74646_Auto_Edited.docx

Name of Journal: *Artificial Intelligence in Gastroenterology*

Manuscript NO: 74646

Manuscript Type: MINIREVIEWS

Application of artificial intelligence in non-alcoholic fatty liver disease and viral hepatitis

Atchayaa Gunasekharan, Joanna Jiang, Ashley Nickerson, Sajid Jalil, Khalid Mumtaz

Abstract

Non-alcoholic fatty liver disease (NAFLD) and chronic viral hepatitis are among the most significant causes of liver-related mortality worldwide. It is critical to develop reliable methods of predicting progression to fibrosis, cirrhosis, and decompensated liver disease. Current screening methods such as biopsy and transient elastography are limited by invasiveness and observer variation in analysis of data. Artificial intelligence (AI) provides a unique opportunity to more accurately diagnose NAFLD and viral hepatitis, and to identify patients at high risk for disease progression. We conducted a literature review of existing evidence for AI in NAFLD and viral hepatitis. Thirteen articles on AI in NAFLD and 14 on viral hepatitis were included in our analysis. We found that machine learning (ML) algorithms were comparable in accuracy to current methods for diagnosis and fibrosis prediction (MELD-Na score, liver biopsy, FIB-4 score, and biomarkers). They also reliably predicted hepatitis C treatment failure and hepatic encephalopathy, for which there are currently no established prediction tools. These studies show that AI could be a helpful adjunct to existing techniques for diagnosing, monitoring, and treating both NAFLD and viral hepatitis.

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) exists on a spectrum from simple hepatocyte steatosis to inflammation, ballooning and fibrosis. Given the lack of efficient screening methods and high rate of asymptomatic disease, it is challenging to identify patients with various stages of NAFLD.(1,9) Non-alcoholic steatohepatitis (NASH) patients with significant fibrosis are at increased risk for cirrhosis and progressive liver failure, which has led NASH to become one of the leading causes of liver transplantation in the United States.(24) **NASH affects approximately 3% to 6% of the US population, and this number continues to increase. It affects approximately 25% of the population worldwide.** (20)

Although liver biopsy remains the gold standard for diagnosing NASH, it is an invasive, costly, and painful procedure. Therefore, serial liver biopsies for surveillance are not always feasible. Conventional imaging modalities including ultrasound, CT, MRI and transient elastography are limited by inter- and intra-observer variability depending on the stage of fibrosis.(1,9) Similarly, despite recent progress in the prevention and treatment of viral hepatitis, predicting sustained virological response (SVR) and disease progression remains challenging.

Artificial intelligence (AI) is an exciting and increasingly pertinent field in medicine as clinicians incorporate augmenting technology into their daily practice. **AI is the concept of teaching a computer to simulate the cognitive abilities of the human brain. Machine learning (ML) entails allowing the computer to simulate the human brain independently. It can either be supervised (through specific feedback from humans) or unsupervised, in which case there is no guidance provided and the computer is able to independently synthesize and analyze the output.(1)** AI is increasingly applied to the diagnosis and prediction of various diseases. Researchers are developing machine learning (ML) algorithms to predict risk and outcomes using multiple demographic, clinical, biochemical, and imaging–parameters for diagnosis and prognosis related to liver fibrosis and steatosis, including NAFLD and viral hepatitis. (1)

Current methods of assessing liver fibrosis progression and mortality in both NAFLD and viral hepatitis have many limitations. **These include the intra- and inter-observer**

variability in staging fibrosis, the inability to place fibrosis along a continuum, and the lack of identifiable markers for disease progression. (1,9) These limitations and the ability of ML models to overcome them will be discussed further in this review. This review will also highlight how ML models have the potential to present opportunities for drug discovery and prediction of therapeutic and toxic effects of drugs. Machine learning models based on AI provide promising features that could not only enhance screening for NAFLD, but also help with fibrosis staging in patients with NASH and viral hepatitis.

