### MINIREVIEWS

80  Machine learning approaches using blood biomarkers in non-alcoholic fatty liver diseases  
* Carteri RB, Grellert M, Borba DL, Marroni CA, Fernandes SA*

88  Artificial intelligence using advanced imaging techniques and cholangiocarcinoma: Recent advances and future direction  
* Brenner AR, Laoveeravat P, Carey PJ, Joiner D, Mardini SH, Jovani M*
ABOUT COVER
Editorial Board Member of Artificial Intelligence in Gastroenterology, Janaina Luz Narciso-Schiavon, PhD, Associate Professor, Department of Internal Medicine, Federal University of Santa Catarina, Florianopolis 88040-900, Santa Catarina, Brazil. janaina.narciso@uol.com.br

AIMS AND SCOPE
The primary aim of Artificial Intelligence in Gastroenterology (AIG, Artif Intell Gastroenterol) is to provide scholars and readers from various fields of artificial intelligence in gastroenterology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

AIG mainly publishes articles reporting research results obtained in the field of artificial intelligence in gastroenterology and covering a wide range of topics, including artificial intelligence in gastrointestinal cancer, liver cancer, pancreatic cancer, hepatitis B, hepatitis C, nonalcoholic fatty liver disease, inflammatory bowel disease, irritable bowel syndrome, and Helicobacter pylori infection.

INDEXING/ABSTRACTING
The AIG is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

RESPONSIBLE EDITORS FOR THIS ISSUE
Production Editor: Wen-Wen Li; Production Department Director: Xiang Lei; Editorial Office Director: Jin-Lei Wang.

NAME OF JOURNAL
Artificial Intelligence in Gastroenterology

ISSN
ISSN 2644-3236 (online)

LAUNCH DATE
July 28, 2020

FREQUENCY
Bimonthly

EDITORS-IN-CHIEF
Rajvinder Singh, Ferruccio Bonino

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3236/editorialboard.htm

PUBLICATION DATE
June 28, 2022

COPYRIGHT
© 2022 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS
https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS
https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH
https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS
https://www.wjgnet.com/bpg/gerinfo/288

PUBLICATION MISCONDUCT
https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE
https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS
https://www.wjgnet.com/bpg/gerinfo/239

ONLINE SUBMISSION
https://www.filipublishing.com

© 2022 Baishideng Publishing Group Inc. All rights reserved. 7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
E-mail: bpgoffice@wjgnet.com https://www.wjgnet.com
Machine learning approaches using blood biomarkers in non-alcoholic fatty liver diseases

Randhall B Carteri, Mateus Grellert, Daniela Luisa Borba, Claudio Augusto Marroni, Sabrina Alves Fernandes

Abstract

The prevalence of nonalcoholic fatty liver disease (NAFLD) is an important public health concern. Early diagnosis of NAFLD and potential progression to nonalcoholic steatohepatitis (NASH), could reduce the further advance of the disease, and improve patient outcomes. Aiming to support patient diagnostic and predict specific outcomes, the interest in artificial intelligence (AI) methods in hepatology has dramatically increased, especially with the application of less-invasive biomarkers. In this review, our objective was twofold: Firstly, we presented the most frequent blood biomarkers in NAFLD and NASH and secondly, we reviewed recent literature regarding the use of machine learning (ML) methods to predict NAFLD and NASH in large cohorts. Strikingly, these studies provide insights into ML application in NAFLD patients' prognostics and ranked blood biomarkers are able to provide a recognizable signature allowing cost-effective NAFLD prediction and also differentiating NASH patients. Future studies should consider the limitations in the current literature and expand the application of these algorithms in different populations, fortifying an already promising tool in medical science.
**Key Words:** Artificial intelligence; Liver diseases; Healthcare; Hepatology; Prognosis; Diagnostics

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core Tip:** The ability of machine learning approaches to process multiple variables, map linear and nonlinear interactions, ranking the most important features, in addition to the capability of building accurate prediction models, sets a future direction to its application in complex diseases such as nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Future studies should consider the limitations in the current literature and expand the application of these algorithms in different populations, fortifying an already promising tool in medical science.

**Citation:** Carteri RB, Grellert M, Borba DL, Marroni CA, Fernandes SA. Machine learning approaches using blood biomarkers in non-alcoholic fatty liver diseases. *Artif Intell Gastroenterol* 2022; 3(3): 80-87


**DOI:** https://dx.doi.org/10.35712/aig.v3.i3.80

**INTRODUCTION**

Nonalcoholic fatty liver disease (NAFLD) affects an expressive part of the population worldwide and is a major cause of liver-disease related morbidity[1]. The most common cause of death in NAFLD patients is related to cardiovascular diseases, which is partially explained by the presence of metabolic comorbidities, such as obesity, type 2 diabetes, dyslipidemia, and hypertension[2]. Recently, there was concordance that the term NAFLD cannot represent the multisystemic metabolic disruption associated with the disease, resulting in the novel term MAFLD - metabolic associated fatty liver disease. Moreover, MAFLD considers the hepatic manifestation of a multimodal disease that is heterogeneous in its causes, symptoms, progression, and outcomes[3]. Nevertheless, the progression of liver fibrosis could lead to Nonalcoholic steatohepatitis (NASH), a condition characterized by histological lobular inflammation and hepatocyte ballooning[2]. Hence, detecting possible elements related to a worse prognosis in these conditions in the early stages of the disease could improve the treatment and its efficiency. Considering the significance of advanced fibrosis in NAFLD patients, differentiating NASH from steatosis is vital, reinforcing the need for cost-effective methods for risk stratification in this population[4]. Although liver biopsy is widely considered the gold standard in liver diseases investigation, it is also invasive, expensive, and prone to sampling error. In this context, the use of non-invasive biomarkers gains considerable importance[5].

