## Contents

**OPINION REVIEW**

1816  Non-alcoholic fatty liver disease in irritable bowel syndrome: More than a coincidence?

*Purssell H, Whorwell PJ, Athwal VS, Vasant DH*

**REVIEW**

1828  Liver-side of inflammatory bowel diseases: Hepatobiliary and drug-induced disorders

*Mazza S, Soro S, Verga MC, Elvo B, Ferretti F, Cereatti F, Drago A, Grassia R*

1850  Gastrointestinal and hepatic side effects of potential treatment for COVID-19 and vaccination in patients with chronic liver diseases

*Law MF, Ho R, Law KWT, Cheung CKM*

1875  Genotype E: The neglected genotype of hepatitis B virus

*Ingasia LAO, Wose Kinge C, Kramvis A*

**MINIREVIEWS**

1892  One stop shop approach for the diagnosis of liver hemangioma

*Sandulescu LD, Urhut CM, Sandulescu SM, Ciurea AM, Cazacu SM, Iordache S*

1909  Liver function in COVID-19 infection

*Przekop D, Gruszewska E, Chrostek L*

1919  Potential role of noninvasive biomarkers during liver fibrosis


1936  Imaging evaluation of the liver in oncology patients: A comparison of techniques

*Freitas PS, Janicas C, Veiga J, Matos AP, Herédia V, Ramalho M*

1956  Liver manifestations and complications in inflammatory bowel disease: A review

*Gaspar R, Branco CC, Macedo G*

1968  Dengue hemorrhagic fever and the liver

*Leowattana W, Leowattana T*

1977  Artificial Intelligence in hepatology, liver surgery and transplantation: Emerging applications and frontiers of research

*Veerankutty FH, Jayan G, Yadav MK, Manoj KS, Yadav A, Nair SRS, Shabeerali TU, Yeldho V, Sasidharan M, Rather SA*

1991  De novo and recurrence of metabolic dysfunction-associated fatty liver disease after liver transplantation

*Han MAT, Olivo R, Choi CJ, Pyrsopoulos N*
## Contents

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Liver dysfunction as a cytokine storm manifestation and prognostic factor for severe COVID-19</td>
<td>Taneva G, Dimitrov D, Velikova T</td>
</tr>
<tr>
<td>2013</td>
<td>COVID-19 and the liver: A brief and core review</td>
<td>Kayaaslan B, Guner R</td>
</tr>
<tr>
<td>2024</td>
<td>Newer variants of progressive familial intrahepatic cholestasis</td>
<td>Vinayagamoorthy V, Srivastava A, Sarma MS</td>
</tr>
<tr>
<td>2039</td>
<td>Deep learning in hepatocellular carcinoma: Current status and future perspectives</td>
<td>Ahn JC, Qureshi TA, Singal AG, Li D, Yang JD</td>
</tr>
</tbody>
</table>

### ORIGINAL ARTICLE

#### Basic Study

- **2052** Gut dysbiosis and systemic inflammation promote cardiomyocyte abnormalities in an experimental model of steatobehatitis  
  Longo L, Rampelotto PH, Filippi-Chiela E, de Souza VEG, Salvati F, Cerski CT, da Silveira TR, Oliveira CP, Uribe-Cruz C, Alvares-da-Silva MR

#### Case Control Study

- **2071** Leukocyte cell-derived chemotaxin-2 and fibroblast growth factor 21 in alcohol-induced liver cirrhosis  
  Sak JJ, Prystupa A, Kiciński P, Luchowska-Kocot D, Kuryś-Denis E, Bis-Wencel H

#### Retrospective Study

- **2081** Biliary complications in recipients of living donor liver transplantation: A single-centre study  

- **2084** Liver function tests and metabolic-associated fatty liver disease: Changes in upper normal limits, does it really matter?  
  Forlano R, Mullish BH, Dhar A, Goldin RD, Thersz M, Manousou P

- **2013** Use of oral vancomycin in children with autoimmune liver disease: A single centre experience  
  Di Giorgio A, Tulone A, Nicaso E, Norsa L, Sonzogni A, D’Antiga L

- **2128** Trends of alcoholic liver cirrhosis readmissions from 2010 to 2018: Rates and healthcare burden associated with readmissions  

### Observational Study

- **2137** New stem cell autophagy surrogate diagnostic biomarkers in early-stage hepatocellular carcinoma in Egypt: A pilot study  
  Yosef T, Ibrahim WA, Matboli M, Swilam AA, El-Nakeepe S

- **2150** Determination of “indeterminate score” measurements in lean nonalcoholic fatty liver disease patients from western Saudi Arabia  
  Khayyat YM
## Contents

**Managing liver transplantation during the COVID-19 pandemic: A survey among transplant centers in the Southeast United States**


**Prospective Study**

**Accuracy of virtual chromoendoscopy in differentiating gastric antral vascular ectasia from portal hypertensive gastropathy: A proof of concept study**

Al-Taee AM, Cubillan MP, Hinton A, Sobotka LA, Befeler AS, Hachem CY, Hussan H

**Non-alcoholic steatohepatitis in liver transplant recipients diagnosed by serum cytokeratin 18 and transient elastography: A prospective study**