This review summarizes recent literature on the application of AI in NAFLD and viral hepatitis. The main objective is to assess the performance of AI as a non-invasive method for the diagnosis and staging of liver fibrosis and steatosis, as well as the detection and treatment of chronic viral hepatitis.

METHODS

A review of current literature in the areas of AI in NAFLD and viral hepatitis was conducted using two separate searches on PubMed. First, we used the search terms "non-alcoholic fatty liver disease", "NAFLD", and "deep learning" in combination with "artificial intelligence", "histology", "omics" and "radiology." The second search was conducted using the search terms "viral hepatitis" in combination with "hepatitis A", "hepatitis B", "hepatitis C", "hepatitis E", "machine learning", "artificial intelligence", "histology" and "radiology".

Most articles on NASH and NAFLD published between 2018 and 2021 were included in this review. **Articles were excluded if they did not offer comparisons between AI modalities and existing methods for screening or prediction (MELD score, elastography, *etc.*). Twenty-seven articles were included in our review, 13 on NAFLD and 14 on chronic viral hepatitis.** For studies on viral hepatitis, described machine learning algorithms fell into one of three categories: predicting prevalence, screening for complications (including fibrosis, HCC, decompensated cirrhosis, and death), and predicting response to treatment.

USE OF AI FOR DIAGNOSING VIRAL HEPATITIS AND NAFLD/NASH

It is estimated that half of patients infected with hepatitis C worldwide are unaware of their diagnosis and only 17% have undergone liver fibrosis staging. [31] This rate is even lower for hepatitis B, for which only 10.5% of infected patients are aware of their status. In March 2020, the USPSTF recommended hepatitis C screening for all adults over 18; however, there are currently no population-based screening recommendations for hepatitis A and B. Primary care offices do not routinely test for hepatitis B. Machine learning has been used both to determine regional prevalence of chronic hepatitis and to identify undiagnosed cases.

Zheng, et al compared two algorithms (Elman neural network and autoregressive integrated moving average, or ARIMA) designed to predict incidence of hepatitis B in Guangxi, China. ARIMA is a type of model that can capture the randomness of data and is often used for infectious disease prediction. Predictions were compared to the reported cases of hepatitis B cases from the Health Commission of Guangxi, China. The neural network was the more predictive model, with a root-mean-square error (RMSE) of 0.89 and mean absolute error (MAE) of 0.70, while the ARIMA had an RSME of 0.94 and an MAE of 0.81.[33]

A 2020 study by Doyle, *et al* aimed to predict HCV positive status by using patient claims data to develop four algorithms, all with a predictive accuracy of over 95%. Algorithms included logistic regression, gradient boosted trees, a stacked ensemble, and random forests. The stacked ensemble performed the best, with a precision of 97% at recall levels >50%. Key predictors of HCV infection included NSAID use, opioids, healthcare utilization, patient age and osteoarthritis or glomerulonephritis treatment. [7] We were unable to find any study to date using AI to screen for NAFLD/NASH.

USE OF AI TO ASSESS FIBROSIS IN VIRAL HEPATITIS AND NAFLD/NASH

Existing histologic models not only rely on scoring of fibrosis by a pathologist but are also unable to place fibrosis along a continuum. Artificial intelligence enables the placement of fibrosis along a continuum, identifies risk factors for progression of fibrosis, allows enhanced scoring of fibrosis stages, leading to better selection of patients for clinical trials This also allows for identification of therapeutic targets. [9]

Lu et al. developed a light gradient-boosting machine (GBM) model to predict liver fibrosis and cirrhosis in treatment-naive chronic hepatitis B patients at four centers in China. The model, named Fibro Box, outperformed transient elastography, APRI, and FIB-4, with AUC 0.88 in external validation sets for significant fibrosis and 0.87 for cirrhosis. Input variables included fibroscan results, platelets, ALT, Prothrombin time (PT), and splenic vein diameter. [21]

