Journal of Gastroenterology and Hepatology

JGHF

doi:10.1111/jgh.15385

SPECIAL SECTION ARTICLE

Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis

Grace Lai-Hung Wong,*,†,‡ Pong-Chi Yuen,§ Andy Jinhua Ma,§ Anthony Wing-Hung Chan,¶ Howard Ho-Wai Leung¶ and Vincent Wai-Sun Wong*,†,‡

*Department of Medicine and Therapeutics, †Medical Data Analytic Centre (MDAC), ‡Institute of Digestive Disease, ¶Department of Anatomical and Cellular Pathology, The Chinese University of Hong Kong, Shatin, and §Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong

Key words
Cirrhosis, Liver fibrosis, Machine learning, Non-alcoholic steatohepatitis (NASH).

Accepted for publication 20 December 2020.

Correspondence
Grace Lai-Hung Wong, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, 9/F Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, Hong Kong.
Email: wonglaihung@cuhk.edu.hk

Pong-Chi Yuen, Department of Computer Science, Hong Kong Baptist University, Room R702, Sir Run Shaw Building, Kowloon Tong, Hong Kong.
Email: pcyuen@comp.hkbu.edu.hk

Declaration of conflict of interest: Vincent Wai-Sun Wong has served as a speaker for Abbvie, Bristol Myers Squibb, Roche, Novartis, Abbott Diagnostics, and Echosens and an advisory committee member for Abbvie, Roche, Novartis, Gilead, and Otsuka.

Abstract

Artificial intelligence (AI) has become increasingly widespread in our daily lives, including healthcare applications. AI has brought many new insights into better ways we care for our patients with chronic liver disease, including non-alcoholic fatty liver disease and liver fibrosis. There are multiple ways to apply the AI technology on top of the conventional invasive (liver biopsy) and noninvasive (transient elastography, serum biomarkers, or clinical prediction models) approaches. In this review article, we discuss the principles of applying AI on electronic health records, liver biopsy, and liver images. A few common AI approaches include logistic regression, decision tree, random forest, and XGBoost for data at a single time stamp, recurrent neural networks for sequential data, and deep neural networks for histology and images.

Grace Lai-Hung Wong has served as a speaker for Abbott, Abbvie, Bristol Myers Squibb, Echosens, Furui, Gilead, and Otsuka and an advisory committee member for Otsuka and Gilead. The other authors declare that they have no competing interests.

Author contribution: All authors were responsible for the conception, interpretation of data, writing, review, and final approval of the manuscript.

Financial support: This study was funded in part by the Health and Medical Research Fund (HMRF) of Food and Health Bureau, Hong Kong (project reference number: 07180216) to Grace Lai-Hung Wong.

Guarantor of the article: Grace Lai-Hung Wong.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is currently the most common chronic liver disease worldwide and affects at least 25% of the global adult population.[1] An updated systematic review suggests that the prevalence of NAFLD in Asia has reached 33.9% in 2012–2017.[2] In Western countries, NAFLD and its active form non-alcoholic steatohepatitis (NASH) have become one of the leading indications for liver transplantation and an important cause of hepatocellular carcinoma.[3]

Despite the importance of NAFLD at the population level, only a small proportion of patients would eventually suffer from cirrhotic complications and die from liver disease.[4] In NAFLD patients at low risk of developing liver-related morbidity and mortality, it is doubtful if liver-specific treatments are warranted. It is therefore important to identify high-risk patients and utilize healthcare resources wisely.

In patients with NAFLD, the presence of type 2 diabetes and the degree of liver fibrosis have the strongest correlation with adverse hepatic outcomes.[5] Although cardiovascular disease is the most common cause of death in NAFLD, a multicenter study of patients with biopsy-proven NAFLD suggests that cirrhotic complications, hepatocellular carcinoma, and liver failure are the leading cause of death in those with advanced liver fibrosis and cirrhosis.[6] Therefore, all phase 3 clinical trials for NASH rightfully focus on patients with at least stage 2 fibrosis.[7] It is also expected that this would be the target treatment population when a drug for NASH becomes available.

Current methods to assess NAFLD and liver fibrosis have various limitations. There is thus much interest in improving assessments using artificial intelligence (AI). Currently, the US Food and Drug Administration Biomarkers, EndpointS and other Tools (BEST) puts biomarkers into different categories, all of which AI may have valuable contributions (Table 1). In this article, we

Table 1 The US Food and Drug Administration Biomarkers, EndpointS and other Tools (BEST) biomarker categories and potential contribution from AI in its application in NAFLD