**CASE REPORT**

**Rare primary mature teratoma of the liver: A case report**

Kovalenko YA, Zharikov YO, Kiseleva YV, Goncharov AB, Shevchenko TV, Gurmikov BN, Kalinin DV, Zhao AV
ABOUT COVER
Editorial Board Member of World Journal of Hepatology, Manuel Luis Rodríguez-Perálvarez, MD, PhD, Consultant Hepatologist and Associate Professor of Medicine, Department of Hepatology and Liver Transplantation, Reina Sofia University Hospital, Córdoba 14014, Spain. ropeml@hotmail.com

AIMS AND SCOPE
The primary aim of World Journal of Hepatology (WJH, World J Hepatol) is to provide scholars and readers from various fields of hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

WJH mainly publishes articles reporting research results and findings obtained in the field of hepatology and covering a wide range of topics including chronic cholestatic liver diseases, cirrhosis and its complications, clinical alcoholic liver disease, drug induced liver disease autoimmune, fatty liver disease, genetic and pediatric liver diseases, hepatocellular carcinoma, hepatic stellate cells and fibrosis, liver immunology, liver regeneration, hepatic surgery, liver transplantation, biliary tract pathophysiology, non-invasive markers of liver fibrosis, viral hepatitis.

INDEXING/ABSTRACTING
The WJH is now abstracted and indexed in PubMed, PubMed Central, Emerging Sources Citation Index (Web of Science), Scopus, China National Knowledge Infrastructure (CNKI), China Science and Technology Journal Database (CSTJ), and Superstar Journals Database. The 2021 edition of Journal Citation Reports® cites the 2020 Journal Citation Indicator (JCI) for WJH as 0.61. The WJH’s CiteScore for 2020 is 5.6 and Scopus CiteScore rank 2020: Hepatology is 24/62.

RESPONSIBLE EDITORS FOR THIS ISSUE
Production Editor: Xu Guo; Production Department Director: Xiang Li; Editorial Office Director: Xiang Li.
Deep learning in hepatocellular carcinoma: Current status and future perspectives

Joseph C Ahn, Touseef Ahmad Qureshi, Amit G Singal, Debiao Li, Ju-Dong Yang

ORCID number: Joseph C Ahn 0000-0001-6994-2870; Touseef Ahmad Qureshi 0000-0002-6683-4556; Amit G Singal 0000-0002-1172-3971; Debiao Li 0000-0001-8560-8231; Ju-Dong Yang 0000-0001-7834-9825.

Author contributions: Yang JD devised the project and the main conceptual ideas for the review; Ahn JC conducted the literature search and identified relevant studies to be included in the review; Qureshi TA and Li D provided the technical expertise on artificial intelligence and deep learning; Ahn JC drafted the manuscript; Yang JD and Singal AG revised the manuscript critically for important intellectual content; and all authors approved the final version to be published.

Conflict-of-interest statement: Dr. Yang provides a consulting service for Exact Sciences and Gilead; Dr. Singal has been on advisory boards and served as a consultant for Genentech, Bayer, Eisai, IMS, Exelixis, AstraZeneca, and TARGET RWE. No other potential conflicts of interest relevant to this article exist.

Country/Territory of origin: United States

Specialty type: Gastroenterology and hepatology

Abstract

Hepatocellular carcinoma (HCC) is among the leading causes of cancer incidence and death. Despite decades of research and development of new treatment options, the overall outcomes of patients with HCC continue to remain poor. There are areas of unmet need in risk prediction, early diagnosis, accurate prognostication, and individualized treatments for patients with HCC. Recent years have seen an explosive growth in the application of artificial intelligence (AI) technology in medical research, with the field of HCC being no exception. Among the various AI-based machine learning algorithms, deep learning algorithms are considered state-of-the-art techniques for handling and processing complex multimodal data ranging from routine clinical variables to high-resolution medical images. This article will provide a comprehensive review of the recently published studies that have applied deep learning for risk prediction, diagnosis, prognostication, and treatment planning for patients with HCC.

Key Words: Hepatocellular carcinoma; Artificial intelligence; Deep learning

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: There are emerging roles for deep learning technology in the field of hepatocellular carcinoma.
Introduction

Hepatocellular carcinoma (HCC) is an aggressive primary liver cancer that develops in the setting of chronic parenchymal liver diseases, and is among the top causes of cancer incidence and mortality worldwide[1,2]. While the burden of HCC has been declining with effective antiviral therapy against hepatitis B virus (HBV) and hepatitis C virus (HCV), HCC incidence related to metabolic syndrome will likely continue to rise due to the dramatic increase in the prevalence of non-alcoholic fatty liver disease (NAFLD) in the general population[3]. Decades of HCC research led to the development of a screening protocol, non-invasive diagnostic modalities based on imaging, and various treatment modalities including surgical, locoregional and systemic therapies[4,5]. However, the overall outcomes of patients with HCC continue to remain poor and there are areas of significant unmet need in risk prediction, early detection, accurate prognostication, and individualized treatments for patients with HCC.

