Artificial Intelligence Applications in Hepatology

Jörn M. Schattenberg,[1] Naga Chalasani,[2] and Naim Alkhouri[3]

[1]Metabolic Liver Research Program, I. Department of Medicine, University Medical Center Mainz, Mainz, Germany; [2]Indiana University School of Medicine and Indiana University Health, Indianapolis, Indiana; and [3]Arizona Liver Health and University of Arizona, Tucson, Arizona

Over the past 2 decades, the field of hepatology has witnessed major developments in diagnostic tools, prognostic models, and treatment options making it one of the most complex medical subspecialties. Through artificial intelligence (AI) and machine learning, computers are now able to learn from complex and diverse clinical datasets to solve real-world medical problems with performance that surpasses that of physicians in certain areas. AI algorithms are currently being implemented in liver imaging, interpretation of liver histopathology, noninvasive tests, prediction models, and more. In this review, we provide a summary of the state of AI in hepatology and discuss current challenges for large-scale implementation including some ethical aspects. We emphasize to the readers that most AI-based algorithms that are discussed in this review are still considered in early development and their utility and impact on patient outcomes still need to be assessed in future large-scale and inclusive studies. Our vision is that the use of AI in hepatology will enhance physician performance, decrease the burden and time spent on documentation, and reestablish the personalized patient-physician relationship that is of utmost importance for obtaining good outcomes.

Keywords: Computer-Based Learning; Ethics; Machine Learning; Deep Learning.

Technological advancements have created the unique opportunity to use artificial intelligence (AI) and more specifically machine learning (ML) in clinical medicine. With available computational power, AI has the potential to transform patient care without losing the patient-centric, physician-guided approach of traditional clinical medicine. This has become even more evident during the COVID-19 pandemic, which provided unprecedented advancements in technology acceptance and availability in all areas of society and in particular in the health care system. Although AI is considered the overarching term that details the rational exploitation of data by a machine, ML more specifically describes the building of models that learn from available data to improve the prediction or performance related to a specific task without actually programming.[1] The ability of ML algorithms to predict outcomes can be exploited based on labeled (supervised) or unlabeled (unsupervised) data. By using the ML algorithm over time and providing more training, the desired output becomes progressively more accurate. Deep learning (DL) refines and narrows ML by using multiple neuronal networks that mimic the human neurologic system to analyze, identify, and learn from complex datasets.[2] The neural networks are organized in multiple layers where the signal travels from the first layer (input) to the last layer (output) after going through multiple intervening layers.

A few important issues have to be considered when aiming to implement AI in the clinical environment today (Figure 1). Beyond investments in technology in the health care sector, the quality of the data that are used to develop algorithms and predict outcome is most critical. In the field of hepatology research, several large prospective studies that are aimed to explore outcome are actively recruiting and will provide the quality and robustness of the data that are required.[3,4] The enormous potential to account for a large number of variables in complex databases and determine the likelihood of specific outcomes in a very short time, will markedly outperform a single physician's capability that operates at the level of personal experience and medical education. Despite high expectations by many stakeholders and receptivity toward AI in the general society and among medical professionals, the level of implementation in clinical practice today is relatively low.[5]

There are several limitations that must be taken into account to allow for a safe application of AI and a higher penetration into clinical care. The assembly of high-quality representative data sets that eliminate unwanted and unconscious biases is a prerequisite for building ML models that do not perpetuate health care disparities. The inability of AI algorithms to account for

Abbreviations used in this paper: **AI, artificial intelligence; ALD, alcohol-related liver disease; ALT, alanine aminotransferase; ANN, artificial neural network; AST, aspartate aminotransferase; AUROC, area under the receiver operating curve; CNN, convolutional neural network; CT, computer tomography; DL, deep learning; ECG, electrocardiogram; EHR, electronic health record; HCC, hepatocellular carcinoma; HVPG, hepatic vein pressure gradient; MELD, Model for End-Stage Liver Disease; ML, machine learning; MRI, magnetic resonance imaging; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; NIT, noninvasive test; PDFF, proton density fat fraction; PSC, primary sclerosing cholangitis; RF, random forest.**

Most current article

1542-3565
https://doi.org/10.1016/j.cgh.2023.04.007

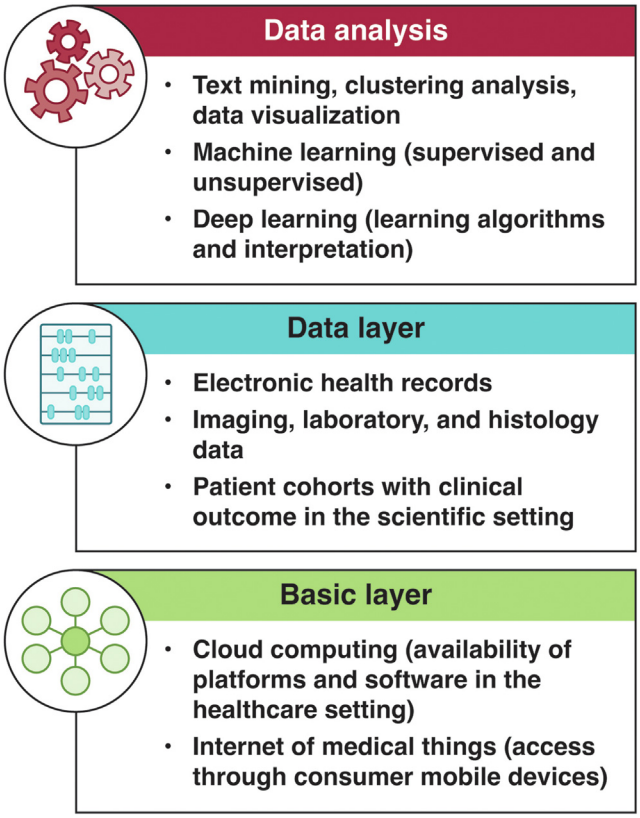

Figure 1. Framework of artificial intelligence in clinical medicine. Adapted from Lu Zx, Qian P, Bi D, et al. Application of AI and IoT in clinical medicine: summary and challenges. Curr Med Sci 2021;41:1134–1150.

