transl Gastroenterol Hepatol* 2022;7:41.