Categories	Meaning	Potential role of AI
Susceptibility/risk	Potential for developing a disease or medical condition	- Predicting incident NAFLD and fibrosis progression with dynamic clinical and laboratory data
Diagnostic	Detect or confirm presence of a disease or condition	- Highlighting patients likely to have NAFLD and significant liver fibrosis based on clinical and laboratory data - Highlighting areas suspicious of hepatocellular carcinoma in imaging studies - Assisting histological interpretation of NAFLD by offering suggested scores or highlighting histological features of interest (e.g. hepatocyte ballooning and fibrosis) - Serving as a high-throughput and reproducible alternative to pathologists' assessment in clinical trials involving thousands of images
Monitoring	Serially for assessing status of a disease or medical condition	- Highlighting suspicion of disease progression based on dynamic clinical and laboratory data - Serving as a high-throughput and reproducible alternative to pathologists' assessment in clinical trials and routine practice to detect changes in key histological features
Prognostic	Likelihood of a clinical event, disease recurrence or progression	- Predicting future cirrhosis, hepatocellular carcinoma, and liver-related complications based on dynamic clinical and laboratory data
Predictive	Favorable or unfavorable effect from exposure to a medical product or an environmental agent	- Identifying potential effective treatments or harmful exposures in real-world big data studies
Pharmacodynamic/ response	Biological response has occurred in an individual who has been exposed to a medical product or an environmental agent	- Identifying markers or signatures indicating favorable treatment response in clinical trials and real-world practice
Safety	Likelihood, presence, or extent of toxicity as an adverse effect	- Identifying markers or signatures predicting adverse events to a therapeutic agent in clinical trials and real-world practice

AI, artificial intelligence; NAFLD, non-alcoholic fatty liver disease.

discuss the principles of AI as it is applied in NAFLD and highlight some initial developments in this exciting field.

Current methods to assess non-alcoholic fatty liver disease and liver fibrosis

In routine practice, the diagnosis of NAFLD is usually based on radiological detection of fatty liver (mostly by abdominal ultrasonography) and the exclusion of excessive alcohol consumption, secondary causes of fatty liver, and concomitant liver diseases.[8] Noninvasive tests or liver biopsy are then performed to assess the severity of liver disease. Details of noninvasive tests of NAFLD and liver fibrosis have been reviewed extensively elsewhere and are beyond the scope of this review.[9] Suffice it to say, among noninvasive tests of fibrosis, specific fibrosis biomarkers (e.g. Pro-C3 and enhanced liver fibrosis biomarker) and imaging studies (ultrasound elastography and magnetic resonance elastography) have already achieved moderate to high accuracies for advanced liver fibrosis and have set a high bar for new diagnostic algorithms by AI. That said, there is certainly room for improvement; the existing noninvasive tests tend to have high negative predictive values in excluding advanced fibrosis but rather modest positive predictive values in confirming the diagnosis. Moreover, while the existing noninvasive tests have a reasonable performance in cross-sectional studies, few have demonstrated sufficient accuracy as a monitoring tool during longitudinal follow-up.

At present, histological improvements in terms of resolution of NASH without worsening of fibrosis or improvement in fibrosis without worsening of NASH remain key endpoints in phase 3 clinical trials for a drug to receive conditional approval from the US Food and Drug Administration subpart H pathway and the European Medicines Agency.[7] Nonetheless, previous studies have demonstrated significant intraobserver and interobserver variability in histological scoring, particularly when suboptimal biopsy specimens are included. Compounded with the variabilities in two serial biopsies, the risk of misclassifying treatment response (or no response) is considerable.[10] In addition, a central pathologist may need to review hundreds to thousands of slides within a tight time window. In this situation, augmentation by AI to produce quick and reproducible scores is particularly attractive.

Principles of artificial intelligence

Basic principles. Artificial intelligence including machine learning algorithms and deep neural networks has been increasingly applied for the detection and prediction of NAFLD and fibrosis outcomes in recent years. With (large-scale) training cohorts for model development by using AI, the prediction on validation cohorts is probably more accurate compared with traditional biostatistical methods.[11] AI is a data-driven and

hypothesis-free approach, which better incorporates any number of clinical factors to detect hidden patterns for disease detection/prediction. Many publicly available platforms can be used to easily build up AI models such as WEKA (University of Waikato, Hamilton, New Zealand), Matlab (MathWorks, Natick, Massachusetts, USA), Tensorflow (Google Brain Team, Menlo Park, California, USA), and Pytorch (Facebook's AI Research lab [FAIR], Menlo Park, California, USA).