information gained from a direct patient-physician interaction is an inherent limitation. The AI algorithms will never be able to substitute for a physician's direct interaction with their patients. We view AI as a complementary tool to significantly enhance patient-provider interaction and patient care. For AI's integration into hepatology clinical practice, multiple currently open questions need to be addressed including the quality of data synthesized, operational procedures, data and systems safety, and ethical challenges. Ethical challenges may arise from clinical decision making based on AI-generated diagnostic algorithms that are not readily recapitulated through medical reasoning. As seen with automated driving, a critical question arises around responsibility and liability in the context of a decision that is based on an AI algorithm. Similarly in medicine, the consequences of false-positive and false-negative results that are generated by AI are far reaching for patients and their providers. This is exacerbated by the fact that AI algorithms are comprised of complex interconnected structures with numerous parameters and a "black box" nature, offering little understanding of their inner working. Explainable AI is a set of processes that allows humans to comprehend the output created by ML algorithms, which help develop trust in the system and meet adherence to regulatory requirements.[6]

The challenges in hepatology are to allow for a safe and evidence-based implementation of AI to support clinical decision making. Several AI approaches have been used in hepatology with a focus on identification of cases (through imaging and noninvasive tests [NIT]), augmentation of histologic analysis, and prediction of outcome (Table 1). This review article focuses on current developments in AI/ML with potential applications in hepatology and defines areas of research that should be addressed in the future. To prepare for this manuscript, a literature search was conducted by the authors using the electronic PubMed databased and the following search terms: "artificial intelligence," "machine learning," and "liver disease." The authors selected the most relevant English language articles to provide an overview of where AI may impact the practice of clinical hepatology.

Current Data on the Use of Artificial Intelligence in Hepatology in Imaging

Imaging modalities represent a cornerstone in the assessment of liver disease. The ability to assess liver morphology and perfusion in addition to masses or hepatic steatosis as point-of-care testing is unprecedented. Although ultrasound is the first-line technology in clinics today, high-end imaging modalities include magnetic resonance imaging (MRI) proton density fat fraction (PDFF) to assess hepatic steatosis, magnetic resonance elastography to assess liver and spleen stiffness, and phase contrast MRI-enhanced imaging to allow for assessment of blood flow. With these technologies, a complete assessment of liver disease stage is possible and the value of more invasive assessment through liver biopsy has declined. Emerging data link these NITs to clinically relevant outcomes.[7] The wealth of data that are acquired from these imaging modalities make these technologies particularly suitable for postacquisition processing using AI. Only a fraction of the available data is actually used to build an image that informs the clinical decision making.

The Application of Artificial Intelligence to Enhance Ultrasound-Based Diagnostics

Most available research data were generated on the use of AI in ultrasound assessment of liver disease. Convolutional neural networks (CNNs) have shown a very high accuracy of replicating the diagnosis of hepatic steatosis made based on ultrasound B-mode images.[8] In a more recent analysis, DL of raw ultrasound data reached an area under the receiver operating curve (AUROC) of 0.98 when compared with the reference standard MRI-PDFF in detecting hepatic steatosis, even in the absence of phantoms to train imaging acquisition.[9] Detection of significant (≥F2) or advanced (≥F3) fibrosis by ultrasound-based elastography was explored in a study that used 3392 images from 328 cases at the Massachusetts General Hospital. Augmentation of shear wave elastography using a CNN improved the AUROC of

Table 1. Summary of Potential AI Application in Hepatology

AI application in hepatology	Examples of AI algorithms	Limitations
Imaging	- CNNs to diagnose hepatic steatosis based on ultrasound images. - CNNs for automated CT and MRI liver segmentation.	- Variations in data acquisition by different scanners and imaging protocols, and image reconstruction methods.
Histology	- ML algorithms to enable quantitative measurement of NASH histologic features. - ML algorithms to predict response to NASH treatment. - ML algorithms to determine the presence of portal hypertension and predict outcomes.	- Lack of universal standards for digitization of slides, data formatting, image data compression, and storage of meta-data.
Identifying at-risk patients using NITs	- Random forest ML model to predict the stage of fibrosis and identify patients with fibrotic NASH. - The AI-Cirrhosis-ECG score to detect cirrhosis.	- The need for high-quality representative datasets to eliminate the potential for bias.
Predicting outcomes	- Cirrhosis Mortality Model to predict cirrhosis mortality. - ML models to predict graft failure within 30 d from liver transplantation. - Random forest ML model to predict incident HCC. - Primary sclerosing cholangitis risk estimate tool (PREsTo) to predict outcomes in patients with PSC.	- Lack of prospective AI-based randomized clinical trials that demonstrate the added value of AI models in improving clinical outcomes for patients.

AI, artificial intelligence; CNN, convolutional neural network; CT, computer tomography; HCC, hepatocellular carcinoma; ML, machine learning; MRI, magnetic resonance imaging; NASH, nonalcoholic steatohepatitis; NIT, noninvasive test; PSC, primary sclerosing cholangitis.

conventional shear wave elastography from 0.74 to 0.89 for detecting the histologic stages $\geq$F2 by improving image quality, the selection of a region of interest, and classifying the region of interest.[10] When considering ultrasound for the detection of hepatocellular carcinoma (HCC), a deep CNN demonstrated an AUROC of 0.92 for distinguishing benign from malignant liver lesions.[11] Importantly, this was superior to the diagnostic sensitivity and specificity of experienced radiologists, comparable with contrast-enhanced computed tomography (CT), and only slightly inferior to contrast-enhanced MRI.[11]

The Application of Artificial Intelligence to Enhance Computed Tomography– and Magnetic Resonance Imaging–Based Diagnostics

The high-performance metrics of AI and ML in refining diagnostic accuracy for liver disease do not overcome the inherent limitations that specific imaging modalities have. Traditionally, the use of CT imaging to detect hepatic steatosis did not exhibit a high accuracy for mild hepatic steatosis.[12] In a recent analysis using a fully automated volumetric hepatosplenic segmentation algorithm and 3-dimensional CNNs with MRI-PDFF as reference standard, the AUROC to detect mild, moderate, and advanced hepatic steatosis exhibited values of 0.669, 0.854, and 0.962, respectively.[13] Thus, even CNN and high-end CT imaging lack the accuracy to detect mild degrees of hepatic steatosis.