Patients with HCC generate enormous amounts of health data. While promising for researchers, ensuring that such high volumes of data are turned into actionable knowledge can be a significant challenge. Artificial intelligence (AI) is thought to be capable of synthesizing and analyzing multimodal data with superhuman degrees of accuracy or reliability, and recent years have seen a rapid growth in the application of AI to many fields of medicine including hepatology[6]. This “AI revolution” over the past decade has been possible due to the advent of deep learning technology. Deep learning algorithms can process a broad spectrum of medical data from structured numeric data such as vital signs and laboratory values, high dimensional data from multi-omics studies, as well as digitized high-resolution images from various radiologic and histopathologic studies. This review aims to provide an overview as well as highlight examples of the many potential applications of deep learning to improve the care of patients with HCC.

AI, Machine Learning, and Deep Learning

AI-based approaches provide a variety of methods for a range of tasks and clinical application including image classification, organ and lesion segmentation, accurate extraction of key imaging features and measurements, tumor detection, stratification of high-risk subjects, prediction of disease and treatment outcome (Figure 1). Advancements in AI in recent years, particularly in the realm of medical image processing and analysis, offer an enormous range of automated tools for extracting precise measurements of biomarkers, revealing complex features, quantifying tissue characteristics and performing radiomics for deep analysis of raw imaging data.

The term “artificial intelligence” encompasses a broad range of technology that enables machines to perform tasks typically thought to require human reasoning and problem-solving skills[7]. “Machine learning” is a branch of AI in which computer algorithms train on sample data to build a mathematical model that makes predictions or decisions without being explicitly programmed to do so[8]. Machine learning algorithms can be broadly divided into supervised and unsupervised learning. Supervised learning algorithms train on sample data with labeled outcome data, and their goal is to learn the relationship between the input data and the outcomes to make accurate predictions.
Figure 1 Schematic representation of the relationships between the terms artificial intelligence, machine learning, and deep learning, and how deep learning can utilize multimodal data to improve care for patients with hepatocellular carcinoma.

According to Ahn JC et al. [9], deep learning (DL) refers to highly complex AI models utilizing multiple layers of artificial neural networks (ANNs) and has recently emerged as a state-of-the-art AI technique for analyzing complex, high-dimensional healthcare data. Some of the commonly used DL techniques include convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [12]. CNNs have connective patterns resembling those of an animal visual cortex and can detect inherent spatial features of high-dimensional images. RNNs have connections forming a directed graph along a temporal sequence, and therefore can be highly useful in time series prediction.

It is crucial to recognize that any AI-based machine learning algorithms require external validation in an independent dataset as models could be overfitted and end up overestimating the performance. In this review article, the performance characteristics of the various DL models are from the validation cohorts, and not the original derivation cohorts used to train the algorithms.

HCC CLINICAL DATA

Despite multiple available risk prediction tools for HCC, none have been rigorously validated or endorsed by major liver societies. Currently, HCC surveillance is recommended for patients with cirrhosis and high risk patients with chronic HBV infection [13]. Accurate prediction models utilizing more specific risk factors for HCC development at individual levels would allow health systems to implement targeted screening strategies. Ioannou et al [14] trained a RNN to predict HCC development within 3 years using 4 baseline variables and 27 longitudinal variables from 48151 patients with HCV-related cirrhosis in the national Veterans Health Administration. The RNN model significantly outperformed logistic regression and exhibited an area under the curve (AUC) of 0.759 among all samples and an AUC of 0.806 among patients with sustained virologic response. Phan et al [15] surveyed 1 million random samples from Taiwan’s National Health Insurance Research Database between 2002 to 2010 to predict liver cancer among patients with viral hepatitis. The disease history of...
Ahn JC et al. Deep learning in HCC

When created a CNN which detects and characterizes benign and malignant focal liver cirrhosis, and HCC with a classification accuracy of 96.6%. Schmauch ultrasound images, which differentiated between normal liver, chronic liver disease, built a CNN using echotexture and roughness of liver surface on 754 segmented ultrasound images with a classification accuracy of 97.2%. Additionally, Bharti auto-encoder, an unsupervised DL technique, to segment and classify liver lesions on 87.1% in the testing set. In 2017, Hassan ANN showed promising performances with accuracies of 94.5% in the training set and 55.0% in the testing set. In 2012, Streba et al. constructed a deep neural network to predict 3-year and 5-year incidence of HCC in 424 patients with HBV-related cirrhosis on entecavir therapy. When applied to an external validation cohort of 316 patients, the DL model achieved a Harrell’s C-index of 0.782 and significantly outperformed 6 previously reported models based on traditional modeling. The same group also developed another DL model called the AI-based Model of Recurrence after Liver Transplantation (MoRAL-AI) to predict HCC recurrence after liver transplantation using variables such as tumor diameter, age, alpha-fetoprotein (AFP), and prothrombin time. The MoRAL-AI showed significantly better predictive performance compared to conventional models such as the Milan, UCSF, up-to-seven, and Kyoto criteria (C-index = 0.75 vs 0.64, 0.62, 0.50, 0.50, respectively; P < 0.001).