Artificial intelligence in electronic health record analysis. Electronic health records (EHRs) contain patient information of sex, ethnicity, body mass index, genetics, laboratory parameters (e.g. viruses and chemicals), comorbidity (such as diabetes), and so forth. For this type of data at a single time stamp, the number of variables is usually small (compared with image data), so traditional machine learning methods as shown in Figure 1a have been widely employed to detect or predict liver diseases including NAFLD and fibrosis.[11] For example, logistic regression, decision tree, random forest, and XGBoost (a gradient-boosted tree model) are evaluated for the task of NASH and NAFLD detection.[12] Logistic regression with logistic loss function is one of the binary classifiers widely used in various medical applications. For development of easily interpretable classification model, decision tree trains a tree-like classifier in which each node depends on a variable to make a decision and assign the input with a class label or the following node. To increase the generalization accuracy on validation cohorts over decision tree, random forest ensembles multiple decision trees trained on randomly selected features and randomly sampled subset of the training cohort. To further increase the detection/prediction performance, XGBoost aggregates each decision tree by correcting the residual error of the preceding ensemble. According to the results in Fialoke *et al.*,[12] XGBoost achieves the highest accuracy compared with other methods.

Many variables in EHRs (e.g. body mass index, viruses, and chemicals) vary with time. Although mean values of EHRs over time could be computed to train traditional machine learning models for detection/prediction, such an approach has not taken full advantages of temporal characteristics for probable better performance. Results in a recent study[13] show that recurrent neural networks (RNNs) for sequential data processing achieve higher accuracy for detecting NASH from NAFLD patients than traditional machine learning models including logistic regression, random forest, and XGBoost. RNNs model temporal dependence over time by using not only the current observation but also the previous state for the current state updating as shown in Figure 1b. As one of the most representative RNN methods, long short-term memory consists of input, output, and forget gates to control the extent on how much short-term and long-term information should be memorized and forgotten, respectively. As a result, the RNN approach (e.g. long short-term memory) is usually a better choice for the analysis of time series EHR data. The advantages and disadvantages of these AI approaches are shown in Table 2.

Artificial intelligence in image interpretation. For medical image data, for example, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography, and histology, deep neural networks have become a promising approach to greatly improve the classification, detection, and segmentation accuracy.[14] Different from traditional machine learning methods constructing classifiers on human-designed features, convolutional neural networks (CNNs) as the most popular technique of deep neural networks for image processing perform end-to-end learning to automatically extract features by learning multiple convolutional filters and train classifiers at the same time. CNNs are composed of convolutional layers (with sparse connectivity and parameter sharing), pooling layers, and fully connected layers. Besides the last few fully connected layers, CNNs stack multiple convolutional blocks consisting of convolutional layers and pooling layers alternatively for feature learning from low to high levels. The idea of CNNs is illustrated in Figure 1c. By using CT images, CNNs have been employed

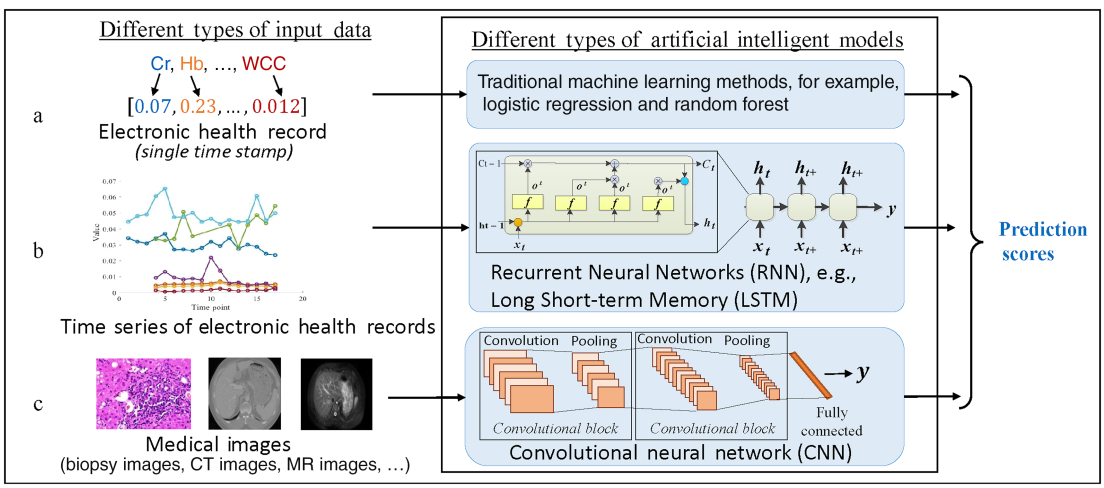

Figure 1 Review of artificial intelligence (AI) methods for non-alcoholic fatty liver disease and fibrosis prediction. (a) AI methods in electronic health record of single time stamp and (b) time series (—o—, Cr; —o—, Hb; —o—, Hct; —o—, PLT; —o—, PT; —o—, urea; —o—, WCC). (c) AI methods in medical images. CT, computed tomography; MR, magnetic resonance. [Color figure can be viewed at wileyonlinelibrary.com]