The more advanced imaging modalities outperform ultrasound in the diagnostic accuracy of hepatic steatosis and detection of advanced fibrosis. Whole-liver segmentation is an automated method that uses CNN for imaging biomarkers. In a recent study, MRI-PDFF detecting hepatic steatosis and transverse relaxometry (R2*) detecting iron overload showed an excellent agreement with the histologic lesions in 165 participants of whom 61% had nonalcoholic fatty liver disease (NAFLD).[14] In a smaller study on 62 participants, texture analysis–derived parameters on non-contrast-enhanced T1-weighting was comparable with magnetic resonance elastography to detect advanced versus early fibrosis.[15] Using a generalized CNN automated liver segmentation was feasible even across CT and MRI for automated liver biometry.[16]

An area of special interest in AI-supported imaging is the augmentation of radiology reports to routinely include aspects of liver health and disease, even if the indication to perform the radiologic examination is not in the context of liver disease. The ability of MRIs to detect changes in the nodularity of the liver surface correlates well with the presence of advanced fibrosis on liver histology.[17] Therefore, one clinical application where AI can run in the backend of an imaging server will be to highlight the presence of increased surface nodularity to trigger the radiologist to include the suspicions of cirrhosis in the structured reports.[18]

Despite the promising results of AI algorithms in liver imaging, several issues limit their widespread use including variations in data acquisition by different scanners, imaging protocols, and image reconstruction

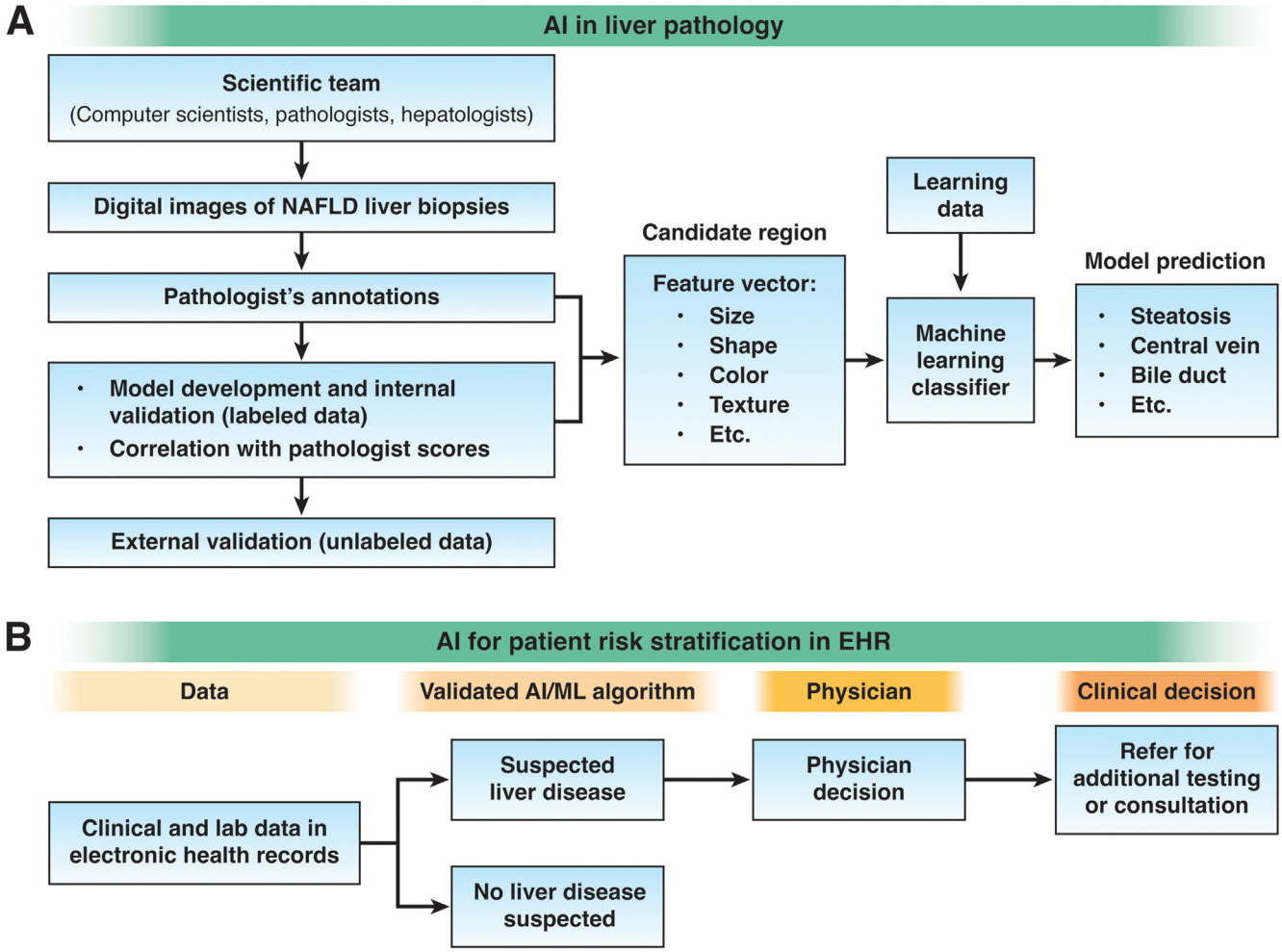

Figure 2. The use of artificial intelligence in hepatology. (*A*) General approach to developing ML models for NAFLD histology analysis. (*B*) Overview of augmentation of clinical decision making based on AI algorithms running in electronic health record systems.

methods to the final selection of radiomic features.[19] Several concrete steps need to be taken to standardize the measurement and analysis of imaging biomarkers. The details of algorithm development including the datasets and computer source code should be shared to ensure transparent translation into the clinical workflow.

Current Data on the Use of Artificial Intelligence in Hepatology in Histology

Determining the Severity of Nonalcoholic Steatohepatitis and Assessing Response to Treatment

Liver biopsy is still considered the gold standard for diagnosing nonalcoholic steatohepatitis (NASH) and fibrosis, although the semiquantitative evaluation of the key histologic features of steatosis, inflammation, ballooning, and fibrosis by the pathologists has been shown to be subjective and prone to major intraobserver and interobserver variability.[20] In 2011, the Food and Drug Administration provided a road map of drug approval including the achievement of 1 of 2 histologic end points as surrogates for outcomes: resolution of NASH without worsening of fibrosis, or regression of fibrosis by 1 stage or more without worsening of NASH.[21] Unfortunately, the reliance on histologic end points has made it difficult to find suitable patients for trials and to reliably assess response to different treatments with clear examples from clinical trials documenting the lack of agreements among expert hepatopathologists.[22] This has created an opportunity to use AI/ML algorithms to develop methods for quantification of the main histologic features of NASH that are less prone to variability by training the algorithm on digitized slides that are annotated by expert pathologists (Figure 2). In 2014, Gawrieh et al[23] published one of the first studies on using supervised ML classifiers to automatically classify white regions in liver biopsies as a method to provide continuous quantitative measurement of macrosteatosis. The ML algorithm performed well with 89% overall accuracy when compared with consensus reading by 2 expert pathologists. The same group developed an AI-based model to quantify liver fibrosis and determine its pattern in patients with NASH

achieving good to excellent correlation between the automatically generated collagen proportionate area and the pathologist fibrosis staging with a coefficient of determination ranging from 0.60 to 0.86.[23] Another group used data from 246 patients with biopsy-proven NASH to develop a high-throughput ML-based quantification of steatosis, ballooning, inflammation, and fibrosis with high interclass correlation coefficient between the manual annotation and the software ranging from 0.97 for steatosis to 0.92 for fibrosis.[24]

