

Artificial Intelligence in Viral Hepatitis: Promise, Reality, And Barriers to Clinical Translation

Sanskruti Sawant¹, Dr. Prabhat Dessai²

¹Second-Year Bachelor of Pharmacy (B.Pharm) Student, Department of Pharmacy, Parul University, Goa Campus, Goa, India

²Associate Professor, Department of Pharmacy, Parul University, Goa Campus, Goa, India

Corresponding author: sawantsanskruti1613@gmail.com

Doi: <https://doi.org/10.5281/zenodo.22302599>

Received: 09th July 2026

Accepted: 19th July 2026

ABSTRACT

Background: Viral hepatitis affects an estimated 311 million people worldwide, with the greatest burden falling on low- and middle-income countries (LMICs). Artificial intelligence (AI) has emerged as a promising tool to improve screening, diagnosis, risk stratification, and clinical management. However, its real-world clinical utility and readiness for implementation remain uncertain.

Methods: We conducted a comprehensive narrative review of artificial intelligence and machine-learning applications in chronic hepatitis B (HBV) and hepatitis C (HCV) across the continuum of care. Evidence was critically appraised with emphasis on methodological quality, external validation, clinical applicability, and equitable implementation, informed by TRIPOD+AI and SPIRIT-AI principles.

Results: Across more than 50 studies, AI demonstrated strong technical performance (AUC approximately 0.80–0.95) for case-finding, fibrosis staging, and hepatocellular carcinoma risk prediction. The Intelligen-C clinical decision support system remains the only reported hepatitis AI tool with documented real-world implementation, demonstrating substantial improvements in HCV case-finding and healthcare efficiency. Nevertheless, external validation was uncommon, evidence from LMICs remained limited, prospective randomized trials were lacking, and economic evaluations were scarce. Important evidence gaps also persist in pediatric hepatitis, hepatitis D virus infection, and HBV/HIV or HCV/HIV coinfection.

Conclusion: Although AI has demonstrated considerable technical potential in viral hepatitis, improvements in model performance alone are insufficient to establish clinical benefit. Future research should prioritize prospective validation,

multicenter implementation studies, equitable evaluation across diverse populations, and rigorous health-economic assessment to enable responsible integration of AI into global hepatitis elimination strategies.

Keywords: Artificial intelligence, Machine learning, Viral hepatitis, Hepatitis B, Hepatitis C, Clinical translation, Clinical decision support systems, Fibrosis staging, Hepatocellular carcinoma prediction, External validation, Randomized controlled trials, Low- and middle-income countries, Health equity, Implementation science, WHO hepatitis elimination

INTRODUCTION

Viral hepatitis remains one of the leading yet under recognized causes of global morbidity and mortality. In 2022, an estimated 257.5 million people were living with chronic hepatitis B virus (HBV) infection, while approximately 57 million were infected with chronic hepatitis C virus (HCV). Together, these infections are responsible for nearly 1.3–1.5 million deaths each year.^[1,2] Chronic HBV is estimated to account for about 42% of cirrhosis cases worldwide, whereas HCV contributes to approximately 21%.^[3] Although highly effective direct-acting antivirals can cure HCV and long-term nucleos(t)ide analogue therapy can effectively suppress HBV replication, access to diagnosis and treatment remains inadequate in many parts of the world. A large proportion of infected individuals are still unaware of their infection, and treatment coverage continues to fall well below the levels needed to achieve the World Health Organization's hepatitis elimination goals.

The World Health Organization (WHO) has set ambitious targets to eliminate viral hepatitis as a public health threat by 2030, including diagnosing 90% of individuals living with chronic HBV or HCV infection and providing treatment to 80% of those who are eligible.^[4] Achieving these goals, particularly in low- and middle-income countries (LMICs), requires innovative approaches to expand screening, improve risk stratification, optimize treatment decisions, and strengthen healthcare delivery. In this context, artificial intelligence (AI) and machine learning (ML) have emerged as promising technologies with the potential to support earlier diagnosis, enhance prognostic assessment, personalize clinical decision-making, and accelerate antiviral drug discovery.

Despite encouraging advances, several important questions remain unresolved. It is still uncertain whether the high diagnostic performance reported by many AI models translates into meaningful improvements in patient outcomes, such as increased cure rates, prevention of cirrhosis, earlier detection of hepatocellular carcinoma, or reduced mortality. Furthermore, many published models have been developed using geographically restricted datasets, raising concerns regarding their external validity across genetically diverse populations and different healthcare systems. Practical challenges related to infrastructure, affordability, data quality, regulatory oversight, and clinical integration also limit the adoption of AI, particularly in resource-constrained settings.