Table 2 Advantages and disadvantages of various AI approaches

AI approach	Advantages	Disadvantages
Logistic regression	• Easy to implement, interpret, and very efficient to train. • No assumptions about distributions of classes in feature space. • Easily extend to multiple classes (multinomial regression) and a natural probabilistic view of class predictions. • Provides both the appropriateness of a predictor (coefficient size) and its direction of association (positive or negative). • Very fast at classifying unknown records. • Good accuracy for many simple datasets, and it performs well when the dataset is linearly separable. • Less inclined to overfitting, but it can overfit in high-dimensional datasets.	• If the number of observations is lesser than the number of features, logistic regression should not be used; otherwise, it may lead to overfitting. • Constructs linear boundaries. • Assumption of linearity between the dependent variable and the independent variables. • Can only be used to predict discrete functions. Hence, the dependent variable of logistic regression is bound to the discrete number set. • Nonlinear problems cannot be solved with logistic regression because it has a linear decision surface. • Requires average or no multicollinearity between independent variables. • Hard to obtain complex relationships using logistic regression. More powerful and compact algorithms such as neural networks can easily outperform this algorithm.
Decision tree	• Requires less effort for data preparation during preprocessing. • Does not require normalization of data. • Does not require scaling of data as well. • Missing values in the data also do not affect the process of building a decision tree to any considerable extent. • Intuitive and easy to explain to technical teams as well as stakeholders.	• A small change in the data can cause a large change in the structure of the decision tree causing instability. • Calculation can go far more complex compared with other algorithms. • Often involves higher time to train the model. • Training is relatively expensive as the complexity and time have taken are more.
Random forest	• Reduces overfitting problem in decision trees and also reduces the variance and therefore improves the accuracy. • Can be used to solve both classification and regression problems. • Works well with both categorical and continuous variables. • Automatically handles missing values. • No feature scaling (standardization and normalization). • Handles nonlinear parameters efficiently. • Robust to outliers and can handle them automatically. • Very stable. Even if a new data point is introduced in the dataset, the overall algorithm is not affected much because the new data may impact one tree, but it is very hard for it to impact all the trees. • Less impacted by noise.	• Complexity: It creates many trees (unlike only one tree in case of decision tree) and combines their outputs. By default, it creates 100 trees in Python sklearn library. To do so, this algorithm requires much more computational power and resources. On the other hand, decision tree is simple and does not require so much computational resources. • Longer training period: It requires much more time to train as compared with decision trees as it generates many trees (instead of one tree in case of decision tree) and makes decision on the majority of votes.
XGBoost	• Easy to read and interpret algorithm. • Efficient prediction capability through the use of its clone methods. • A resilient method that curbs overfitting easily.	• Sensitive to outliers because every classifier is obliged to fix the errors in the predecessors. • Almost impossible to scale up, as every estimator bases its correctness on the previous predictors, thus making the procedure difficult to streamline.
Recurrent neural networks (RNNs)	• Can process inputs of any length. • Modeled to remember each information throughout the time, which is very helpful in any time series predictor. • Model size does not increase with input size. • Weights can be shared across the time steps. • Can use their internal memory for processing the arbitrary series of inputs, which is not the case with feedforward neural networks.	• Because of its recurrent nature, the computation is slow. • Training of RNN models can be difficult. • Difficult to process sequences that are very long. • Prone to problems such as exploding and gradient vanishing.
Convolutional neural networks (CNNs)	• Better performance for image-related vision tasks. • Automatic feature learning compared with the traditional manually designed feature engineering approach. • Reduced size of parameters by weight sharing compared with fully connected neural networks.	• Requires a large-scale dataset for training. • The computational complexity is higher than traditional machine learning methods. • There are more hyper-parameters to determine.

AI, artificial intelligence.

for fibrosis staging by segmenting the liver region and predicting the stage label.[15] With histology images, histological features of ballooning, inflammation, steatosis, and fibrosis are automatically scored by using CNNs.[16]

Artificial intelligence and non-alcoholic fatty liver disease

Most of the prediction models for NAFLD were derived and validated against liver biopsy or radiological tests using conventional statistical logistic regression analysis to identify factors associated with NAFLD and *c*-statistics to evaluate the overall accuracy (Table 2).[17] Optimal cutoff values are often defined according to those with high sensitivity, specificity, a maximum sum of sensitivity and specificity, and a maximum diagnostic accuracy (the sum of true positives and true negatives over the total number of patients), according to the diagnostic target.

Artificial intelligence may be applied in either histological or clinical scoring systems. AI assessment of histological features of NAFLD may reduce human variability and provide continuous rather than semiquantitative measurement of these features, with accuracy ranging from 82% to 89%.[18] AI software may identify histological features of NAFLD with high levels of interobserver and intraobserver agreement (0.95 to 0.99).[19]

Electronic health records can provide robust and comprehensive demographic and laboratory data in thousands and millions of patients, yet clinical observations and anthropometric measurements are often missing. A clinical decision support system has an accuracy of 91.7% to diagnose NAFLD.[20] NAFLD ridge score was developed with machine learning approach based only on five laboratory parameters and one comorbidity—alanine aminotransferase, high-density lipoprotein cholesterol, triglyceride, hemoglobin A1c, white blood cell count, and the presence of hypertension—using proton magnetic resonance spectroscopy (^{1}H-MRS) as the reference standard.[21] The score has an excellent negative predictive value of 96% to exclude NAFLD.