In the largest study to date, the PathAI team (Boston, MA) used liver biopsy samples from 3 large NASH clinical trials to build and validate a deep CNN to enable quantitative measurement of NASH histologic severity.[25] The ML parameters also predicted clinical outcomes, such as progression to cirrhosis and hepatic decompensation, and by quantification of more complex features, such as portal inflammation and the ratio area of steatosis to ballooning. Furthermore, a Deep Learning Treatment Assessment (DELTA) Liver Fibrosis score was developed to capture changes in fibrosis severity from baseline to the end of treatment and showed correlation with other noninvasive fibrosis markers, such as the enhanced liver fibrosis score and liver stiffness by transient elastography. The same team at PathAI developed a new ML score to predict hepatic vein pressure gradient (ML-HVPG score) by using biopsies and HVPG measurements from a phase 2b trial. The ML-HVPG score had a stronger correlation with traditional HVPG than collagen proportionate area by morphometry and was able to identify patients with clinically significant portal hypertension with good accuracy (AUROC of 0.85 and 0.76 in the training and test sets, respectively).[26] The advantage of the ML-based score is the fact that HVPG can be estimated from a standard percutaneous liver biopsy without the need for highly specialized interventional radiology procedures and human expertise to interpret HVPG tracing.

In a series of abstracts presented at international meetings, Noureddin et al developed histologic scores via ML in a post hoc analysis of patients from the belapectin phase 2a trial (NCT02462967). This trial provided a cohort of patients with NASH cirrhosis (n = 143) with liver biopsies and phenotype data including HVPG and clinical outcomes. This analysis consisted of discovery and validation cohorts. A second harmonic generation/2-photon excitation fluorescence imaging-based tool provided an automated quantitative assessment of histologic features related to cirrhosis: 252 features related to septa, 21 related to nodules, and 184 related to fibrosis (SNOF). The investigators developed a ML score, SNOF, which significantly correlated with HVPG as a continuous variable ($r = 0.57$ for training and $r = 0.70$ for validation; $P < .05$ for both) and significantly clinically significant portal hypertension (AUROC of 0.85 for training and 0.74 for validation). Investigators also created 2 companion scores: SNOF-V score, which significantly predicted the presence of varices (AUROC of 0.86 for discovery and AUROC of 0.73 for validation cohorts); and SNOF-C score, which identified patients who had >20% change in HVPG 12 months apart with an AUROC of 0.89. Collectively, these data offer a compelling proof-of-concept that ML tools can be applied to liver histology to derive clinically important data that are otherwise difficult to collect from patients.[27,28]

Since then, several AI/ML technologies have been described to assess liver histology in the context of conditional drug approval in the indication NASH. These include second harmonic generation/2-photon excitation to provide quantitative assessment of NASH histologic features on unstained liver histology (Histoindex, Singapore), automated fibrosis quantification from stained slides (Pharmanest, Princeton, NJ), and multiparametric image analysis using proprietary software tools on digitalized histologic slides of entire lobe sections (Biocellvia, Marseille, France).

These examples assert the integration of AI in the interpretation of NASH histology given the decreased variability in interpretation, fast processing of samples, and decreased pathologist workload. It will be important to have universal standardization of digitized slides in terms of data formatting, image data compression, and storage of metadata that will enable future discovery of histopathologic biomarkers.

Diagnosing and Predicting Recurrence of Hepatocellular Carcinoma

Multiple ML algorithms have been generated recently that can diagnose HCC on liver histology and provide risk stratification for recurrence of HCC after surgical resection.[29] For example, Lal et al[30] developed a DL network architecture called NucleiSegNet to grade HCC nuclei on hematoxylin-eosin-stained liver cancer histopathology, which yielded superior results compared with traditional nuclei segmentation methods. An important tool for HCC surgery evaluation is segmentation of hematoxylin-eosin-stained slides by pathologists to assess tumor load before surgical resection and monitor treatment response. Wang et al[31] developed a neural network–based DL model for automatic HCC segmentation that produced high accuracy in 3 public databases.

Unfortunately, 50%–70% of patients with HCC experience tumor recurrence at 5 years postsurgical resection. Saillard et al[32] developed 2 DL models based on whole-slide digitized images for predicting survival after HCC surgical resection and demonstrated better performance of the DL models in comparison with composite scores that used various clinical, pathologic, and biologic factors. Similarly, Yamashita et al[33] developed and validated a DL system called HCC-SurvNet from hematoxylin-eosin-stained digitized slides that could

stratify patients into high- and low-risk groups according to their survival.

Current Data on the Use of Artificial Intelligence in Hepatology in Noninvasive Tests

NITs are the basis of clinical decision making in many patients with liver disease. The interpretation of NITs can be complex based on clinical aspects and requires experience and training. Refinement and augmentation of the interpretation of NITs by AI and ML has been explored in several studies. A study from the Mayo Clinic Rochester used ML to distinguish between alcohol-associated hepatitis and acute cholangitis based on 10 commonly used laboratory variables including white blood cell count, hemoglobin, mean corpuscular volume, platelet count, aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase, total bilirubin, direct bilirubin, and albumin.[34] Exploring 265 case of alcohol-associated hepatitis and 194 cases of acute cholangitis, the ML algorithm demonstrated an excellent performance with an AUROC of 0.932. Interestingly, an online survey of physicians using the same 10 variables resulted in an inferior diagnostic accuracy with an AUROC of 0.790 highlighting the strength of ML.[34] A comparable approach was taken to improve the overall low rate of identification of patients with NAFLD and NASH. An AI algorithm that uses clinical data and standard laboratory studies available in electronic health care records (EHRs) is the NASHmap algorithm. NASHmap was built using the National Institute of Diabetes and Digestive and Kidney Disease dataset on 704 patients with histologically defined NASH and validated in the larger OPTUM EHR using an eXtreme Gradient Boosting model (XGBoost) consisting of 14 features. The performance of this 14-feature model to predict cases with histologic NASH was good with an AUROC of 0.82.[35] When decreasing the number of variables to 5 using only hemoglobin A_{1c}, AST, ALT, total protein, and triglycerides, the AUROC to predict NASH was still 0.79.[35] In a smaller analysis from Japan, the separation of NAFLD from non-NAFLD cases was achieved by using 11 clinical variables including age, sex, height, weight, waist circumference, AST, ALT, γ-glutamyltransferase, cholesterol, triglyceride, and platelet count. Here the AUROC in the validation cohort was 0.950. Through extension of clinical data with multiomics (genomic, transcriptomic, proteomic, and metabolomic) data, a recent analysis that aimed to detect liver fat content >5% on MRI in 1514 participants used LASSO (least absolute shrinkage and selection operator) and yielded an AUROC of 0.84.[36] More recently, ML models including logistic regression, random forest (RF), and artificial neural network (ANN) were implemented to predict the histologic stages of fibrosis and the presence of fibrotic NASH using demographic/clinical features in 1370 patients with biopsy-proven disease. The ML models performed better overall than traditional NITs for fibrosis[37] with the AUROC for the RF model versus vibration controlled transient elastography by FibroScan and the FIB-4 index for ≥F2, ≥F3, and F4 being (0.86 vs 0.81; 0.78), (0.89 vs 0.83; 0.82), and (0.89 vs 0.86; 0.85), respectively. The ML model included 17 readily available clinical variables, such as age, body mass index, AST, platelet count, and the presence of diabetes. This model can be implemented easily in the EHR system to identify patients with fibrotic NASH that may benefit from pharmacologic treatment. In fact, the RF model performed as well as the FibroScan-based FAST score in predicting fibrotic NASH with no statistically significant difference in accuracy, AUROC, sensitivity, specificity, and positive- and negative-predictive values.