The interest in artificial intelligence (AI) methods in different medical specialties, including hepatology, has dramatically increased during the last decade[6]. Advances in technology and data acquisition have simplified the collection and storage of large data sets with long time series, leading to increasingly varied fields of application, including biomedical areas. In this context, large-volume data mining evaluations had been showing promising results in recent clinical studies using machine learning methods[7-9]. More specifically, supervised machine learning (SML), can automatically detect patterns in existing training data and then use the detected patterns to predict future data[6]. Rather than considering differences between groups (as traditional statistical comparisons do), SML methods address individual differences, classifying individuals in ways that contribute to the clinical decision-making process.

The commonly late diagnosis of liver disorders contributes to suboptimal treatment and poor results. More specifically, as the prevalence of NAFLD is an important public health concern, early diagnosis of NAFLD and potential progression to NASH, could reduce the further advance of the disease, and improve patient outcomes. Using SML methods allows for collecting patient data and identifying their profile regarding the risk of developing comorbidities associated with liver damage, such as the development of metabolic syndrome or even predicting the patient’s prognosis. Several recent reviews highlighted the application of artificial intelligence in hepatology, while broadly discussing how different approaches present potential applications in several areas of hepatology[10-12]. However, specific discussion of machine learning approaches using cost-effective biomarkers could help to guide future studies towards the improvement of NAFLD diagnosis. Therefore, the objective of this mini-review is to discuss the application of SML approaches using biomarkers for the diagnosis of NAFLD and the prediction of NASH presence.
BLOOD BIOMARKERS IN NAFLD

Biomarkers are a “defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention”. This includes a plethora of possible assessments commonly investigated in NAFLD, such as blood profile, imaging (histological/radiographic) exams, specific anthropometric characteristics (body composition), and also phase angle derived from bioimpedance[13]. Noteworthy, blood biomarkers are a less invasive approach from a biological point of view and could complement imaging techniques to improve disease monitoring. In clinical settings, liver biopsy is the diagnostic gold standard for NAFLD, allowing the assessment of lipid content, inflammation, hepatocellular ballooning, and fibrotic alterations, which can also determine NASH diagnostics[14]. However, non-invasive techniques provide limited inflammation and hepatocellular ballooning determination, making objective biomarker panels for the assessment and monitoring of NAFLD or NASH a current challenge[14,15].

Nevertheless, abnormal liver function is often initially identified by nonspecific hepatocellular damage through elevations in alanine aminotransferase (ALT) and aspartate aminotransferase (AST) in addition to alkaline phosphatase and gamma-glutamyl transferase (GGT)[16]. However, ALT and AST can present normal levels while GGT can present a 1.5- fold elevation, and this response does not reflect hepatic inflammation, fibrosis, or patient metabolic risks[17,18]. Recently, cytokeratin (CK)-18 gained attention as a more specific approach for hepatocyte apoptosis since CK-18 is a major intermediate filament protein cleaved by caspases creating fragments during the apoptotic processes[19]. Assays of CK-18 fragments provide moderate accuracy due to high variability between cut-offs and respective diagnostic accuracy among studies[19]. More specifically, M30 measures caspase-cleaved CK18 produced during apoptosis, and M65 measures the total levels of (both cleaved and intact) CK18[20]. The CK-18 fragments could independently predict NAFLD severity and detect the presence of NASH with a specificity close to 90%[21,22]. In a large and heterogeneous cohort, the blood concentration of CK-18 fragments of patients with NAFLD was higher when compared with healthy volunteers and correlated to several biomarkers of liver damage and steatosis[22]. Moreover, several "biomarker panels" to grade NAFLD patients’ steatosis and fibrosis through specific scores comprise different biomarker combinations, summarized in Table 1. Notably, the FibroTest, Fibrometer, Hepascore, and Enhanced Liver Fibrosis scores are patented and commercially available panels. Nevertheless, most of the biomarker panels for the diagnosis of NAFLD and NASH, lack validation in specific cohorts, such as bariatric patients and patients with varying ethnicities[23,24]. Further, recent evidence reinforces that a combination of different commonly assessed blood-based biomarkers in addition to direct fibrogenesis markers can provide higher diagnostic accuracy in detecting advanced fibrosis when compared to current protocols. The study of Vilar-Gomez et al.[25], reviewed the diagnostic accuracy of several blood-based biomarkers, suggesting an algorithm to diagnose NAFLD patients at risk of fibrosis development. Additionally, the European guidelines recommend the combination of different tests to assess NAFLD, stating that the FibroMeter is a non-invasive alternative to liver biopsy, albeit the guidelines are not clear regarding which specific version of the FibroMeter is preferred[26]. Also, the commercially available biomarker panels and other complementary methods are not accessible for most health services, justifying the search for alternative approaches[25].

The validation study by Wu et al.[27] compared different panels of biomarkers in 417 NAFLD patients (156 with advanced fibrosis), showing that when predicting liver fibrosis scores Fibrosis-4 (FIB-4), NAFLD Fibrosis Score (NFS), AST to Platelet Ratio Index (APRI) and BARD score (BARD), it is possible to obtain a prediction of moderate fibrosis based on the receptor operator area under the curve (AUROC: 0.724, 0.671 and 0.609, respectively). The authors argued that FIB-4 and NFS performed better compared to both APRI and BARD scores, which resulted in high false-positive rates. Importantly, this study evaluated NAFLD patients based on the new definition of MAFLD, highlighting that the investigated biomarker panels provided poor performance in this setting[27]. In conclusion, the fact that the aforementioned biomarkers come from different types of procedures makes it hard for human experts to jointly analyze all this information, which motivates the use of machine learning techniques. These models can work with different types of data and discovering the relationship between them to obtain a better prediction.