The histology-based AI models provide automated and objective diagnosis of NAFLD in selected patients who have undergone liver biopsy, whereas EMR-based AI models suit the need of large-scale epidemiological screening and risk stratification. We should be cautious in applying these AI models, as the reference standard may not be 100% correct (as in the case of histological features). Also, the general population of whom these AI models to be applied may have very different risk profile and clinical characteristics, in particular alcohol consumption and metabolic profile, when compared with the derivation cohorts. These models are mostly derived from cross-sectional cohorts at a particular time point, while in real life, all the factors included in the models are highly dynamic.

Non-alcoholic fatty liver disease used to be a diagnosis of exclusion. Recently, a group of expert hepatologists proposed to rename the condition as metabolic dysfunction-associated fatty liver disease (MAFLD) and require the inclusion of metabolic risk factors in its diagnosis.[22] A handful of studies discussed the superiority of MAFLD over NAFLD in terms of diagnosis and prognosis.[23–26] In a population study from our group, we found that the prevalence of MAFLD and NAFLD was similar despite the slightly different definitions, whereas the new MAFLD definition may reduce the incidence of new-onset disease from 13.8% to 10.4% in 3–5years.[27] Importantly, people who have fatty liver but no or mild metabolic risk factors are unlikely to have severe liver disease. It is thus reasonable not to classify them as having a disease. How this new definition of MAFLD affecting the performance of AI models remains unclear. Future studies may compare the diagnostic performance of AI models according to the definitions and diagnostic criteria of NAFLD and MAFLD.

Artificial intelligence and fibrosis

Assessing the severity of NAFLD is not only prognostically important but also a mandatory step to select patients for novel pharmacological treatments for NASH in clinical practice and research studies (Table 3).[9] AI models for liver fibrosis may be built on histology, clinical and laboratory parameters, and imaging. A handful of laboratory parameters, ranging from common parameters (such as aminotransferases) to specialized fibrosis biomarkers (e.g. hyaluronic acid and Pro-C3), are found to have modest to moderate accuracy to detect advanced fibrosis.[28]

An integrated AI-based automated tool is able to detect and quantify liver fibrosis through collagen proportionate area and assess its architectural pattern in liver biopsy with accuracy above 90% for bridging fibrosis and histological cirrhosis.[29] A refined AI-based algorithm (qFibrosis) incorporated 26 novel histological periportal parameters that showed a good discriminatory ability for earlier stages of liver fibrosis in NAFLD (F1 *vs* F2).[30] This would be particularly useful for screening patients for NAFLD clinical trials, which specifically target patients with F2 fibrosis.

There are a number of platforms to evaluate liver fibrosis and other histological features in patients with NAFLD. The dual-photon microscopy system looks at unstained slides and identifies various features of fibrosis including the length, thickness, and morphology of different fibrils (Fig. 2). In a study of 344 patients with biopsy-proven NAFLD, the system had 88–96% sensitivity and 78–91% specificity for diagnosing different fibrosis stages, using histological scoring as the reference standard.[31] Importantly, these AI-detected fibrosis patterns predicted cirrhotic complications and hepatocellular carcinoma during follow-up. Recently, the system has been expanded to quantify other histological features including steatosis, lobular inflammation, and ballooning (coined qFIBS), although the discrimination for severe inflammation and ballooning requires further refinement.[32]

Other investigators and companies have relied instead on digitized images of stained biopsy slides.[20,29,30,33] While most systems provide a consensus score based on the most likely fibrosis stage, the PathAI system describes the proportion of areas with features most resembling different particular fibrosis stages. In a combined analysis of a phase 2b and a phase 3 NASH trial, the change in AI-based features between baseline and follow-up liver biopsies correlated better with changes in the other noninvasive tests than traditional histological scoring, suggesting that AI may be more sensitive and specific as a method to evaluate treatment response.[33]

One notable difficulty in developing AI-based histology assessment is the lack of a real gold standard. Consensus scoring by more than one pathologist can reduce but not eliminate interobserver variability. In the dual-photon microscopy example, investigators adopted the "fair umpire" method using transient elastography as an additional reference.[31] Using this concept, the