Importantly, these algorithms are not fit to substitute for a clinical diagnosis but will allow to flag cases that are otherwise not recognized as "at-risk of liver disease" and could be particularly useful in primary care to support the selection of referral cases (Figure 2).

One of the big challenges in nonspecialty clinics is the identification of patients with compensated cirrhosis. These cases are typically difficult to diagnose using standard clinical variables and laboratory tests. In a large multicenter cohort with validation in a clinical trial cohort, a model including international normalized ratio, γ-glutamyltransferase, ALT, platelets, and age discriminated best between patients with bridging fibrosis and cirrhosis with an AUROC of 0.733 (95% confidence interval, 0.671–0.795).[38] An approach that explored the liver-heart axis to detect patients with advanced cirrhosis was developed based on distinct abnormalities that are detectable on electrocardiogram (ECG) tracings. The AI-Cirrhosis-ECG score was developed in 5212 patients that underwent liver transplantation at the 3 Mayo Clinic transplant centers between 1988 and 2019. The model distinguished advanced cirrhosis from control subjects with an AUROC of 0.908 (84.9% sensitivity; 83.2% specificity) based on an analysis of ECG tracings independent of comorbidities. These approaches could lead to the development of low-cost tools in the care of patients when expert hepatology advise might not be available.[39] A comparable approach aiming at identifying liver disease in a primary care setting was explored by implementation of automated analysis in a laboratory management system. The Intelligent liver function testing (iLFT) study introduced reflex testing of blood samples when increased liver enzymes were detected.[40] Through this approach, detection rates of liver disease increased by 43% compared with historic rates. In addition, an automated diagnosis and management plan was introduced resulting in a cost-effective implementation in the UK National Health Service mostly through decreasing unnecessary referrals.[40]

Prediction of HCC by AI is another intensively studied area.[29] In a large-scale analysis using clinical data and laboratory values in 48,151 patients with hepatitis C virus–related cirrhosis in the Veterans Affairs cohorts, the ability of recurrent neural network models to predict

HCC exhibited an AUROC of 0.759.[41] Importantly, this recurrent neural network used 4 baseline variables and 27 longitudinal variables.[41]

Other Potential Uses of Artificial Intelligence and Machine Learning in Hepatology

Predicting Cirrhosis Outcomes and Mortality

Several traditional risk scores have been developed to estimate mortality risk in cirrhosis with the Model for End-Stage Liver Disease (MELD) score being the most commonly used.[42] MELD was developed to predict 90-day mortality, which may be too short of a window for most patients with cirrhosis except for those listed for transplantation. Better understanding of long-term prognosis can inform patients and caregivers preferences and goals of care and improve decision making for clinicians. ML models have potential to greatly enhance the predictive accuracy and improve prognostication compared with traditional models but their black box analytics have limited their use. Therefore, Kanwal et al[43] used a hybrid approach that tested 3 different ML algorithms in a retrospective large cohort from 130 hospitals to predict cirrhosis mortality, and then developed a blended model called the Cirrhosis Mortality Model that used selected variables that were implemented in an accessible platform. Cirrhosis Mortality Model outperformed MELD for predicting 1-, 2-, and 3-year mortality with the AUROC for 1-year mortality being 0.78 for Cirrhosis Mortality Model and 0.67 for MELD ($P < .001$).

Readmissions are common in patients with cirrhosis and represent major challenge for families and the health care system overall, making adequate prediction of patients at high-risk for readmission within 90 days from discharge a high priority.[44] However, a recent study that trained 3 AI models including logistic regression, RF, and kernel support vector machine to predict 90-day readmission rates demonstrated only modest accuracy with AUROC of 0.59, 0.62, and 0.62, respectively.[45] These findings underscore the fact that readmissions in cirrhosis are influenced by the complex interactions of clinical, psychosocial, and financial factors that may not be captured by the EHR data. In the future, the use of AI to integrate data from socioeconomic and biologic variables may help improve prognostication and classification accuracy.[46]

Predicting Graft Survival after Liver Transplantation

Given the increasing number of patients in need for liver transplantation and the limited supply of deceased donor livers, there is an urgent need to optimize listing and allocation decisions to maximize utility.[47] Traditional predictive models, such as donor risk index and the survival outcomes after liver transplantation, do not capture the complex interactions among donor, recipient, and the transplantation process that are known to affect outcomes.[48] Therefore, there is an opportunity for AI/ML techniques to improve predictive accuracy based on donor and recipient variables that are known at the time of organ allocation. Lau et al[49] developed ML models (RF and ANN) to predict graft failure within 30 days from transplantation in a dataset that included 180 transplants and 276 available donor and recipient variables. They also directly compared their ML models with traditional liver scores and found that their ANN model was superior to donor risk index and survival outcomes after liver transplantation (AUROC of 0.84 compared with 0.68 and 0.64, respectively). The difference of 0.16 in AUROC values between the ANN model and the donor risk index can be considered clinically significant to provide proof of concept that the ML model could be used to support decision making in organ allocation. More recently, a systematic review evaluated the use of ML methodology to predict graft outcome following liver transplantation.[50] Nine studies met the inclusion criteria and reported outcomes from 18,771 liver transplants with ANNs being the most commonly used ML methodology (7/9 studies).