ARTIFICIAL INTELLIGENCE APPLICATION IN NAFLD

Briefly, AI is an umbrella term, referring to a structured utilization of software and algorithms that analyze a wide range of data, ultimately simulating human cognition and intelligence[6]. Machine learning (ML) is one of the subdisciplines of AI, focusing on learning from data and associating specific patterns with different outcomes. An important advantage of ML techniques is that they allow the modeling of complex problems that depend on multiple input variables, justifying the application of ML methods to potentially fill several gaps in the study of complex diseases, such as NAFLD[6]. This is especially important in the case of NAFLD, which is closely related to metabolic disturbances associated with obesity and metabolic syndrome[28]. Given its complexity, NAFLD presents in different forms,
from simple asymptomatic lipid accumulation to symptomatic non-alcoholic steatohepatitis (NASH) characterized by several factors, including steatosis, hepatocellular ballooning, lobular inflammation, and often fibrosis[28]. Machine learning methods are becoming increasingly popular, which has also motivated an increase in the complexity of these models. Particularly, deep learning (DL) models, like convolutional neural networks (CNN), showed promising results in hepatology, especially with high-resolution data such as images and spectrograms[29]. Likewise, CNN models encompass several layers that involve operations like convolution, pooling, and nonlinear activations, making their decisions difficult to understand. Therefore, they represent black-box models, as opposed to interpretable (white-box) techniques, such as regression/decision trees and Bayesian networks[30,31]. Hence, ML could identify patients at risk and guide clinical treatments, whilst considering that the clinical manifestations of NAFLD appear in advanced disease status and the availability and cost of screening methods for the clinicians. Also, ML can help to rank and categorize specific biomarkers and help to elaborate specific “disease signatures”, contributing not only to clinical diagnostics, but also provide mechanistic insights for the study of the disease and the development of specific treatments.

MACHINE LEARNING APPROACHES USING BLOOD BIOMARKERS IN HEPATOLOGY

As stated above, the interest in using AI approaches to support clinical decision-making processes in hepatology has increased, albeit current literature is still scarce. Table 2 summarizes the specific studies addressing NAFLD and NASH classification. Initially, the study of Sowa et al[32] showed no differences in the investigated biomarkers (ALT, AST, and apoptotic signaling) between patients with a fibrosis score of 1 or 2. However, combining these parameters using random forests (RF) reached 79% accuracy in fibrosis prediction with a sensitivity of more than 60% and specificity of 77%. Moreover, RF identified the cell death markers M30 and M65 as more important for the decision than the classic liver parameters. Similarly, Yip et al[33] built a model to predict steatosis in a study including 922 individuals with assessment for NAFLD. The four models developed presented good diagnostic precision for steatosis (AUROC was 0.87-0.9), albeit the authors claimed that the “NAFLD ridge score” offered the best balance between efficacy and simplicity. This model included six parameters (serum triglycerides, alanine aminotransferase, high-density lipoprotein cholesterol, hemoglobin A1c, white cell count, and...
the presence of hypertension) that are routinely available for individuals undergoing medical checkups, and it does not require anthropometric measures, which are not always available. Although there is evident feasibility of the NAFLD ridge score to screen individuals, it still needs additional validation in other ethnicities. The study of Ma et al.\(^ {34} \), investigated the predictive power for NAFLD of eleven machine learning techniques, demonstrating that the Bayesian network model had the best performance, revealing that the five most discriminating features (based on information gain scores) to be weight, TG, ALT, GGT, and serum uric acid levels. Thus, in practice, users could focus on these features. Furthermore, Canbay et al.\(^ {35} \) compared different scores for the non-invasive detection of NASH. Briefly, using an ensemble feature selection approach for biomarker selection, the authors built a logistic regression model and validated in an independent study cohort of 122 patients. The logistic regression model generated from age, GGT, hemoglobin A1c, M30, and adiponectin had a strong correlation with the non-alcoholic steatohepatitis activity score and demonstrated reasonable performance to discriminate between NAFL and NASH. Likewise, Liu et al.\(^ {36} \) performed a retrospective cross-sectional study on 15315 Chinese subjects, where 5878 patients presented NAFLD. The biomarker ranking indicated the body mass index as the most valuable indicator to predict NAFLD, followed by waist circumference, triglycerides, waist-to-height ratio, and alanine aminotransferase. Notably, among seven machine learning models, the extreme gradient boosting (XGBoost) model demonstrated the best prediction ability. Similarly, the XGBoost also presented the highest AUC (0.93), accuracy (0.94), and sensitivity value (0.90) in the study of Pei et al.\(^ {37} \), comparing different models for predicting fatty liver Disease risk in 3419 participants, of which 845 had diagnostic confirmation. Importantly, regarding the biomarkers, uric acid, body mass index, and triglycerides were the most decisive risk factors for the ML models, whilst high-density lipoprotein and hemoglobin also counted as important risk factors for prediction. Strikingly, these studies provide insights into ML application in a complex context such as NAFLD patients’ prognostics. Notably, while there are investigations using AI techniques and common biomarkers to predict NAFLD and NASH, approaches using AI and novel proposed biomarkers are scarce. For instance, a recent meta-analysis showed that CK-18 is the only marker for NASH presenting external validation, with an AUROC of 0.82\(^ {38} \). Conversely, a large study conducted by the multicenter NASH Clinical Research Network demonstrated that the addition of routinely available clinical-laboratory parameters to CK-18 measurement did not significantly improve its diagnostic performance\(^ {22} \). However, it remains unknown whether the use of AI techniques combining different biomarkers in a large and diverse cohort could provide different results. Taken together, the data suggests that ranked blood biomarkers can provide a recognizable signature allowing cost-effective NAFLD prediction and also differentiating NASH patients.