HCC MULTI-OMICS

Serum AFP has been widely used as a predictive and prognostic biomarker for HCC, but AFP has limited sensitivity for detecting early-stage HCC and its levels do not reliably correlate with disease progression. Recent advances in multi-omics related to HCC are expected to address this unmet need for novel biomarkers. Multi-omics refers to an approach to biological analysis which utilizes data sets from multiple ‘omics’, such as the genome, epigenome, transcriptome, proteome, metabolome and microbiome. Multi-omics experiments generate an enormous amount of information, and various machine learning techniques including DL that can help with the computational challenges of processing and analyzing such high dimensional data. Xie et al. used gene expression profiling of peripheral blood to build an ANN model that classifies HCC patients from a control group. Using a nine-gene expression system, the ANN was able to distinguish HCC patients from controls with an AUC of 0.943, 98% sensitivity, and 85% specificity, although it should be noted that the control group was healthy individuals rather than patients with cirrhosis, which could have overestimated the performance of the model. Choi et al. proposed a novel network-based DL method to identify prognostic gene signatures via G2Vec, a modified Word2Vec model originally used for natural language processing (NLP). When applied to gene expression data for HCC from the Cancer Genome Atlas (TCGA), G2Vec showed superior prediction accuracy for patient outcomes compared to existing gene selection methods and was able to identify two distinct gene modules significantly associated with HCC prognosis. Chaudhary et al. used RNA sequencing, miRNA, and methylation data of 360 HCC patients from TCGA to build an autoencoder, which is an unsupervised feed-forward neural network. Using this DL model, they were able to distinguish patients with survival differences and identify specific mutations and pathways as predictors of aggressive tumor behavior.

RADIOLOGY

HCC diagnosis and segmentation

In recent years, there have been remarkable advances in the application of AI for the interpretation of medical imaging, primarily due to the use of DL algorithms using CNNs. CNN algorithms trained on ultrasound, computed tomography (CT), or magnetic resonance imaging (MRI) images have shown excellent performances in detection of lesions, classification of lesions, segmentation of organs or anatomic structures, and imaging reconstruction.

In 2012, Streba et al. prospectively studied contrast-enhanced ultrasound images of 112 patients to train an ANN that classified five different types of liver tumors. The ANN showed promising performances with accuracies of 94.5% in the training set and 87.1% in the testing set. In 2017, Hassan et al. reported using the stacked sparse auto-encoder, an unsupervised DL technique, to segment and classify liver lesions on ultrasound images with a classification accuracy of 97.2%. Additionally, Bharti et al. built a CNN using echotexture and roughness of liver surface on 754 segmented ultrasound images, which differentiated between normal liver, chronic liver disease, cirrhosis, and HCC with a classification accuracy of 96.6%. Schmauch et al. also created a CNN which detects and characterizes benign and malignant focal liver lesions on 2-D ultrasound images from 367 patients from various institutions. When
applied to a new dataset of 177 patients, the model achieved a weighted mean AUC of 0.891. Recently, Brehar et al.[29] conducted a study comparing CNN’s performance for HCC detection on ultrasound images against conventional machine learning algorithms including multi-layer perceptron, support vector machines, random forest and AdaBoost. The CNN achieved an AUC of 0.95% with 91.0% accuracy, 94.4% sensitivity, and 88.4% specificity and significantly outperformed the conventional machine learning algorithms. Beyond detecting the actual presence of HCC on ultrasound images, studies have also attempted to predict the risk of future HCC development based on analyzing the ultrasound images of liver parenchyma in patients without HCC. For example, Jin et al.[30] performed a DL radiomics analysis on 2-D shear wave elastography and corresponding B-mode ultrasound images of 434 chronic HBV patients, which predicted 5-year HCC development with AUC of 0.900 in the test cohort.

In addition to ultrasound images, cross-sectional imaging from CT or MRI studies serve as an extremely abundant and promising source of data for DL. In 2018, Yasaka et al.[31] used CT image sets of liver masses from 460 patients to train a CNN that can classify liver lesions into five categories of: (1) HCC; (2) Other malignant tumors; (3) Indeterminate masses; (4) Hemangiomas; and (5) Cysts with a median AUC of 0.92. Shi et al.[32] showed that incorporation of a CNN enabled identification of HCC using a three-phase CT imaging protocol with a diagnostic accuracy similar to that of a four-phase protocol, which would allow patients to receive lower doses of radiation. Segmentation of HCC, liver parenchyma, and other organs on CT scan is very important for determination of tumor extent and treatment planning, but manual contouring of the images is highly time-consuming and subject to inter-observer variability. The 2017 International Conference On Medical Image Computing Computer Assisted Intervention called for a Liver Tumor Segmentation Benchmark (LITS) challenge, encouraging researchers to develop automatic segmentation algorithms to segment liver lesions using 200 CT scans (training: 130; testing: 70) provided by clinical sites around the world. Several teams participating in the challenge have developed DL algorithms with promising performances for HCC segmentation using CT images.[33-37]. Beyond the LITS challenge, there are ongoing research efforts to improve segmentation using different architectures of DL networks.[38-42]

Hamm et al.[43] used MRI images from 494 patients to train a CNN which can classify hepatic lesions into six different categories. When applied to random cases in the test set, the CNN outperformed expert radiologists (90% sensitivity and 98% specificity vs 82.5% sensitivity and 96.5% specificity) and especially for HCC detection (90% sensitivity vs 60%-70% sensitivity). The same group conducted additional studies to make their CNN interpretable by generating highlighted feature maps corresponding to liver lesions.[44]. Wu et al.[45] built a CNN using multiphase MRI images and achieved an AUC of 0.95 for distinguishing Liver Imaging Reporting and Data System (LI-RADS) grade 3 from LI-RADS 4 and 5 lesions for HCC diagnosis. Zhen et al. [46] also trained a CNN model combining unenhanced MRI images and clinical variables from 1210 patients with liver tumors, which demonstrated diagnostic performances on par with three experienced radiologists using enhanced MRI images.