A 2013 study by *Zheng et al*, used an artificial neural network (ANN) to predict 3-month mortality of individuals with acute-on-chronic liver failure due to hepatitis B (HBV-ACLF). Patient characteristics included in this model were age, PT, serum sodium, total bilirubin, E antigen positivity status and hemoglobin. The ANN predicted mortality more accurately than MELD-based scoring systems, with area under the curve receiver operating characteristic (AUCROC) 0.765 in the validation cohort compared to 0.599 for MELD. [32]

Similarly, Huo *et al* developed ANNs to predict 28- and 90-day mortality in HBV-ACLF. Data were retrospectively reviewed from 684 patients admitted for ALF at 8 hospitals in various Chinese provinces with 423 cases in the training cohort and 261 in the validation cohort. In the training cohorts, the neural network had a significantly higher accuracy than MELD, MELD-Na, CLIF-ACLF, and Child-Pugh score, with AUC 0.948 and 0.913 for 28- and 90-day mortality, respectively. In the validation cohort, the model performed significantly better than MELD and insignificantly better than other scoring systems, with AUC 0.748 and 0.754 for 28- and 90-day mortality. Significant mortality predictors included age, presence of HE, sodium, PT, GGT, e antigen, ALP, and bilirubin. [16]

7

In another study, *Wang et al* used Deep learning Radiomics of Elastography (DLRE) to assess stages of liver fibrosis in patients with chronic hepatitis B. DLRE was compared to 2D shear wave elastography and biomarkers (AST: platelet ratio, fibrosis index), with liver biopsy as the reference standard. 1990 images from 398 patients were used to develop the models. AUCROCs for DLRE were 0.97 for cirrhosis, 0.98 for advanced fibrosis, and 0.85 for significant fibrosis; this performed better than other methods except for elastography in severe fibrosis. [28]

Like viral hepatitis, there are several studies establishing the role of AI in **assessing** fibrosis in NAFLD/NASH. In one study by *Forlano et al*, [9] liver biopsy specimens were annotated by two expert pathologists using the clinical research network (CRN) score as a measurable scale of degree of steatosis, inflammation, ballooning and fibrosis. The machine learning model was built using 100 patients with NAFLD in the derivation group and 146 patients in the validation group. There was good concordance when the machine learning model was compared to the scoring of the expert histopathologist on the liver biopsy **specimens**; the interclass correlation coefficients (ICC) were 0.97 (95%CI, 0.95–0.99; p-value < .001) for steatosis, 0.96 (95%CI, 0.9–0.98; p-value < .001) for inflammation, 0.94 (95%CI, 0.87–0.98; p-value < .001 for ballooning, and 0.92 for fibrosis (95%CI, 0.88–0.96; p-value <.001). A subgroup analysis showed that quantitative analysis performed better than the clinical research network (CRN) score in differentiating between the various stages of NAFLD. Another CNN model developed by *Qu et al,* showed that a computational neural network (CNN) model had an area under the curve (AUC) of 63% for all four subsets of the NAFLD scoring, while the AUC's were 90.48% for steatosis, 81.06% for ballooning, 70.18% for inflammation and 83.85% for fibrosis. These studies underscore the utility of ML models in illustrating the heterogeneity of liver pathology in NAFLD. [9,26]

In another study by *Taylor-Weiner et al*, a convolutional neural network (CNN) model was developed that allowed for assessment of fibrosis along a continuum, which is not possible with pathologist scoring alone. The CRN and Ishak scores were applied to each pixel within a given image, allowing for evaluation of heterogeneity in fibrosis as well. In addition, the CNN served as a prediction model allowing for identification of features associated with disease progression. The model's predictions correlated significantly with the pathologist scoring in all three studies, the STELLAR-3, STELLAR-4, and ATLAS, whose participants were used to build and validate the ML model - steatosis, ρ = 0.60; p-value < 0.001; lobular inflammation, ρ = 0.35; p-value < 0.001; and HB, ρ= 0.41; p-value < 0.001.[27] The model's level of agreement with pathologist scoring was within the range of agreement between individual pathologists. The weighted Cohen's kappa was 0.801 for NASH CRN and 0.817 for the Ishak classifications.