### Table 2 Machine learning studies in nonalcoholic fatty liver disease and nonalcoholic steatohepatitis patients

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Patients</th>
<th>Investigated biomarker</th>
<th>Model with best performance</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sowa et al(^ {32} ), 2013</td>
<td>126 patients</td>
<td>Alanine aminotransferase; Aspartate aminotransferase; M30; M60; Hyaluronic acid</td>
<td>Random forest</td>
<td>79% Accuracy in fibrosis prediction; 60% sensitivity; 77% specificity</td>
</tr>
<tr>
<td>Yip et al(^ {33} ), 2017</td>
<td>922 patients</td>
<td>Alanine aminotransferase; High-density lipoprotein cholesterol; Triglycerides; HbA1c; White blood cells; Hypertension</td>
<td>Ridge score</td>
<td>88% Accuracy in steatosis prediction; 92% sensitivity; 90% specificity</td>
</tr>
<tr>
<td>Ma et al(^ {34} ), 2018</td>
<td>10.508 patients; 2522 NAFLD patients</td>
<td>Age; Sex; Body mass index; Alanine aminotransferase; Aspartate aminotransferase; Alkaline phosphatase; Gamma-glutamyl transpeptidase; Triglycerides; Blood urea nitrogen; Bilirubin; Cholesterol; Creatinine; Fasting glucose; Uric acid</td>
<td>Bayesian network model</td>
<td>83% Accuracy in NAFLD prediction; 68% sensitivity; 94% specificity</td>
</tr>
<tr>
<td>Canbay et al(^ {35} ), 2019</td>
<td>164 patients; 122 (validation)</td>
<td>Age; HbA1c; Gamma-glutamyl transpeptidase; M30; Adiponectin</td>
<td>Logistic regression</td>
<td>70% Accuracy in separate NAFLD and NASH</td>
</tr>
<tr>
<td>Liu et al(^ {36} ), 2021</td>
<td>15.315 patients(^ {37} ) with NAFLD</td>
<td>Body mass index; Waist circumference; Waist-to-height ratio; Alanine aminotransferase; Fasting blood glucose; Gamma-glutamyl transpeptidase; Very-low-density lipoprotein cholesterol; Low-density lipoprotein cholesterol; High-density lipoprotein cholesterol; Systolic blood pressure; Alkaline phosphatase; Diastolic blood pressure</td>
<td>XGBoost model</td>
<td>79% Accuracy in NAFLD prediction; 61% sensitivity; 90% specificity</td>
</tr>
<tr>
<td>Pei et al(^ {37} ), 2021</td>
<td>3.419 patients; 845 with fat liver diseases</td>
<td>Age; Height; Hemoglobin; Aspartate aminotransferase; Glucose; Uric acid; Low-density lipoprotein; Alpha-fetoprotein; Triglycerides; High-density lipoprotein; Carcinoembryonic antigen</td>
<td>XGBoost model</td>
<td>94% accuracy of prediction; 90% sensitivity; 95% specificity</td>
</tr>
</tbody>
</table>

NAFLD: Nonalcoholic fatty liver disease; XGBoost: Extreme gradient boosting.
CURRENT CHALLENGES IN SML APPROACHES IN HEPATOLOGY

The term "AI-Chasm" describes the gap between developing and testing an algorithm and the definitive application of the algorithm in clinical practice[39]. Unequivocally, the AI application in medical sciences is auspicious, and current literature is shedding light on a plethora of potential applications; however, many challenges for SML approaches using biomarkers in hepatology still await scrutiny.

Firstly, the collection, curation, and preprocessing of patient data is a major concern, since SML methods are data-driven[10]. Notably, the cited studies in this mini-review provide relatively small data from specific populations which could lead to sampling bias whilst limiting the generalization of the obtained results. Further, data collection should be standardized and precise, but should also be monitored for privacy and data security breaches. Secondly, as recently discussed by Quinn et al.[40], one of the main aspects of concern in future studies is the understanding that transdisciplinary approaches require cooperation to build a conceptually appropriate framework while also focusing on evaluating the performance of SML algorithms in terms of clinical endpoints and not just predictive accuracy. In addition to these technical challenges, there is also an increasing demand for transparency concerning the predictions of these models, especially in areas that have no computing background. For instance, healthcare professionals and other stakeholders that can benefit from these solutions are still reluctant to the idea of employing these methods, evidencing the necessity of educational programs aimed to explicit information about the involved decision processes. Nevertheless, the field of explainable AI has emerged to address these issues, with the purpose of creating ML techniques that produce explainable models while maintaining a high level of learning performance, enabling humans to understand and trust the predictions to support their decisions[41].

CONCLUSION

Recent advances in the field of biosciences applying machine learning algorithms resulted in promising results for the diagnosis of disease and biomarker study. The main idea is that SML could overcome the limitations of common statistical techniques. For instance, SML identifies data patterns for classification, considering multiple features at once, allowing the ranking and selection of the available blood biomarkers related to disease pathogenesis for the prediction of NAFLD or NASH, minimizing potential errors between the predicted values and the real data. Although the cited studies provide promising results, there are specific limitations that future studies should reduce. For example, most of the studies involved the Chinese population, and these algorithms still need additional validation in heterogeneous populations. The strong association between NAFLD and metabolic syndrome, obesity, and alcohol consumption may be a confounding factor in previous studies, and the application of these methods in diabetic patients with and without NAFLD could shed light on the influence of specific treatments on the performance of these ML methods. Nevertheless, the ability of ML approaches to process multiple variables, map linear and nonlinear interactions, and rank the most important features, in addition to the capability of building accurate prediction models, sets a future direction to its application in complex diseases, including NAFLD and NASH. Future studies should consider the limitations in the current literature and expand the application of these algorithms in different populations, fortifying an already promising tool in medical science.