Despite promising results, there are several challenges to implementing ML models for organ allocation. First, the issue of biologic plausibility when the algorithm ends up including variables that are not considered biologically or clinically relevant. Second, the issue of generalizability and the fact that the ML algorithms perform best for predicting outcomes on the dataset from which they were derived and may not perform as well on a global scale.

Improving Hepatocellular Carcinoma Early Detection and Outcomes

AI has the potential to transform the full spectrum of HCC clinical care by providing improved HCC risk prediction, diagnosis, and prognostication/response to treatment assessment. For more detailed discussion on the role of AI in HCC management, we refer readers to an excellent recent review by Calderaro et al.[29] Several studies have applied AI techniques to longitudinal EHR data to predict incident HCC in high-risk populations including an RF ML algorithm that had AUROC of 0.64, which significantly outperformed the conventional HALT-C model for predicting HCC.[51] Another area where ML algorithm may improve HCC care includes the prognostication of established HCC and predicting response to locoregional therapies. Wu et al[52] built an ANN model that included 15 clinical variables to predict 1- and 2-year survival of patients with HCC who received radiofrequency ablation and showed good accuracy with AUROC of 0.84. Similarly, another group developed an ANN model to predict patient survival at 1 year after transarterial chemoembolization for HCC with promising performance (AUROC of 0.77).[53]

Early Diagnosis and Predicting Outcomes in Other Liver Diseases

Primary sclerosing cholangitis (PSC) is a progressive cholestatic liver disease that is associated with advanced fibrosis, portal hypertension, and higher risk for colorectal and hepatobiliary malignancies.[54] Given the heterogeneous nature of the disease and the overall low event rate over a short period of time, developing predictive models that can simplify the design of clinical trials is a high priority.[55] The Primary Sclerosing Cholangitis Risk Estimate Tool (PREsTo) leveraged gradient boosting, an ML technique, to predict hepatic decompensation (ascites, variceal hemorrhage, or encephalopathy) in 2 large cohorts of patients with PSC.[56] PREsTo included 9 clinical variables (bilirubin, albumin, alkaline phosphatase times the upper limit of normal, platelets, AST, hemoglobin, sodium, age, and the duration of PSC diagnosis in years) and was able to accurately predict decompensation compared with other traditional methods (AUROC of 0.90 compared with 0.72 for MELD and 0.85 for Mayo PSC risk score).

Similarly, ML was applied to a large international database of patients with primary biliary cholangitis to predict the outcome of liver-related death or transplantation.[57] ML identified 4 new subsets of patients with primary biliary cholangitis with different phenotypes and prognosis and highlighted that ursodeoxycholic acid–induced increase of albumin was associated with improved transplant-free survival.

Hereditary hemochromatosis is associated with iron accumulation in the liver, which if left untreated can result in progression to cirrhosis and the development of HCC.[58] Hence, early diagnosis is crucial and can be achieved by using AI/ML approaches to heterogeneous large datasets. In fact, a recent study tested 7 popular ML algorithms in the Hemochromatosis and Iron Overload Screening (HEIRS) cohort, which consists of 254 cases and 701 control subjects, to determine which combination of risk factors and algorithm provided the best performance in the diagnosis of hereditary hemochromatosis.[59] The final model was developed using extreme gradient boosting and included the following variables: HFE C282Y homozygosity, age, mean corpuscular volume, iron level, serum ferritin level, transferrin saturation, and unsaturated iron-binding capacity. The AUROC for the new model was 0.94 and outperformed the traditional iron overload screening (IRON) tool (AUROC of 0.60).

Alcohol-related liver disease (ALD) is now the leading indication for liver transplantation in the United States[60] and most patients are diagnosed after decompensation.[61] Therefore, the identification of ALD at an earlier stage could provide the opportunity to prevent disease progression through alcohol cessation treatment programs and the management of metabolic comorbidities, such as obesity and type 2 diabetes. Niu et al[62] explored the use of mass spectrometry/plasma proteomics supported by ML to identify biomarkers for the diagnosis and prognostication of early ALD. The ML models selected a panel of proteins that detected significant fibrosis with high accuracy (AUROC of 0.92) and outperformed standard-of-care clinical tests, such as the FIB-4 index and APRI. The same proteomic-based panel was predictive of the development of major adverse liver outcomes and overall mortality. Another study from Denmark used a novel AI approach to identify significant fibrosis in populations with low prevalence.[63] The authors used routine clinical variables that were available in a prospective cohort of 3352 asymptomatic subjects, of whom 35% were at risk for ALD. The AI models accurately identified patients with elevated liver stiffness more than 8 kPa (AUC of 0.86–0.94) and were superior to conventional blood-based indices (AUC, 0.60–0.76: $P < .01$).

Challenges to the Use of Artificial Intelligence in Hepatology and Immediate Research Priorities

Despite the great potential for AI-based models to improve care for patients with liver disease, several challenges need to be addressed before transitioning these models from research to bedside. There is a rising concern that applying computer algorithms in health care may introduce and further amplify health inequities and biases against vulnerable populations.[64] Several biases can be introduced through AI algorithms including biases related to the selection of the clinical research problem, biases in data collection and variable selection, and biases in algorithm development and postdevelopment use.[65]

To increase equity in AI use in health care, tools need to be developed to debias data collection, model training and output, and clinical application. Engaging health equity experts early in the process of algorithm development is of utmost importance. In addition, ethical frameworks and regulatory standards need to be developed by the Food and Drug Administration and other agencies, such as the mandatory reporting of the racial and socioeconomic status of the patient population used in the AI model development. This will ensure adequate representation of certain minorities and the generalizability of the results. The ideal scenario is the development of an augmented intelligence approach that combines the physician decision making with the AI tools to eliminate bias and ensure health equity.[66]

To fully implement AI in hepatology, prospective AI-based randomized clinical trials are necessary to unequivocally demonstrate the added value of AI models in improving clinical outcomes for patients. In the endoscopy field, several trials have demonstrated the impact of computer-aided algorithms on increasing polyp and adenoma detection rates during colonoscopy.[67]

In conclusion, we firmly believe that AI/ML models will play an integral part in the identification and management of liver diseases in the near future. Therefore, providing adequate training for practicing physicians on the concepts of AI and how to interpret the output of AI models is of paramount importance. In addition, investments in health service infrastructure, data safety, and ethical studies of AI are required to generate the level of trust with physicians and patient to accepted AI-augmented clinical decision making.