Another study by *Gawrieh et al* built a ML model using support vector machines (SVM) to better characterize architectural patterns in fibrosis. [10] This ML model was built to differentiate between six different patterns of fibrosis and had a strong correlation with the pathologist's semi-quantitative scores for fibrosis, with a coefficient of determination of automated CPA ranging between 0.60 to 0.86 when compared with the pathologist score. The model was built using a trichrome-stained liver biopsy specimen which was **marked** with 987 annotations for different fibrosis types. As noted in the study, the model's AUROCs were 78.6% for detection of periportal fibrosis, 83.3% for pericellular fibrosis, 86.4% for portal fibrosis, and >90% for detection of normal fibrosis, bridging fibrosis and presence of nodules/cirrhosis.

AI USING METABOLOMICS FOR NAFLD/NASH

There is an increasing number of studies focusing on **metabolomics** that allow for non-invasive identification of targets associated with development and progression of NAFLD. **These biomarkers may differentiate between patients with and without cirrhosis, and between a healthy liver and NAFLD or NASH**. [2,6,24] Several direct and indirect blood-based biomarkers currently exist to assess fibrosis. These have been incorporated to form scoring systems such as NAFLD fibrosis score (NFS), Fibrosis-4 (FIB-4), AST to Platelet Ratio Index (APRI), BARD Score, FibroSURE and Enhanced liver fibrosis score. [24] ML allows for analysis of many multi-omics and clinical variables to screen for NASH and NAFLD and to build models for disease progression.

An eXtreme Gradient Boosting Model (XG Boost) was developed using the NIDDK database by Docherty *et al*, which contains a large real-world patient population**. This**

model used confirmed NASH and non-NASH patients within this subset.[6] The unique feature of this study is that it used several demographic variables and clinical biomarkers run through recursive feature elimination (RFE) in combination with confirmed histologic cases to build an efficient model **with** a high specificity. When a greater number of markers were used **in predicting** patients with NASH, the AUROC was 0.82, sensitivity 81%, and precision 81%.

In a study of adults of European ancestry by *Atabaski-Pastar et al,* patients with type 2 diabetes and others with high-risk features for the development of NASH were assessed for liver fat content using MRI.[2] Several multi-omics and clinical data, including laboratory markers, were entered into the least absolute shrinkage and selection operator (LASSO) to select the most relevant features, which then underwent random forest analysis for the development of the algorithm. The model developed using this method produced a cross-validated AUROC of 0.84 (95%CI 0.82, 0.86; p-value < 0.001) and outperformed existing prediction tools for NAFLD. However, unlike other studies, the model was built in comparison to MRI fat content, which is not reflective of the continuum of NAFLD, and thus cannot be used to monitor disease progression.

Another study based in China by *Ma et al* identified BMI, triglycerides, gamma-glutamyl transpeptidase (GGT), the serum alanine aminotransferase (ALT) and uric acid as the most common features contributing to NAFLD when a Bayesian network model was used.[22] The model had an accuracy of 83%, specificity of 0.878, sensitivity of 0.675, and F-measure score of 0.655. The F-measure score is an indicator of whether there can be a balance between precision and recall of these variables, and it was higher than for logistic regression models in machine learning.

AI IN IMAGE INTERPRETATION FOR NAFLD/NASH

Like markers discussed previously, **many studies have combined machine learning with imaging modalities to more effectively assess liver fat content and to better define fibrosis scores**. This would allow for more accurate monitoring of patients for disease progression and their selection for clinical trials.