FOOTNOTES

Author contributions: Carteri RB organized the structuring of the article in the order of the subheadings covered; Grellert M and Borba DL contributed to manuscript writing; Maroni CA and Fernandes SA supervised and revised the manuscript; All authors contributed to the writing of the article and review of the scientific article.

Conflict-of-interest statement: All authors declare that there are no conflicts of interest related to this article.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

Country/Territory of origin: Brazil

ORCID number: Randhall B Carteri 0000-0003-4124-9470; Mateus Grellert 0000-0003-0600-7054; Daniela Luisa Borba (0000-0002-2141-3993; Claudio Augusto Maroni 0000-0002-1718-6548; Sabrina Alves Fernandes 0000-0001-8504-603X.

S-Editor: Liu JH
REFERENCES


26 European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016; 64: 1388-1402 [PMID: 27062661 DOI: 10.1016/j.jhep.2015.11.004]
39 Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med* 2018; 1: 40 [PMID: 31304321 DOI: 10.1038/s41746-018-0048-x]
Artificial intelligence using advanced imaging techniques and cholangiocarcinoma: Recent advances and future direction

Aaron R Brenner, Passisd Laoveeravat, Patrick J Carey, Danielle Joiner, Samuel H Mardini, Manol Jovani

Abstract

While cholangiocarcinoma represents only about 3% of all gastrointestinal tumors, it has a dismal survival rate, usually because it is diagnosed at a late stage. The utilization of Artificial Intelligence (AI) in medicine in general, and in gastroenterology has made gigantic steps. However, the application of AI for biliary disease, in particular for cholangiocarcinoma, has been sub-optimal. The use of AI in combination with clinical data, cross-sectional imaging (computed tomography, magnetic resonance imaging) and endoscopy (endoscopic ultrasound and cholangioscopy) has the potential to significantly improve early diagnosis and the choice of optimal therapeutic options, leading to a transformation in the prognosis of this feared disease. In this review we summarize the current knowledge on the use of AI for the diagnosis and management of cholangiocarcinoma and point to future directions in the field.

Key Words: Cholangiocarcinoma; Artificial intelligence; Cholangioscopy; Artificial neural network; Machine learning; Therapeutic endoscopy; Endoscopic ultrasound

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.
Core Tip: Artificial intelligence (AI) aided by multiple imaging modalities is accurate and effective for diagnosis and characterization of biliary masses. The advancement and incorporation of imaging into artificial intelligence will help to decrease delay in diagnosis of cholangiocarcinoma and potentially decrease mortality. This review examines studies showing that AI can assist in real-time diagnosis of cholangiocarcinoma and predict outcomes of treatment. Current data suggests that AI will soon become an indispensable part of the armamentarium for the management of cholangiocarcinoma and other biliary diseases.

INTRODUCTION
The concept of AI is best explained as a computer program that possesses the ability to perform functions such as data analysis, learning, and problem solving. Medical artificial intelligence involves the development of AI programs to assist in diagnosis and prognosis, therapeutic decision making, drug development, as well as development and data mining from the electronic medical records (EMR)[1-3]. In fact, artificial intelligence is utilized in almost every field of medicine[4-8], including radiology[9], gastroenterology[10], ophthalmology[11], cardiology[12], and surgery[13].

There are many different types of AI. The foundation of the most used form, Artificial Neural Networks (ANN), takes inspiration from the human nervous system[1,3]. The neurons of ANNs are individual computer processors that interconnect and possess the capability of processing and analyzing large amounts of data[1]. ANNs are composed of links of multiple layers of these ‘neurons’, an input layer linked to multiple hidden layers, which are in turn linked to an output layer[1]. All the layers in an ANN communicate in a feed forward manner with the ability to ‘learn’ by repeatedly adjusting their links[2]. Thus, one of the attractive qualities of ANNs is in their analytical and pattern recognition ability. One of the first applications of ANNs in medicine was to aid in the diagnosis of myocardial infarction[14]. Since that time, ANNs have been widely used[1]. Support Vector Machines (SVM) is another type of machine learning which uses data analysis algorithms for classification and regression analysis[15]. SVMs are widely used in drug development and cancer detection[16,17].

Convolutional neural networks (CNN) are a type of deep learning network, a network that incorporates three or more layers, that is commonly employed in medicine, in particular because of its easy applicability to imaging[18]. Convolutional neural networks are multi-layer analyses which work by taking an image (e.g. from CT, MRI, US) and extracting layers or features at each step of the process. These features are then characterized further by complex mathematical equations to break them down and compare them to similar images, leading to pattern recognition[18]. CNNs also can place weight on the value of a specific feature, thus allowing for the presence or absence of a given variable to have a greater influence on the overall outcome.

The application of AI has grown at a rapid pace in all fields of medicine, and gastroenterology is no exception[19]. AI has been utilized in gastroenterology to identify esophageal neoplasms[20,21], diagnosis of Helicobacter pylori[22], predict gastric bleeding in patients on anti-thrombotics[23], predict the length of hospitalization for acute pancreatitis[24], differentiate between chronic pancreatitis and pancreatic cancer[25], stratify the need for ERCP[26], and characterize of colonic polyps[27]. These and many other ongoing developments will significantly impact the future of both diagnostic and therapeutic gastroenterology. One area of that has been somewhat neglected in the application of AI in gastroenterology is that of biliary disease, in particular cholangiocarcinoma. In this paper, we will review the current knowledge of the application of artificial intelligence in cholangiocarcinoma and point to the future directions in the field.