Table 3 Studies on the use of AI to evaluate NAFLD

Author (year)	Country/region	Type of parameters	Parameters in the model	Gold standard	*N*	Accuracy
Prediction of NAFLD by clinical features						
Douali *et al.* (2013)	France	Laboratory/clinical	(Details not reported)	Liver biopsy	162	92%
Yip *et al.* (2017)	Hong Kong	Laboratory/clinical	• Alanine aminotransferase • High-density lipoprotein cholesterol • Triglyceride • Hemoglobin A1c • White blood cell count • Presence of hypertension	^{1}H-MRS	922	87–88%
Prediction of liver fibrosis by clinical features						
Sowa *et al.* (2013)	Germany	Laboratory/clinical	• Alanine aminotransferase • Aspartate aminotransferase • Hyaluronic acid • Apoptosis marker M30 • Apoptosis marker M65	Liver biopsy	126	79%
Schawkat *et al.* (2020)	Switzerland/USA	MRI	• T1-weighted fat-saturated images • T2-weighted fat-saturated images with MR elastography	Liver biopsy	62 (16/25.8% NAFLD)	82%
AI-based histology assessment						
Vanderbeck *et al.* (2013)	USA	Histology	• Morphological features • Texture • GLCM and its statistics • N-jet • Nuclei density	Liver biopsy	47	82–89%
Forlano *et al.* (2019)	UK	Histology	Automated quantification (in %) • Fat • Inflammation • Ballooning • Collagen proportional area (CPA) for fibrosis	Liver biopsy	246	92–97%
Gawrieh *et al.* (2020)	USA	Histology	Fibrosis quantification • CPA Fibrosis pattern • Morphological structures • Textural properties	Liver biopsy	987 annotations	Periportal fibrosis 78.6% Pericellular fibrosis 83.3% Portal fibrosis 86.4% Bridging fibrosis/cirrhosis > 90%
Leow *et al.* (2020)	China/Singapore	Histology	qFibrosis algorithm by dual-photon microscopy • 28 newly created periportal parameters • 28 altered perisinusoidal parameters	Liver biopsy	160	95–99% (F1 *vs* F2)
Wang *et al.* (2020)	China	Histology	• 25 fibrosis-related parameters (q-FPs) by dual-photon microscopy	Liver biopsy	344 patients (428 liver biopsies)	87.6–96.5% (F0–F1 *vs* F2–F4) 89.7–96.5% (F0–F2 *vs* F3–F4) 88.3–96.6% (F0–F3 *vs* F4)
Taylor-Weiner *et al.* (2020)	America/Asia/Europe	Histology	PathAI • Digitized images of liver biopsy • Assigns a fibrosis stage to each small areas • Percentage tissue area of predicted fibrosis stage	Liver biopsy	605 patients	Not reported in the abstract

^{1}H-MRS, proton magnetic resonance spectroscopy; AI, artificial intelligence; GLCM, gray-level co-occurrence matrix; MR, magnetic resonance; MRI, magnetic resonance imaging; NAFLD, non-alcoholic fatty liver disease.

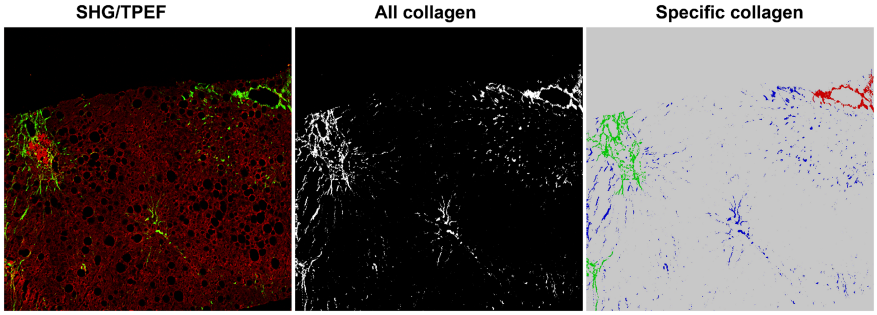

Figure 2 Examples of artificial intelligence-based image analysis of histological slides. The second-harmonic generation (SHG)/two-photon excitation fluorescence (TPEF) imaging-based tool examines unstained slides and identifies collagen fibrils (shown in green in the left panel and white in the middle panel). The system quantifies various features of liver fibrosis (right panel) for the determination of overall fibrosis stages (courtesy of Dr Dean Tai, HistoIndex Pte Ltd). ▬, central vein collagen; ▬, portal tract collagen; ▬, perisinusoidal collagen. [Color figure can be viewed at wileyonlinelibrary.com]

reference test does not have to be perfect; it only needs to have reasonable discriminating power for the outcome of interest and is not mechanistically related to the tests under evaluation. In this situation, the reference test should have the strongest correlation with the best diagnostic test. Nevertheless, these AI-based algorithms would eventually need to demonstrate an association with clinical outcomes, as is currently required for other surrogate endpoints in clinical trials.