**HCC prognostication, treatment planning, and response to treatment**

In addition to serving as accurate and efficient tools for diagnosis of HCC, DL models utilizing radiology data can also be used for prognostication, treatment planning, and assessing tumor response to therapy. Vascular invasion is a key prognostic element in patients with HCC. Recent studies developed CNN models with promising ability to detect microvascular invasion on MRI images of HCC patients undergoing surgical resection.[47-49]. An et al.[50] used an unsupervised CNN-based deformable image registration technique to assess the relationship between ablative margins and local tumor progression in 141 patients with single HCC who underwent microwave ablation, and demonstrated that patients with ablative margins < 5 mm were at significantly higher risk of local tumor progression. Liu et al.[51] developed a DL radiomics model to predict responses to trans-arterial chemoembolization (TACE) using ultrasound images of 130 HCC patients, which accurately predicted TACE response with an AUC of 0.93. The same group also assessed their ultrasound-based DL radiomics model to predict 2-year progression-free survival among 419 HCC patients and facilitate optimized treatment selection. Peng et al.[52] trained a residual CNN model to predict response to TACE using CT images from 562 patients with intermediate-stage HCC undergoing TACE, which showed accuracies of 85.1% and 82.8% in two external validation cohorts. Another study developed a DL score for disease-specific survival by using CT images in a cohort of 243 patients with HCC.
Ahn JC et al. Deep learning in HCC

Large-scale, available, limiting external validation. Independent validation of the proposed model using their own patients’ data may over-represent certain groups of patients and not accurately reflect the real-world population of patients seen at local community hospitals. Finally, AI models, like other prediction models, are often not publicly released, and their performance can be overestimated in external validation. For example, Wang et al. [60] found that the performance of an AI model for predicting HCC recurrence was significantly overestimated in an independent validation set.

HCC PATHOLOGY

Automated interpretation of histopathologic images from liver biopsy is another major area of medical imaging in patients with HCC where DL can be utilized. In addition to effectively replicating the human pathologists’ jobs of diagnosing and grading HCC, DL models can help identify and analyze additional complex imaging features and patterns which are related to specific mutations and disease prognosis. Lin et al. [55] used images from multiphoton microscopy of 113 HCC patients to train a CNN with over 90% accuracy for determining HCC differentiation. Kiani et al. [56] developed a CNN-based “Liver Cancer Assistant” which accurately differentiated hematoxylin and eosin (H&E) images of HCC and cholangiocarcinoma and helped improve the diagnostic performance of nine pathologists. Liao et al. [57] used TCGA dataset for training a CNN that distinguished HCC from adjacent normal tissues with perfect performance (AUC: 1.00) and predicted the presence of specific somatic mutations with AUCs over 0.70. Wang et al. [58] trained a CNN for automated segmentation and classification of individual nuclei at single-cell levels on H&E-stained tissue sections of HCC tumors from TCGA, and performed feature extraction to identify 246 quantitative image features. Then, a clustering analysis by an unsupervised learning approach identified three distinct histologic subtypes which were independent of previously established genomic clusters and had different prognosis. Chen et al. [59] trained a CNN for automatic grading of HCC tumors on histopathological H&E images, which showed 96% accuracy for benign and malignant classification and 89.6% accuracy for the degree of tumor differentiation, and predicted the presence of specific genetic mutations.

Lu et al. [60] applied three pre-trained CNN models to extract imaging features from HCC histopathology and performed Cox proportional hazards analysis to predict overall survival and disease-free survival, and observed significant correlations between the imaging features and established biological pathways. Saillard et al. [61] used two DL algorithms based on whole-slide digitized histological slides from 194 patients with HCC to predict the survival of patients treated by surgical resection. When tested on an independent validation set from TCGA, both DL models had a higher discriminatory power than a score combining all baseline variables associated with survival. Shi et al. [62] built an interpretable DL framework using pathologic images from 1445 patients with HCC and developed a “tumor risk score” which showed prognostic performances independent of and superior to clinical staging systems and stratified patients into five groups of different prognosis. A recent study by Yamashita et al. [63] developed a histopathology-based DL based system which stratified patients with risk scores for postsurgical recurrence of HCC.