Current modalities for estimation of liver fat content include conventional ultrasound (US), which is limited by variable accuracy, operator dependency, and its qualitative nature. The measurement of proton density fat fraction (PDFF) by MRI is proving to be an effective method for quantification of hepatic steatosis, but it is expensive and there is variability in results due to dependence on calibration. In a study by *Han et al,* one-dimensional convolutional neural networks (CNN) was applied to ultrasound radiofrequency signals for the diagnosis of NAFLD and quantitation of hepatic fat content with an AUC of 0.98 (95%CI: 0.94, 1.00). (13) In diagnosing NAFLD, the model had an accuracy of 96%, sensitivity of 97%, and specificity of 94%, PPV of 97% and NPV of 94%. The ML model also correlated with MRI-PDFF with a Pearson correlation coefficient of 0.85 (p-value < .001). The same method was applied to animal models in a study by *Nguyen et al* and it showed that CNN outperformed quantitative ultrasound in differentiating between NAFLD and normal liver.(23) Further support for ML comes from a recent study by Das *et al*. on pediatric patients which used an ensemble model comprising SVM, Neural Net and XG Boost that had an AUC of 0.92 (95%CI, 0.91–0.94) when tested in an external validation cohort. (5)

Nonenhanced CT also remains superior to histopathologic quantification of liver fat content like MRI-PDFF, but it is also more commonly performed in clinical practice for other reasons when compared to MRI. It currently uses a manual region-of-interest (ROI) for estimation of liver fat content. A study by *Graffy et al* developed a deep-learning based automated liver segmentation tool and applied it to estimate liver fat content using three-dimensional CNN, without having to depend on manual ROI. The pearson correlation coefficient was 0.93.(11) This allows for large population level estimation of liver fat content to determine the prevalence of NAFLD. It would also determine normal liver fat content based on a large sample. Used in combination with other non-invasive modalities such as serum biomarkers, it could help identify patients who will need closer monitoring for NAFLD progression to cirrhosis. In a similar study by *Hou et al*, the automated liver attenuation ROI-based measurement (ALARM) model had a pearson coefficient of 0.94 when compared with manual ROI. (14)

In addition to differentiating healthy liver from NAFLD, ML models have also been used to reduce variability in detecting fibrosis, specifically F2 fibrosis, which is a limiting feature of shear wave elastography. A study by Brattain *et al* combined the use of shear wave elastography with CNN to better assess F2 fibrosis. (3) This approach not only assessed image quality, but also selected ROI, unlike the previous studies. This ML model detected F2 fibrosis with AUC of 0.89 compared to AUC of 0.74 when image quality and ROI were not incorporated into a ML model. This demonstrates the importance of ML models once again in selecting patients for clinical trials, and in assessing response to treatment.

AI IN VIRAL HEPATITIS TREATMENT

The rate of sustained viral response (SVR) for hepatitis C with modern direct acting antiviral (DAA) regimens is estimated to be over 90%; however, variability remains in treatment length and efficacy. Patients with prior DAA exposure, cirrhosis, and other risk factors may require a longer treatment course. (12, 31) Machine learning has been applied to predicting treatment response and duration based on patient-specific factors. *Haga et al* applied nine machine learning algorithms to identify the optimized combination of HCV genotypic variants that predict SVR after DAA therapy. HCV genomes were sequenced from the serum of 173 patients (including 64 without SVR). The support vector machine algorithm was found to be the most predictive, with a validation accuracy of 0.95.(12) *Feldman et al* used data from 60 million beneficiaries of a managed care plan (including 3943 cases of hepatitis C who received sofosbuvir/ ledipasvir), to identify demographic and medical factors that may predict a prolonged course of DAA. Machine learning algorithms included extreme gradient boosting (XG Boost), random forest and support vector machine, with XG Boost being the optimal predictive model at an AUC of 0.745. Patient age, comorbidity burden, and type 2 diabetes status were significant predictors.(8) *Wei et al* developed an ANN and logistic regression model to predict fibrosis reversal after 78 wk of hepatitis B treatment. Significant predictors included AST and ALT, platelets, WBC, gender, and Fibroscan results. The ANN outperformed the logistic regression model, with an AUC of 0.81 *vs* 0.75.(30)

The only approved treatment for NAFLD is weight reduction. We were unable to find AI based algorithms and predictive models for NAFLD due to lack of pharmacologic management options.