References

1. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019;28:73–81.
2. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019;25:24–29.
3. Sanyal AJ, Shankar SS, Calle RA, et al. Non-invasive biomarkers of nonalcoholic steatohepatitis: the FNIH NIMBLE project. Nat Med 2022;28:430–432.
4. Hardy T, Wonders K, Younes R, et al. The European NAFLD Registry: a real-world longitudinal cohort study of nonalcoholic fatty liver disease. Contemp Clin Trials 2020;98:106175.
5. Xiang Y, Zhao L, Liu Z, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. Artif Intell Med 2020;102:101780. https://doi.org/10.1016/j.artmed.2019.101780.
6. Nazir S, Dickson DM, Akram MU. Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks. Comput Biol Med 2023;156:106668.
7. Gidener T, Ahmed OT, Larson JJ, et al. Liver stiffness by magnetic resonance elastography predicts future cirrhosis, decompensation, and death in NAFLD. Clin Gastroenterol Hepatol 2021;19:1915–1924.
8. Zamanian H, Mostaar A, Azadeh P, et al. Implementation of combinational deep learning algorithm for non-alcoholic fatty liver classification in ultrasound images. J Biomed Phys Eng 2021;11:73–84.
9. Han A, Byra M, Heba E, et al. Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat with radiofrequency ultrasound data using one-dimensional convolutional neural networks. Radiology 2020;295:342–350.
10. Brattain LJ, Telfer BA, Dhyani M, et al. Objective liver fibrosis estimation from shear wave elastography. Annu Int Conf IEEE Eng Med Biol Soc 2018;2018:1–5.
11. Yang Q, Wei J, Hao X, et al. Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: a multicentre study. EBioMedicine 2020;56:102777.
12. Lawrence DA, Oliva IB, Israel GM. Detection of hepatic steatosis on contrast-enhanced CT images: diagnostic accuracy of identification of areas of presumed focal fatty sparing. AJR Am J Roentgenol 2012;199:44–47.
13. Pickhardt PJ, Blake GM, Graffy PM, et al. Liver steatosis categorization on contrast-enhanced CT using a fully automated deep learning volumetric segmentation tool: evaluation in 1204 healthy adults using unenhanced CT as a reference standard. AJR Am J Roentgenol 2021;217:359–367.
14. Martí-Aguado D, Jiménez-Pastor A, Alberich-Bayarri Á, et al. Automated whole-liver MRI segmentation to assess steatosis and iron quantification in chronic liver disease. Radiology 2022;302:345–354.
15. Schawkat K, Ciritsis A, von Ulmenstein S, et al. Diagnostic accuracy of texture analysis and machine learning for quantification of liver fibrosis in MRI: correlation with MR elastography and histopathology. Eur Radiol 2020;30:4675–4685.
16. Wang K, Mamidipalli A, Retson T, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. Radiol Artif Intell 2019;1:180022.
17. Dioguardi Burgio M, Sartoris R, Beaufrere A, et al. Liver surface nodularity on non-contrast MRI identifies advanced fibrosis in patients with NAFLD. Eur Radiol 2022;32:1781–1791.
18. Schattenberg JM, Emrich T. Refining imaging tools to detect advanced fibrosis: could liver surface nodularity address an unmet need in the NAFLD epidemic? Eur Radiol 2022;32:1757–1759.
19. Nam D, Chapiro J, Paradis V, et al. Artificial intelligence in liver diseases: improving diagnostics, prognostics and response prediction. JHEP Rep 2022;4:100443.
20. Gawrieh S, Knoedler DM, Saeian K, et al. Effects of interventions on intra- and interobserver agreement on interpretation of nonalcoholic fatty liver disease histology. Ann Diagn Pathol 2011;15:19–24.
21. Sanyal AJ, Brunt EM, Kleiner DE, et al. Endpoints and clinical trial design for nonalcoholic steatohepatitis. Hepatology 2011;54:344–353.
22. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol 2020;73:1322–1332.
23. Gawrieh S, Sethunath D, Cummings OW, et al. Automated quantification and architectural pattern detection of hepatic fibrosis in NAFLD. Ann Diagn Pathol 2020;47:151518.
24. Forlano R, Mullish BH, Giannakeas N, et al. High-throughput, machine learning-based quantification of steatosis, inflammation, ballooning, and fibrosis in biopsies from patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2020;18:2081–2090.
25. Taylor-Weiner A, Pokkalla H, Han L, et al. A machine learning approach enables quantitative measurement of liver histology and disease monitoring in NASH. Hepatology 2021;74:133–147.
26. Bosch J, Chung C, Carrasco-Zevallos OM, et al. A machine learning approach to liver histological evaluation predicts clinically significant portal hypertension in NASH cirrhosis. Hepatology 2021;74:3146–3160.
27. Noureddin M, Tai D, Chng E, et al. Derivation of machine learning histologic scores correlating with portal pressures and the development of varices in NASH patients with cirrhosis. J Hepatol 2022;77:S623–S624.
28. Noureddin M, Goodman Z, Tai D, et al. Machine learning liver histology scores correlate with portal hypertension assessments in nonalcoholic steatohepatitis cirrhosis. Aliment Pharmacol Ther 2023;57:409–417.
29. Calderaro J, Seraphin TP, Luedde T, et al. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. J Hepatol 2022;76:1348–1361.
30. Lal S, Das D, Alabhya K, et al. NucleiSegNet: robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. Comput Biol Med 2021;128:104075.