FUTURE DIRECTION

There are several key issues to address before DL-based AI models can be universally implemented in real world clinical practice settings. Due to their complexity, DL models are traditionally considered to be “black-box” models, meaning humans cannot understand how the DL models make their predictions. Interpretability of the DL models are crucial for physicians to accept and trust them in everyday clinical practice, and for troubleshooting and improving the models for rare cases. This is being addressed by recent developments in various “explainable AI” techniques but currently there is no clear consensus on the best methodology. Another potential limitation is the generalizability of the individual DL algorithms. Concerns have been raised that AI algorithms developed at highly specialized academic medical centers using their own patients’ data may over-represent certain groups of patients and not accurately reflect the real-world population of patients seen at local community hospitals. Finally, AI models, like other prediction models, are often not publicly available, limiting external validation. Independent validation of the proposed model and comparison to old models are as important as deriving new models. Large-scale,
<table>
<thead>
<tr>
<th>Study</th>
<th>Cohort</th>
<th>Data source</th>
<th>Deep learning</th>
<th>Input</th>
<th>Output</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ioannou et al[14] 2020</td>
<td>48151 HCV cirrhosis (T: 90%, V: 10%)</td>
<td>VHA database</td>
<td>RNN</td>
<td>Clinical variables</td>
<td>Risk of HCC development</td>
<td>RNN predicted HCC development with AUC of 0.759, and AUC of 0.806 among those who achieved SVR</td>
</tr>
<tr>
<td>Phan et al[15] 2020</td>
<td>6052 HBV and HCV (T: 70%, V: 30%)</td>
<td>Taiwanese NHIRD</td>
<td>CNN</td>
<td>Disease history data</td>
<td>Risk of HCC development</td>
<td>CNN achieved an accuracy of 0.980 and AUC of 0.886 for predicting HCC development among viral hepatitis patients</td>
</tr>
<tr>
<td>Nam et al[16] 2020</td>
<td>T: 424 HBV cirrhosis; V: 316 HBV cirrhosis</td>
<td>2 Korean centers</td>
<td>ResNet</td>
<td>Clinical variables</td>
<td>Risk of HCC development</td>
<td>DL model achieved an accuracy of 0.763 and AUC of 0.782 in the validation cohort and outperformed previous models</td>
</tr>
<tr>
<td>Nam et al[17] 2020</td>
<td>T: 349 LT recipients; V: 214 LT recipients</td>
<td>3 Korean LT centers</td>
<td>ResNet</td>
<td>Clinical variables</td>
<td>Recurrent HCC after LT</td>
<td>DL model significantly outperformed conventional models in prediction of post-T HCC recurrence with AUC of 0.75</td>
</tr>
<tr>
<td>Xie et al[20] 2018</td>
<td>T: 133 HCC/54 HV; V: 52 HCC/34 HV</td>
<td>1 center in China</td>
<td>ANN</td>
<td>Gene expression</td>
<td>HCC detection</td>
<td>ANN using nine genes had an AUC of 0.943, 98% sensitivity, and 85% specificity for classifying HCC</td>
</tr>
<tr>
<td>Choi et al[21] 2018</td>
<td>135 HCC (10-fold CV)</td>
<td>TCGA</td>
<td>G2Vec</td>
<td>Gene expression</td>
<td>HCC prognosis</td>
<td>G2Vec showed significantly higher prediction accuracy for patient outcomes compared to existing gene selection tools</td>
</tr>
<tr>
<td>Chaudhary et al[22] 2018</td>
<td>T: 360 HCC; V: 220, 221, 166, 40, 27 HCC</td>
<td>TCGA; 5 external datasets</td>
<td>Auto-encoder</td>
<td>RNA-seq, miRNA-seq, methylation</td>
<td>HCC prognosis</td>
<td>DL model distinguished groups with survival differences and identified mutations and pathways predicting aggressive tumor behavior</td>
</tr>
<tr>
<td>Streba et al[25] 2012</td>
<td>112 FLL (10-fold CV)</td>
<td>1 center in Romania</td>
<td>ANN</td>
<td>US images</td>
<td>FLL type</td>
<td>ANN had 87.12% testing accuracy, 93.2% sensitivity, and 89.7% specificity for classifying 5 classes of liver lesions</td>
</tr>
<tr>
<td>Hassan et al[26] 2017</td>
<td>110 FLL (10-fold CV)</td>
<td>1 center in Egypt</td>
<td>Auto-encoder</td>
<td>US images</td>
<td>FLL type</td>
<td>The proposed system had 97.2% accuracy, 98% sensitivity, and 95.70% specificity for classifying liver lesions</td>
</tr>
<tr>
<td>Bharti et al[27] 2018</td>
<td>24 normal, 25 CLD, 25 cirrhosis, 20 HCC</td>
<td>1 center in India</td>
<td>CNN</td>
<td>US images</td>
<td>Liver stages</td>
<td>CNN achieved 96.6% classification accuracy for differentiating normal liver, CLD, cirrhosis, and HCC</td>
</tr>
<tr>
<td>Schmauch et al[28] 2019</td>
<td>T: 367 FLL; V: 177 FLL</td>
<td>Centers in France</td>
<td>ResNet</td>
<td>US images</td>
<td>FLL type</td>
<td>DL model reached mean AUC of 0.955 for focal liver lesion detection and 0.916 for focal liver lesion characterization</td>
</tr>
<tr>
<td>Beerar et al[29] 2020</td>
<td>T: 200 HCC; V: 68 HCC</td>
<td>1 center in Romania</td>
<td>CNN</td>
<td>US images</td>
<td>HCC detection</td>
<td>CNN achieved AUC of 0.95, accuracy of 0.91, 94.4% sensitivity and 88.4% specificity for HCC detection</td>
</tr>
<tr>
<td>Jin et al[30] 2021</td>
<td>454 HBV (3:1 split)</td>
<td>1 center in China</td>
<td>DL radiomics</td>
<td>US images</td>
<td>Risk of HCC development</td>
<td>DL radiomics model predicted 5-yr HCC development risk with AUC of 0.900 in the test set</td>
</tr>
<tr>
<td>Yasaka et al[31] 2018</td>
<td>T: 460 liver masses; V: 100 liver masses</td>