DISCUSSION

Among the algorithms described, more complex models performed better, with machine learning consistently outperforming more basic logistic regression models. The highest-performing models incorporated both demographic and radiologic/serologic variables. AI models also predicted complications more accurately than biomarkers and scoring systems like MELD and FIB-4. **These models could be used to predict the incidence and prevalence of viral hepatitis in regions without robust, widespread screening programs. Additionally, they** could be helpful in the initiation of treatment and predicting response to antivirals **for individual patients**, for which no gold standard currently exists.

Limitations of the current AI models are notably due to the lack of large scale, randomized controlled trials. Further research is necessary to demonstrate the utility of AI. With further advancements, ML models could potentially be incorporated into all aspects of a patient's care, from screening the general population for NAFLD or NASH, to monitoring disease progression and treatment response in clinical trials by enhancing classification of steatosis, ballooning, inflammation, and fibrosis. **In this regard, more population-based studies are needed to study the applications of ML models in screening. Additionally, large scale, randomized controlled trials are needed to study serologic and histologic markers for disease progression.** Further studies are also warranted to explore the potential of ML algorithms to provide target-specific medications, yielding efficacious pharmacotherapy in a disease such as NASH where good treatment options are lacking at this time. **Though AI is promising in terms of its**

potential to develop therapeutic targets, we were unable to find any studies to date describing the use of AI in drug discovery.

Future directions also include using AI to actively improve outcomes with viral hepatitis by increasing adherence to DAAs or identifying individuals at risk for contracting viral hepatitis. Machine learning models could also help identify barriers to accessing treatment.

CONCLUSION

Machine learning models focus on various aspects of liver disease, including demographics, biochemical labs, histologic assessment and patterns, identification of non-invasive biomarkers, and liver imaging techniques (Figure 1). Overall, the studies outlined above are promising in their reliance on non-invasive methods as opposed to conventional liver biopsy to study the stages of fibrosis, **as well as their ability to place fibrosis along a continuum and identify markers for disease progression. This could reduce healthcare costs by allowing better selection of patients in whom a liver biopsy is performed. It would also benefit patients by decreasing the number of them who undergo this invasive procedure.** AI can also improve efficiency of pathologist and sonographer scoring of samples when added to existing methods. **This will allow for a better understanding of the pathophysiology of diseases like NAFLD, which would not only allow for appropriate screening for disease progression, but also improve the ability to develop therapeutic targets.**

74646_Auto_Edited.docx

ORIGINALITY REPORT

4%

SIMILARITY INDEX

PRIMARY SOURCES

1 "Posters (Abstracts 264-2239)", Hepatology, 2017
Crossref
40 words — 1%

2 Pankaj Aggarwal, Naim Alkhouri. "Artificial Intelligence in Nonalcoholic Fatty Liver Disease: A New Frontier in Diagnosis and Treatment", Clinical Liver Disease, 2021
Crossref
37 words — 1%

3 Samer Gawrieh, Deepak Sethunath, Oscar W. Cummings, David E. Kleiner, Raj Vuppalanchi, Naga Chalasani, Mihran Tuceryan. "Automated quantification and architectural pattern detection of hepatic fibrosis in NAFLD", Annals of Diagnostic Pathology, 2020
Crossref
31 words — 1%

4 Amaro Taylor ‐ Weiner, Harsha Pokkalla, Ling Han, Catherine Jia et al. "A Machine Learning Approach Enables Quantitative Measurement of Liver Histology and Disease Monitoring in NASH", Hepatology, 2021
Crossref
17 words — < 1%

5 www.ncbi.nlm.nih.gov
Internet
16 words — < 1%

6 Theodore C Feldman, Jules L. Dienstag, Kenneth D Mandl, Yi-Ju Tseng. "Machine-Learning-Based Predictions of Direct Acting Antiviral Therapy Duration for
13 words — < 1%

Patients with Hepatitis C", International Journal of Medical Informatics, 2021
Crossref

7	Www.semanticscholar.org Internet	12 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		