31. Wang X, Fang Y, Yang S, et al. A hybrid network for automatic hepatocellular carcinoma segmentation in H&E-stained whole slide images. Med Image Anal 2021;68:101914.
32. Saillard C, Schmauch B, Laifa O, et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. Hepatology 2020;72:2000–2013.
33. Yamashita R, Long J, Saleem A, et al. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. Sci Rep 2021;11:2047.
34. Ahn JC, Noh YK, Rattan P, et al. Machine learning techniques differentiate alcohol-associated hepatitis from acute cholangitis in patients with systemic inflammation and elevated liver enzymes. Mayo Clin Proc 2022;97:1326–1336.
35. Docherty M, Regnier SA, Capkun G, et al. Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. J Am Med Inform Assoc 2021; 28:1235–1241.
36. Atabaki-Pasdar N, Ohlsson M, Viñuela A, et al. Predicting and elucidating the etiology of fatty liver disease: a machine learning modeling and validation study in the IMI DIRECT cohorts. PLoS Med 2020;17:e1003149.
37. Chang D, Truong E, Mena EA, et al. Machine learning models are superior to noninvasive tests in identifying clinically significant stages of NAFLD and NAFLD-related cirrhosis. Hepatology 2023;77:546–557.
38. Labenz C, Toenges G, Zheng MH, et al. Derivation and validation of the nonalcoholic fatty liver disease cirrhosis score (NCS) to distinguish bridging fibrosis from cirrhosis. Eur J Intern Med 2022;98:53–60.
39. Ahn JC, Attia ZI, Rattan P, et al. Development of the AI-Cirrhosis-ECG Score: an electrocardiogram-based deep learning model in cirrhosis. Am J Gastroenterol 2022;117:424–432.
40. Dillon JF, Miller MH, Robinson EM, et al. Intelligent liver function testing (iLFT): a trial of automated diagnosis and staging of liver disease in primary care. J Hepatol 2019;71:699–706.
41. Ioannou GN, Tang W, Beste LA, et al. Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis C cirrhosis. JAMA Netw Open 2020;3:e2015626.
42. Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. N Engl J Med 2008;359:1018–1026.
43. Kanwal F, Taylor TJ, Kramer JR, et al. Development, validation, and evaluation of a simple machine learning model to predict cirrhosis mortality. JAMA Netw Open 2020;3:e2023780.
44. Tapper EB. Challenge accepted: Confronting readmissions for our patients with cirrhosis. Hepatology 2016;64:26–28.
45. Hu C, Anjur V, Saboo K, et al. Low predictability of readmissions and death using machine learning in cirrhosis. Am J Gastroenterol 2021;116:336–346.
46. Bajaj JS, Reddy KR, Tandon P, et al. Association of serum metabolites and gut microbiota at hospital admission with nosocomial infection development in patients with cirrhosis. Liver Transpl 2022;28:1831–1840.
47. Volk ML, Roney M, Merion RM. Systematic bias in surgeons' predictions of the donor-specific risk of liver transplant graft failure. Liver Transpl 2013;19:987–990.
48. Flores A, Asrani SK. The donor risk index: a decade of experience. Liver Transpl 2017;23:1216–1225.
49. Lau L, Kankanige Y, Rubinstein B, et al. Machine-learning algorithms predict graft failure after liver transplantation. Transplantation 2017;101:e125–e132.
50. Wingfield LR, Ceresa C, Thorogood S, et al. Using artificial intelligence for predicting survival of individual grafts in liver transplantation: a systematic review. Liver Transpl 2020; 26:922–934.
51. Singal AG, Mukherjee A, Elmunzer BJ, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol 2013;108:1723–1730.
52. 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;116:765–773.
53. Mahringer-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.
54. Bowlus CL, Arrive L, Bergquist A, et al. AASLD practice guidance on primary sclerosing cholangitis and cholangiocarcinoma. Hepatology 2023;77:659–702.
55. Ponsioen CY, Chapman RW, Chazouilleres O, et al. Surrogate endpoints for clinical trials in primary sclerosing cholangitis: review and results from an International PSC Study Group consensus process. Hepatology 2016;63:1357–1367.
56. Eaton JE, Vesterhus M, McCauley BM, et al. Primary Sclerosing Cholangitis Risk Estimate Tool (PREsTo) predicts outcomes of the disease: a derivation and validation study using machine learning. Hepatology 2020;71:214–224.
57. Gerussi A, Verda D, Bernasconi DP, et al. Machine learning in primary biliary cholangitis: a novel approach for risk stratification. Liver Int 2022;42:615–627.
58. Bacon BR, Adams PC, Kowdley KV, et al. Diagnosis and management of hemochromatosis: 2011 practice guideline by the American Association for the Study of Liver Diseases. Hepatology 2011;54:328–343.
59. Martins Conde P, Sauter T, Nguyen TP. An efficient machine learning-based approach for screening individuals at risk of hereditary haemochromatosis. Sci Rep 2020;10:20613.
60. Noureddin M, Vipani A, Bresee C, et al. NASH leading cause of liver transplant in women: updated analysis of indications for liver transplant and ethnic and gender variances. Am J Gastroenterol 2018;113:1649–1659.
61. Nilsson E, Anderson H, Sargenti K, et al. Clinical course and mortality by etiology of liver cirrhosis in Sweden: a population based, long-term follow-up study of 1317 patients. Aliment Pharmacol Ther 2019;49:1421–1430.
62. Niu L, Thiele M, Geyer PE, et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. Nat Med 2022; 28:1277–1287.
63. Blanes-Vidal V, Lindvig KP, Thiele M, et al. Artificial intelligence outperforms standard blood-based scores in identifying liver fibrosis patients in primary care. Sci Rep 2022;12:2914.
64. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447–453.
65. Uche-Anya E, Anyane-Yeboa A, Berzin TM, et al. Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. Gut 2022; 71:1909–1915.
66. Rajkomar A, Hardt M, Howell MD, et al. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169:866–872.

67. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–1819.

Correspondence
Address correspondence to: Naim Alkhouri, MD, Arizona Liver Health, 3051 East Rock Wren Road, Phoenix, Arizona 85048. e-mail: nalkhouri@azliver.com.

CRediT Authorship Contributions

Jörn M. Schattenberg (Conceptualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Naga Chalasani (Conceptualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Naim Alkhouri, MD (Conceptualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Conflicts of interest

The authors disclose the following: Jörn M. Schattenberg reports consultancy for Apollo Endosurgery, Albireo Pharma Inc, Bayer, BMS, Boehringer Ingelheim, Gilead Sciences, GSK, Intercept Pharmaceuticals, Ipsen, Inventiva Pharma, Madrigal, MSD, Northsea Therapeutics, Novartis, Novo Nordisk, Pfizer, Roche, Sanofi, and Siemens Healthcare GmbH; research funding from Gilead Sciences, Boehringer Ingelheim, and Siemens Healthcare GmbH; and Speaker Honorarium from Boehringer Ingelheim, Echosens, MedPublico GmbH, Novo Nordisk, and Madrigal Pharmaceuticals. Naga Chalasani has paid consulting agreements within the past 24 months with AbbVie, Lilly, Madrigal, Zydus, Galectin, Boehringer-Ingelheim, GSK, Foresite, and Altimmune; and research grants from Exact Sciences and DSM. Naim Alkhouri has research funding from 89Bio, Akero, AbbVie/Allergan, Better Therapeutics, Boehringer Ingelheim, Bristol-Myers Squibb, Corcept, DSM, Galectin, Genentech, Genfit, Gilead, Hepagene, Healio, Intercept, Inventiva, Ionis, Madrigal, Merck, NGM, Noom, NorthSea, Novo Nordisk, Perspectum, Pfizer, Poxel, Viking, and Zydus; is on the speaker bureau for AbbVie/Allergan, Alexion, Echosens, Eisai, Exelixis, Gilead, Intercept, Perspectum, Salix, and Theratechnologies; and is a consultant for AbbVie/Allergan, Echosens, Fibronostics, Gilead, Intercept, Madrigal, Novo Nordisk, Perspectum, Pfizer, and Zydus.