<td>1 center in Japan</td>
<td>CNN</td>
<td>CT images</td>
<td>Liver mass type</td>
<td>CNN classified liver lesions into five categories with a median AUC of 0.92</td>
</tr>
<tr>
<td>Shi et al[32] 2020</td>
<td>449 FLL; (T: 80%, V: 20%)</td>
<td>1 center in China</td>
<td>CNN</td>
<td>CT images</td>
<td>FLL type</td>
<td>CNN applied to three-phase CT protocol images achieved AUC of 0.925 for differentiating HCC</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>T:</td>
<td>V:</td>
<td>Center</td>
<td>CNN</td>
<td>MRI Images</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>----</td>
<td>---</td>
<td>-------</td>
<td>------</td>
<td>------------</td>
</tr>
<tr>
<td>Hamm et al.</td>
<td>2019</td>
<td>434 FLL;</td>
<td>60 FLL</td>
<td>1 center in United States</td>
<td>CNN</td>
<td>MRI images</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2019</td>
<td>434 FLL;</td>
<td>60 FLL</td>
<td>1 center in United States</td>
<td>CNN</td>
<td>MRI images</td>
</tr>
<tr>
<td>Wu et al.</td>
<td>2020</td>
<td>89 liver tumors;</td>
<td>60: 20: 20</td>
<td>1 center in United States</td>
<td>CNN</td>
<td>MRI images</td>
</tr>
<tr>
<td>Zhen et al.</td>
<td>2020</td>
<td>1210 liver tumors;</td>
<td>201 liver tumors</td>
<td>1 center in China</td>
<td>CNN</td>
<td>MRI images</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2019</td>
<td>434 FLL;</td>
<td>60 FLL</td>
<td>1 center in United States</td>
<td>CNN</td>
<td>MRI images</td>
</tr>
<tr>
<td>Hamm</td>
<td>2019</td>
<td>434 FLL;</td>
<td>60 FLL</td>
<td>1 center in United States</td>
<td>CNN</td>
<td>MRI images</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2020</td>
<td>60 HCC;</td>
<td>40 HCC</td>
<td>1 center in China</td>
<td>CNN</td>
<td>MRI images</td>
</tr>
<tr>
<td>Jiang et al.</td>
<td>2021</td>
<td>405 HCC; (T: 80%, V: 20%)</td>
<td>1 center in China</td>
<td>CNN</td>
<td>CT images</td>
<td>MVI in HCC</td>
</tr>
<tr>
<td>An et al.</td>
<td>2020</td>
<td>141 single HCC resect MWA</td>
<td>1 center in China</td>
<td>CNN</td>
<td>MRI images</td>
<td>Ablative margin</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2020</td>
<td>89 HCC resect TACE; V: 41 HCC rec. TACE</td>
<td>1 center in China</td>
<td>CNN</td>
<td>Ultrasound images</td>
<td>Response to TACE</td>
</tr>
<tr>
<td>Peng et al.</td>
<td>2020</td>
<td>562 HCC resect TACE; V:227 HCC rec. TACE</td>
<td>3 centers in China</td>
<td>CNN</td>
<td>CT images</td>
<td>Response to TACE</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2020</td>
<td>243 HCC resect TACE (6:1:3 split)</td>
<td>1 center in China</td>
<td>CNN</td>
<td>CT images</td>
<td>Post-TACE survival</td>
</tr>
<tr>
<td>Zhang et al.</td>
<td>2020</td>
<td>201 HCC resect TACE + sorafenib (T: 120, V: 81)</td>
<td>3 centers in China</td>
<td>CNN</td>
<td>CT images</td>
<td>OS on TACE + sorafenib</td>
</tr>
<tr>
<td>Lin et al.</td>
<td>2019</td>
<td>113 HCC</td>
<td>1 center in China</td>
<td>CNN</td>
<td>Histopath images</td>
<td>HCC differentiation</td>
</tr>
<tr>
<td>Kiani et al.</td>
<td>2020</td>
<td>70 WSI (35 HCC, 35 CC)</td>
<td>TCGA</td>
<td>CNN</td>
<td>Histopath images</td>
<td>HCC vs CC</td>
</tr>
<tr>
<td>Liao et al.</td>
<td>2020</td>
<td>491 HCC; V: 455 HCC</td>
<td>TCGA; 1 center in China</td>
<td>CNN</td>
<td>Histopath images</td>
<td>HCC detection, mutations</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2020</td>
<td>99 HCC; V: 205 HCC</td>
<td>TCGA</td>
<td>CNN</td>
<td>Histopath images</td>
<td>Histological HCC subtype</td>
</tr>
<tr>
<td>Study</td>
<td>T: HCC/Normal</td>
<td>V: Normal</td>
<td>Method</td>
<td>Data Source</td>
<td>Images</td>
<td>Outcome</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>-----------</td>
<td>--------</td>
<td>-------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Shen et al. [59] 2020</td>
<td>T: 402 HCC/89 normal; V: 67 HCC/34 normal</td>
<td>GDC portal; 1 center in China</td>
<td>CNN</td>
<td>Histopath images</td>
<td>HCC grade mutations</td>
<td>CNN achieved 89.4% accuracy for tumor differentiation stage and predicted presence of specific gene mutations</td>
</tr>
<tr>
<td>Lu et al. [60] 2020</td>
<td>T: 421 HCC/105 normal (6-fold CV)</td>
<td>GDC portal</td>
<td>CNN</td>
<td>Histopath images</td>
<td>HCC prognosis</td>
<td>Pre-trained CNN predicted OS using pathology images and identified HCC subgroups with different prognosis</td>
</tr>
<tr>
<td>Saillard et al. [61] 2020</td>
<td>T: 194 HCC; V: 328 HCC</td>
<td>1 French center TCGA</td>
<td>CNN</td>
<td>Histopath images</td>
<td>Survival after HCC resection</td>
<td>CNN models using pathology images predicted survival with C-index 0.75-0.78 and outperformed conventional models</td>
</tr>
<tr>
<td>Shi et al. [62] 2021</td>
<td>T: 1125 HCC; V: 3201HCC</td>
<td>1 center in China; TCGA</td>
<td>CNN</td>
<td>Histopath images</td>
<td>HCC outcomes</td>
<td>Deep learning-based “tumor risk score” was superior to clinical staging and stratified 5 groups of different prognosis</td>
</tr>
<tr>
<td>Yamashita et al. [63] 2021</td>
<td>T: 36 WSI; V: 30 WSI</td>
<td>1 center in United States; TCGA</td>
<td>CNN</td>
<td>Histopath images</td>
<td>Post-surgical recurrence</td>
<td>CNN risk scores outperformed TNM system for predicting recurrence and identified high- and low-risk subgroups</td>
</tr>
</tbody>
</table>