CHOLANGIOCARCINOMA
Cholangiocarcinoma (CCA) is a malignant neoplasm that can arise from anywhere along the biliary tree, including within the liver parenchyma, and is classified as distal, perihilar or intrahepatic[28]. Risk factors for CCA usually include long-term inflammatory states, like those associated with primary sclerosing cholangitis (PSC) and helminthic infection, or the continued presence of choledocholithiasis, but the majority for cases are idiopathic[29]. Cholangiocarcinoma accounts for about 3% of all gastrointestinal tumors and 10%-15% of hepatobiliary tumors[30]. Although rare, CCA has a very poor
prognosis, with 5-year survival rates following surgery rarely exceeding 35% [21]. Additionally, CCA incidence and mortality rates are increasing worldwide [31, 32]. CCA is usually detected late in the disease stage and found incidentally due to poor screening methods for early detection [32]. Early diagnosis is relatively rare, limiting the possibility of curative surgery to < 30% of patients [33]. Furthermore, even among these, about 20%-50% of patients deemed candidates for resection via preoperative evaluation are found to have unresectable disease burden during surgery [34, 35].

Given the importance of assessing disease burden, staging and location in determining a patient’s treatment plan, it is imperative to have proper preoperative imaging in CCA [36]. While pathological examination remains the gold standard of diagnosis, grading and staging for CCA, advancements in imaging and detection of biomarkers have paved the way for further preoperative predictability of malignancy type and responsiveness to therapies. These advancements have allowed for the incorporation of AI into the sphere of cholangiocarcinoma for a more accurate and personalized management of the disease [37, 38].

ARTICLE IDENTIFICATION PROCESS

The article search process was conducted in Medline and Embase [JM1]. Initial search was using different combinations of keywords such as “cholangiocarcinoma”, “biliary disease”, “cholangiography”, “artificial intelligence”, “artificial neural networks” and “convolutional neural networks”. Abstracts of major conferences, such as Digestive Disease Week and United European Gastroenterology Week were also reviewed. Finally, a comprehensive search on clinicaltrial.gov was also conducted using the same keywords to search for active clinical trials involving cholangiocarcinoma and artificial intelligence.

ARTIFICIAL INTELLIGENCE IN BILIARY DISEASES AND CHOLANGIOCARCINOMA

Artificial intelligence has been employed to advance the classification and detection of cholangiocarcinoma by aiding in creating a histopathologic database [39] and characterizing bile acid assays to better predict malignancy [40]. The use of AI to optimize the predictive value of multivariable models, and in improving the diagnostic yield of cross-sectional imaging and endoscopy has been rapidly expanding. Table 1 summarizes currently available studies.

Use of artificial intelligence in aiding the predictive abilities of multivariable models

Artificial Intelligence models have been successfully used to improve the predictive abilities of multivariable models both in the pre-interventional diagnostic phase, as well as in post-operative or post-procedural outcomes in CCA patients. Many of these studies has utilized the area under the curve (AUC), the ability of a test to diagnose a differentiate a disease state from non-disease state, to assess the added benefit of the incorporation of AI in improving the effectiveness of multivariable models.

In the preoperative phase, multiple studies have used AI/radiographic model to predict lymph node metastasis (LNM) in CCA. One study developed and validated a radiographic model for LNM detection in intrahepatic cholangiocarcinoma (ICC) based on computed tomography (CT) imaging features combined with CA19-9 values [43]. In this study, an acceptable calibration and discrimination was observed in the primary study cohort (AUC 0.8462) and in a validation cohort (AUC 0.8921) [41]. Another study developed support vector machine model utilizing magnetic resonance imaging (MRI) imaging to preoperatively evaluate for LNM in ICC. This study found that an SVM model combining CA19-9 levels and select MRI features resulted in better predictive capabilities compared to a model based on imaging features alone (AUC of 0.842 vs 0.788, P = 0.0219) [42].

One retrospective study was able to use pre-operative MRI combined with post-operative immunohistochemical results to predict early recurrence of ICC after partial hepatectomy [43]. The model that combined AI with pathology and imaging features had a higher AUC (0.949 vs 0.889, P = 0.247) compared to the model that included only the pathology and imaging features, as well as better sensitivity (0.938 vs 0.875), and specificity (0.839 vs 0.774) [43]. In another study, inclusion of AI improved the ability of a multivariable model to predict early occlusion of bilateral plastic stents placed in patients with inoperable ICC [44]. In this study, the ANN built with the multivariable model was compared to a multivariable logistic regression model alone that included age, sex, stent diameter, cancer stage, and presence of liver metastasis [44]. Overall, 288 patients were analyzed, and the ANN model outperformed the logistic regression model (AUC 0.9647 vs 0.8763, P = 0.021) [44]. Artificial intelligence has also been used to identify which serum biomarkers can have higher diagnostic power for CCA [45]. An ANN model analyzed eight biochemical markers of CCA in 85 subjects with CCA and in 82 controls [45]. Alkaline phosphatase and CCA-associated carbohydrate antigen had a higher predictive value for the distinguishing CCA patients from controls [45]. Finally, in a recent study, Müller et al. [46] developed an ANN utilizing known risk factors for ICC to predict survival in ICC patients. Using 293 patients, the ANN trained model achieved a higher AUC in predicting the 1 year survival
Use of AI in aiding cross-sectional imaging performance

Artificial Intelligence has been used to aid in the interpretation of cross-sectional imaging for nearly two decades. In a 2006 study, an artificial neural network applied to contrast-enhanced computed tomography (CE-CT) images helped differentiate four types of hepatic masses (intrahepatic peripheral cholangiocarcinoma, hepatocellular carcinoma, hemangioma, and metastatic lesions) from one-another [47]. The study then employed radiologists to evaluate CT scans with and without the assistance of ANN. There was marked improvement in diagnosis the hepatic masses with assistance from ANN compared to traditional radiologic evaluation (AUC 0.934 vs 0.888, \( P = 0.02 \), respectively) [47]. Another CT-based study was designed to predict survival outcomes and LNM in biliary tract cancers, and CT images were taken from 177 subjects who had previously undergone surgery [48]. An ANN based on CT characteristics was then built to classify the subjects into high risk or low risk for lymph node metastasis [48]. Patients who were classified as high risk based on the ANN model had a significantly lower survival rate compared to those classified as low risk [hazard ratio (HR) 3.37, 95%CI: 1.92, 5.91], underlying the importance of AI in improving prediction of disease course after treatment [48].