Combination of imaging modalities and AI improves the accuracy in staging liver fibrosis. The performance of combining machine learning with CT and MRI has been evaluated.[20] Their predictions are found to correlate with histological-evaluated liver fibrosis stages. Machine learning-based radiomics also play a role in image analysis. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis.[34] Besides that, imaging-based texture analysis-derived parameter, which is a tool of radiomics, combined with machine learning of non-contrast-enhanced T1-weighted magnetic resonance images could be as accurate (82%) as magnetic resonance elastography for liver fibrosis quantification.[35] These approaches may act as alterative for staging liver fibrosis in different resource settings.

The histology-based or imaging-based AI models provide automated and objective detection of significant (F2) to advanced fibrosis (F3–F4) of NAFLD in selected patients who have undergone liver biopsy or MRI. Laboratory-based AI models have higher applicability as some common laboratory parameters are part of the routine liver panel. Here, applicability has to be balanced with accuracy. As histological fibrosis staging remains the gold standard, it is subjected to sampling error. Such validated studies are often skewed towards patients with more advanced disease so that the accuracies may differ in the general population.

Artificial intelligence and non-alcoholic fatty liver disease-related hepatocellular carcinoma

There has been an increasing trend of NAFLD-related HCC.[8] Novel data mining approaches were used to develop prognostic algorithm for these patients.[36] Random forest analysis identified treatment for HCC and serum albumin level as the first and second prognostic factors. A decision tree algorithm revealed that patients who underwent curative treatments, namely, surgical resection and radiofrequency ablation, together with a higher serum albumin level had the best prognosis. The beauty of these data mining approaches is common clinical parameters and well-defined decision process are involved, so that it can be easily applied in clinical settings.

Conclusions and the future

Among the technologies discussed earlier, AI-based image analysis is the closest to clinical application and may transform clinical trials in NASH. Apart from using AI to suggest histological scores like what pathologists are currently doing, quantification of the AI features may provide more granular data to determine treatment effects. If the latter is to be used as surrogate endpoints, studies should demonstrate an association between these features and clinical outcomes. Nevertheless, such an approach still requires liver biopsies. The future should be noninvasive tests and algorithms, which may also be AI-based. At present, the use of AI to predict NAFLD, fibrosis, and clinical outcomes is still in its infancy. Most studies only used cross-sectional data. In real life, clinicians do not see patients once but repeatedly. AI is in fact well positioned to handle this type of dynamic data, and this should be explored further. With concerted effort, these new developments in AI would hopefully improve clinical care and patient experience and outcomes.

References

1 Younossi Z, Tacke F, Arrese M *et al*. Global perspectives on nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Hepatology* 2019; **69**: 2672–82.
2 Li J, Zou B, Yeo YH *et al*. Prevalence, incidence, and outcome of non-alcoholic fatty liver disease in Asia, 1999–2019: a systematic review and meta-analysis. *Lancet Gastroenterol. Hepatol.* 2019; **4**: 389–98.
3 Younossi ZM, Stepanova M, Ong J *et al*. Nonalcoholic steatohepatitis is the most rapidly increasing indication for liver transplantation in the United States. *Clin. Gastroenterol. Hepatol.* 2020. https://doi.org/10.1016/j.cgh.2020.05.064. Online ahead of print.
4 Brunt EM, Wong VW, Nobili V *et al*. Nonalcoholic fatty liver disease. *Nat. Rev. Dis. Primers* 2015; **1**: 15080.
5 Taylor RS, Taylor RJ, Bayliss S *et al*. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Gastroenterology* 2020; **158**: 1611–25, e12.