ANN: Artificial neural network; AUC: Area under the curve; CC: Cholangiocarcinoma; CNN: Convolutional neural network; CV: Cross-validation; FLL: Focal liver lesion; GDC: Genomic Data Commons; HBV: Hepatitis B virus; HCC: Hepatocellular carcinoma; HCV: Hepatitis C virus; HV: Healthy volunteers; LT: Liver transplant; MVI: Microvascular invasion; MWA: Microwave ablation; OS: Overall survival; PFS: Progression-free survival; RFA: Radio-frequency ablation; RNN: Recurrent neural network; SR: Surgical resection; STS-net: Spatial transformed similarity network; SVR: Sustained virologic response; T: Training; TCGA: The Cancer Genome Atlas; V: Validation; VHA: Veterans Health Administration; WSI: Whole slide image; CT: Computed tomography; MRI: Magnetic resonance imaging; NHIRD: National Health Insurance Research Database; TNM: Tumor, Nodes, Metastasis; TACE: Trans-arterial chemoembolization; LI-RADS: Liver Imaging Reporting and Data System.

prospective, multi-centered studies involving diverse populations with external validation will be necessary before DL algorithms can be widely accepted.

A currently under-explored, but highly promising and exciting area for the application of DL is the field of autonomous robotics. In a recent editorial, Gumbs et al. [64] state that while the current form of robotic surgery seems like a form of minimally invasive surgery, the true power of robotic surgery exists in its potential to create autonomous actions. Recently, a DL-based surgical instrument tracking algorithm was able to closely track the instruments during robotic surgery and evaluate the surgeons’ performance, demonstrating that DL algorithms can learn the correct steps of robotic surgery [65]. With the help of DL and other AI technologies, it may be possible to imagine a future where fully autonomous robots perform resection of large, complex HCC in ways that no human surgeons can mimic. However, there are significant barriers before the idea of fully autonomous robotic surgery can become a reality, including the current technical limitations of autonomous surgical robotics, as well as the hesitation of patients and providers to fully trust autonomous robots to perform invasive operations. “Explainability” of the DL algorithms will be critical here, as humans would need to be able to understand and correct every single mistake that an autonomous robot makes during surgery. Therefore, for the foreseeable future, DL will most likely remain as a helpful, adjunctive tool to assist human surgeons.

**CONCLUSION**

This review has provided a comprehensive overview of various ways in which DL algorithms can be employed to assist medical providers and enhance the care of patients with HCC (Table 1). DL algorithms not only can efficiently and accurately replicate the same jobs performed by human physicians, but more importantly can help discover novel biologic pathways and disease subgroups with clinical significance by processing and analyzing complex high-dimensional data in ways impossible for the human brain.

Despite some important limitations to overcome, application of state-of-the-art AI technologies such as DL for the care of patients with HCC is no longer a futuristic idea but is rapidly becoming a reality. Most of the studies covered in this review were published within the past two years, and the number of studies utilizing DL continues...
to increase exponentially. We anticipate that DL algorithms will soon take a major role in the diagnosis, prognostication, and treatment of patients with HCC.

REFERENCES


prognostic genes. Sci Rep 2018; 8: 13729 [PMID: 30213980 DOI: 10.1038/s41598-018-32180-0]


37 Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. IEEE Trans Med Imaging 2018; 37: 2663-2674 [PMID: 29994201 DOI: 10.1109/TMI.2018.2845918]