Artificial intelligence has also been used with MRI to improve its diagnostic/predictive power in several studies. One such study investigated the ability for an MRI based AI model to predict LNM in extrahepatic cholangiocarcinoma [37]. This was a proof-of-concept study to display the viability of a pre-operative prediction of both LMN and degree of differentiation, which could influence treatment approach. Images from 100 subjects with CCA were analyzed for the degree of CCA differentiation and rates compared to one of the most commonly used scoring system, the Fudan score (0.89 vs 0.77, \( P = 0.24 \)). In all of these studies, the addition of AI to commonly used multivariable models significantly improved their predictive abilities, improving therefore the diagnostic and post-procedural management of patient with suspected or diagnosed cholangiocarcinoma.

### Table 1 Summary of studies assessing computed tomography, magnetic resonance, and endoscopic ultrasound using artificial intelligence-based approach for pancreatic cancer

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Type of AI</th>
<th>Imaging modality</th>
<th>Training (#)</th>
<th>Testing (#)</th>
<th>AUC</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mataké et al[47], 2006</td>
<td>2006</td>
<td>ANN</td>
<td>CT</td>
<td>120 - patients</td>
<td>120 - patients</td>
<td>0.934</td>
<td>81.9</td>
<td>94.4</td>
</tr>
<tr>
<td>Ji et al[48], 2019</td>
<td>2019</td>
<td>ANN</td>
<td>CT</td>
<td>177 - patients</td>
<td>70 - patients</td>
<td>0.961</td>
<td>72</td>
<td>76.2</td>
</tr>
<tr>
<td>Logeswaran[49], 2009</td>
<td>2009</td>
<td>MLP</td>
<td>MRI</td>
<td>120 - images</td>
<td>593 - images</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Yang et al[37], 2020</td>
<td>2020</td>
<td>ANN</td>
<td>MRI</td>
<td>80 - patients</td>
<td>20 - patients</td>
<td>0.9 (LMN)</td>
<td>85.8 (LMN)</td>
<td>81.8 (LMN)</td>
</tr>
<tr>
<td>Ghandour et al[51], 2021</td>
<td>2021</td>
<td>CNN</td>
<td>Cholangioscopy</td>
<td>254 - patients</td>
<td>95 - patients</td>
<td>0.86</td>
<td>0.81</td>
<td>0.91</td>
</tr>
<tr>
<td>Robles-Medrana et al[38], 2021</td>
<td>2021</td>
<td>ML</td>
<td>Cholangioscopy</td>
<td>1714 - images</td>
<td>198 - images</td>
<td>N/A</td>
<td>92</td>
<td>N/A</td>
</tr>
<tr>
<td>Pereira et al[50], 2022</td>
<td>2022</td>
<td>CNN</td>
<td>Cholangioscopy</td>
<td>5180 - images</td>
<td>1295 - images</td>
<td>1</td>
<td>99.3</td>
<td>99.4</td>
</tr>
<tr>
<td>Pattampaisaroj et al[45], 2015</td>
<td>2015</td>
<td>ANN</td>
<td>Multivariate</td>
<td>85 - patients</td>
<td>22 - patients</td>
<td>N/A</td>
<td>98.71</td>
<td>96.94</td>
</tr>
<tr>
<td>Shao et al[44], 2018</td>
<td>2018</td>
<td>ANN</td>
<td>Multivariate</td>
<td>231 - patients</td>
<td>57 - patients</td>
<td>0.9544</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Ji et al[41], 2019</td>
<td>2019</td>
<td>N/A</td>
<td>Multivariate</td>
<td>103 - patients</td>
<td>52 - patients</td>
<td>0.8462</td>
<td>86.8</td>
<td>76.3</td>
</tr>
<tr>
<td>Xu et al[42], 2019</td>
<td>2019</td>
<td>SVM</td>
<td>Multivariate</td>
<td>106 - patients</td>
<td>42 - patients</td>
<td>0.842</td>
<td>89.36</td>
<td>57.63</td>
</tr>
<tr>
<td>Zhao et al[43], 2019</td>
<td>2019</td>
<td>N/A</td>
<td>Multivariate</td>
<td>92 - patients</td>
<td>33 - patients</td>
<td>0.949</td>
<td>0.938</td>
<td>0.839</td>
</tr>
<tr>
<td>Müller et al[46], 2021</td>
<td>2021</td>
<td>ANN</td>
<td>Multivariate</td>
<td>233 - patients</td>
<td>60 - patients</td>
<td>0.89</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

ANN: Artificial neural network; MLP: Multi-layer perceptron; CNN: Convolutional neural network; ML: Machine learning; SVM: Support vector machine; CT: Computed tomography; MRI: Magnetic resonance imaging; N/A: Not applicable.
lymph node metastasis. The AI model had an AUC of 0.9 (95%CI: 0.66, 1.0) for predicting LMN while the AUC for degree of differentiation was 0.80 (95%CI: 0.58, 0.97)\cite{37}. In another study, an ANN model based on MRCP images was able to distinguish between patients with CCA from those without CCA \cite{49}. A total of 309 images were processed, 248 of which were normal and 61 were taken from patient with CCA. The ANN model achieved an accuracy of 94% for distinguishing between them. Furthermore, ANN achieved an accuracy of 88% in distinguishing between images of CCA and images of other common biliary diseases, such as cholecystitis, choledocholithiasis, PSC, and cholangitis\cite{49}.