6 Vilar-Gomez E, Calzadilla-Bertot L, Wong VW *et al*. Fibrosis severity as a determinant of cause-specific mortality in patients with advanced nonalcoholic fatty liver disease: a multi-national cohort study. *Gastroenterology* 2018; **155**: 443–57, e17.
7 Wong VW, Chitturi S, Wong GL *et al*. Pathogenesis and novel treatment options for non-alcoholic steatohepatitis. *Lancet Gastroenterol. Hepatol.* 2016; **1**: 56–67.
8 Wong VW, Chan WK, Chitturi S *et al*. Asia–Pacific Working Party on Non-alcoholic Fatty Liver Disease guidelines 2017—part 1: definition, risk factors and assessment. *J. Gastroenterol. Hepatol.* 2018; **33**: 70–85.
9 Wong VW, Adams LA, de Ledinghen V *et al*. Noninvasive biomarkers in NAFLD and NASH—current progress and future promise. *Nat. Rev. Gastroenterol. Hepatol.* 2018; **15**: 461–78.
10 Davison BA, Harrison SA, Cotter G *et al*. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J. Hepatol.* 2020; **73**: 1322–32.
11 Spann A, Yasodhara A, Kang J *et al*. Applying machine learning in liver disease and transplantation: a comprehensive review. *Hepatology* 2020; **71**: 1093–105.
12 Fialoke S, Malarstig A, Miller MR *et al*. Application of machine learning methods to predict non-alcoholic steatohepatitis (NASH) in non-alcoholic fatty liver (NAFL) patients. *AMIA Annu. Symp. Proc.* 2018; **2018**: 430–9.
13 Suresha PB, Wang Y, Xiao C *et al*. A deep learning approach for classifying nonalcoholic steatohepatitis patients from nonalcoholic fatty liver disease patients using electronic medical records. *In Explainable AI in Healthc. Med.* 2020: 107–13.
14 Zhou LQ, Wang JY, Yu SY *et al*. Artificial intelligence in medical imaging of the liver. *World J. Gastroenterol.* 2019; **25**: 672–82.
15 Choi KJ, Jang JK, Lee SS *et al*. Development and validation of a deep learning system for staging liver fibrosis by using contrast agent-enhanced CT images in the liver. *Radiology* 2018; **289**: 688–97.
16 Heinemann F, Birk G, Stierstorfer B. Deep learning enables pathologist-like scoring of NASH models. *Sci. Rep.* 2019; **9**: 18454.
17 Wong VW, Vergniol J, Wong GL *et al*. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology* 2010; **51**: 454–62.
18 Vanderbeck S, Bockhorst J, Komorowski R *et al*. Automatic classification of white regions in liver biopsies by supervised machine learning. *Hum. Pathol.* 2014; **45**: 785–92.
19 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**: 2081–90, e9.
20 Douali N, Abdennour M, Sasso M *et al*. Noninvasive diagnosis of nonalcoholic steatohepatitis disease based on clinical decision support system. *Stud. Health Technol. Inform.* 2013; **192**: 1178.
21 Yip TC, Ma AJ, Wong VW *et al*. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment. Pharmacol. Ther.* 2017; **46**: 447–56.
22 Eslam M, Newsome PN, Sarin SK *et al*. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. *J. Hepatol.* 2020; **73**: 202–9.
23 Lin S, Huang J, Wang M *et al*. Comparison of MAFLD and NAFLD diagnostic criteria in real world. *Liver Int.* 2020; **40**: 2082–9.
24 Yamamura S, Eslam M, Kawaguchi T *et al*. MAFLD identifies patients with significant hepatic fibrosis better than NAFLD. *Liver Int.* 2020. https://doi.org/10.1111/liv.14675. Online ahead of print.
25 Vilar-Gomez E, Calzadilla-Bertot L, Wong VW *et al*. Type 2 diabetes and metformin use associate with outcomes of patients with nonalcoholic steatohepatitis-related, Child–Pugh A cirrhosis. *Clin. Gastroenterol. Hepatol.* 2021; **19**: 136–45.e6.
26 Sun DQ, Jin Y, Wang TY *et al*. MAFLD and risk of CKD. *Metabolism* 2020; **115**: 154433.
27 Wong VW, Wong GL, Woo J *et al*. Impact of the new definition of metabolic associated fatty liver disease on the epidemiology of the disease. *Clin. Gastroenterol. Hepatol.* 2020. https://doi.org/10.1016/j.cgh.2020.10.046. Online ahead of print.
28 Sowa JP, Heider D, Bechmann LP *et al*. Novel algorithm for non-invasive assessment of fibrosis in NAFLD. *PLoS ONE* 2013; **8**: e62439.
29 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.
30 Leow WQ, Bedossa P, Liu F *et al*. An improved qFibrosis algorithm for precise screening and enrollment into non-alcoholic steatohepatitis (NASH) clinical trials. *Diagnostics (Basel)* 2020; **10**: 643.
31 Wang Y, Wong GL, He FP *et al*. Quantifying and monitoring fibrosis in non-alcoholic fatty liver disease using dual-photon microscopy. *Gut.* 2020; **69**: 1116–26.
32 Liu F, Goh GB, Tiniakos D *et al*. qFIBS: an automated technique for quantitative evaluation of fibrosis, inflammation, ballooning, and steatosis in patients with nonalcoholic steatohepatitis. *Hepatology* 2020; **71**: 1953–66.
33 Taylor-Weiner A, Pokkalla H, Han L *et al*. Validation of a machine learning-based approach (DELTA liver fibrosis score) for the assessment of histologic response in patients with advanced fibrosis due to NASH. *Hepatology* 2020; **72**: 949–51.
34 Wang K, Lu X, Zhou H *et al*. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut.* 2019; **68**: 729–41.
35 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**: 4675–85.
36 Kawaguchi T, Tokushige K, Hyogo H *et al*. A data mining-based prognostic algorithm for NAFLD-related hepatoma patients: a nationwide study by the Japan Study Group of NAFLD. *Sci. Rep.* 2018; **8**: 10434.