Use of AI in aiding endoscopic evaluation of biliary diseases/cholangiocarcinoma

Artificial intelligence has also more recently been used to aid in endoscopic diagnosis of cholangiocarcinoma or other biliary diseases, even though most studies are currently in abstract form only. A study by Pereira et al\cite{50} developed a CNN that differentiates biliary strictures as benign or malignant based on images from digital single operator cholangioscopy. After an evaluation of 6475 images from 85 patients with indeterminate biliary strictures, the authors found a sensitivity of 99.3%, specificity of 99.4%, and AUC of 1.00 for a correct diagnosis. In another study, currently available only as an abstract, the authors developed a CNN to detect abnormal biliary features \textit{via} cholangioscopy images\cite{51}. They defined abnormal features as presence of papillary mass, tortuous vessels, or ulcerations. Over 1000000 images were from 528 patients were evaluated for the study. The CNN showed an AUC of 0.86 (95%CI: 0.80, 0.92), sensitivity of 0.81 (95%CI: 0.72, 0.91), and specificity of 0.75 (95%CI: 0.86, 0.97)\cite{51}. In another recent study, the utility of AI to perform real-time diagnosis of biliary strictures during cholangioscopy was assessed. This model was built using 25 cholangioscopy videos and was then tested on known cases (20 live cholangioscopy and 20 videos of cholangioscopy) of malignant biliary strictures. It accurately predicted malignancy in every case\cite{38}. These initial results suggests that introduction of AI into standard clinical practice could potentially decrease time to diagnosis of indeterminate biliary strictures and allow for better diagnostic accuracy.

Endoscopic ultrasound (EUS) in combination with AI has been used in the assessment of pancreatic disease and may be beneficial in assisting in real-time differentiation between pancreatic masses and other solid masses during endoscopy\cite{52}. However, there has been limited use of AI during EUS evaluations for cholangiocarcinoma. One recent study developed an AI system to recognize standard stations of EUS for biliary duct evaluation. In this study, AI had comparable accuracy to that of expert endosonographers, and significantly improved the learning curve of trainees\cite{53}.

CHOLEDOCHOLITHIASIS

Artificial intelligence has also been useful for the study of possible risk factors for CCA, such as choledocholithiasis. Several studies have demonstrated that AI can be used to risk-stratify patients with possible choledocolithiasis and therefore aid in the decision-making of the need for ERCP\cite{54,55}. One study showed that a machine learning model using pre-ERCP imaging, including US and CT, in addition to select demographic features and laboratory findings can achieve a sensitivity of 97.7% and specificity of 100% in identifying choledocholithiasis\cite{55}. Another study found that an AI model outperformed ASGE guidelines for proper indication for an ERCP (AUC 0.79 \textit{vs} 0.59, respectively)\cite{54}. In addition, the use of AI would avoid the need for ERCP in 36% of cases who would have undergone the procedure according to the ASGE guidelines\cite{54}. Once more, the addition of AI can help providers achieve an individualized management program for patients in daily clinical practice.

CONCLUSION

The diagnosis and staging of cholangiocarcinoma is challenging, leading to potential major non-curative surgeries and/or dismal survival rate because of late diagnosis and inadequate prediction of metastases or recurrence using standard diagnostic methods. The introduction of AI technologies to traditional cross-sectional imaging and endoscopy, can create a major shift in the diagnosis and management of CCA. As mentioned above, many studies have already incorporated AI with significant improvement over traditional clinical data. While most of these studies are retrospective in nature, and therefore provide relatively poor quality data, they are very encouraging.

In addition, new studies are currently ongoing in which AI technologies are used to diagnose and risk-stratify patients with cholangiocarcinoma. The Synergy-AI clinical trial for example, is a non-interventional prospective observational study currently enrolling participants with cholangiocarcinoma, along with other malignancies. This trial is employing an Application Programming Interface to help match participants with personalized treatment protocols based on CT imaging, biomarkers, and laboratory results. In this setting, AI is expected to identify both the most cost effective, appropriate, and personalized treatment approach to each individual’s malignancy\cite{56}. Considering that most hospitals have incorporated electronic medical records (EMR) for their patients, it is easy to see how AI can be
used to select different patient variables (biochemical, histological or cross-sectional imaging) and use them to help develop personalized management strategies which optimize outcomes. Combining biomarkers, genetic sequencing, and imaging through AI models could lead to new approaches to the diagnosis and treatment of cholangiocarcinoma, including decreasing the need for unnecessary invasive endoscopic procedures for procurement of biopsies, as well as help develop a more targeted approach for therapy[57]. While more research and fine tuning of current AI systems is needed before reaching this stage, the future of AI in the management of cholangiocarcinoma seems clearly within reach.

FOOTNOTES

Author contributions: All authors contributed to the paper with regard to conception and design of the study, literature review and analysis, drafting the manuscript and all authors approved the final version of the manuscript.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Open-Access: This is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

Country/Territory of origin: United States

ORCID number: Aaron R Brenner 0000-0001-8816-9182; Passisd Laoveeravat 0000-0001-6855-0437; Patrick J Carey 0000-0001-6521-4825; Manol Jovani 0000-0001-6803-8521.

Corresponding Author’s Membership in Professional Societies: American Gastroenterological Association, No. 1612153.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

Brenner AR et al. Review of artificial intelligence in cholangiocarcinoma