Previous reviews have comprehensively summarized the expanding role of AI in viral hepatitis and hepatology; however, comparatively less attention has been given to the methodological quality of published studies, external validation, implementation barriers, cost-effectiveness, and equity considerations.^[5,6] Therefore, this narrative review critically evaluates current evidence on AI and ML across the viral hepatitis care continuum, discusses key challenges

to clinical translation, and highlights governance and implementation strategies needed to facilitate safe, equitable, and sustainable integration of AI into hepatitis elimination programs.

CONVENTIONAL MANAGEMENT: BASELINE FOR AI COMPARISON

Current management of chronic HBV infection follows well-established evidence-based algorithms that integrate virological, biochemical, and clinical assessments. Initial screening with hepatitis B surface antigen (HBsAg) is followed by HBV DNA quantification, evaluation of liver injury and function, and non-invasive assessment of hepatic fibrosis using validated tools such as the Fibrosis-4 (FIB-4) index, the Aspartate Aminotransferase-to-Platelet Ratio Index (APRI), or transient elastography. Decisions regarding antiviral therapy are guided by HBV DNA levels, alanine aminotransferase (ALT) activity, the severity of liver fibrosis, and other patient-specific clinical factors, including the presence of cirrhosis and risk of disease progression.^[7,8]

For HCV infection, current guidelines recommend universal screening using anti-HCV antibody testing, followed by reflex HCV RNA testing to confirm active infection. The widespread availability of direct-acting antivirals (DAAs) has transformed HCV management, with sustained virological response (SVR) rates exceeding 95% in most patient populations, making viral eradication an achievable goal for the majority of treated individuals.^[9]

Table 1: Non-Invasive Fibrosis Assessment Tools

NIT	AUC (F $\geq$ 2)	Cost / Accessibility	LMIC Availability
APRI / FIB-4	0.72 – 0.80	Free / universally accessible	Widely available
Transient elastography	0.84 – 0.90	Specialized equipment (moderate to high cost)	Limited availability
MRI elastography	0.90 – 0.96	Research setting (high cost)	Rarely available

*AUC: Area under the receiver operating characteristic curve; F $\geq$ 2: Significant fibrosis (METAVIR $\geq$ F2);
LMIC: Low- and middle-income countries.*

Abbreviations

APRI – Aspartate Aminotransferase-to-Platelet Ratio Index

FIB-4 – Fibrosis-4 Index

AUC – Area Under the Receiver Operating Characteristic Curve

F $\geq$ 2 – Significant Fibrosis (METAVIR stage F2 or higher)

MRI – Magnetic Resonance Imaging

LMIC – Low- and Middle-Income Countries

NITs – Non-Invasive Fibrosis Tools

Despite the availability of several non-invasive methods for fibrosis assessment, accurate staging of liver disease remains challenging. Although tools such as the Fibrosis-4 (FIB-4) index, the Aspartate Aminotransferase-to-Platelet Ratio Index (APRI), and transient elastography have substantially reduced the need for liver biopsy, a considerable proportion of patients still fall into indeterminate diagnostic categories that require additional evaluation.^[10] Likewise, established hepatocellular carcinoma (HCC) risk prediction models, including PAGE-B and REACH-B, demonstrate only moderate predictive performance across different patient populations, limiting their ability to consistently identify individuals at highest risk while avoiding unnecessary surveillance in lower-risk groups^[11]

These limitations provide an important opportunity for artificial intelligence (AI)-based approaches. However, meaningful clinical adoption will depend not only on improved statistical performance but also on rigorous external validation, use of robust reference standards, demonstration of clinical benefit, and successful integration into diverse healthcare settings.

MACHINE LEARNING FOR SCREENING AND CASE-FINDING

Case-Finding Using EHR and Laboratory Data

Machine learning has shown considerable promise in improving targeted screening and case-finding for viral hepatitis by leveraging electronic health records (EHRs) and routinely collected laboratory data. In a large U.S. healthcare system, Rigg et al. (2023) developed gradient boosting models using EHR-derived clinical variables to identify individuals at increased risk of previously undiagnosed HCV infection. The models achieved an area under the receiver operating characteristic curve (AUC) of approximately 0.82–0.85, demonstrating improved identification of high-risk individuals compared with untargeted screening strategies.^[12] A key strength of this study was its integration into real-world clinical workflows using routinely available healthcare data. However, the findings were derived from a single healthcare system without external validation, and the study did not evaluate downstream clinical outcomes such as linkage to care, treatment uptake, sustained virological response, or cost-effectiveness. Consequently, the applicability of these models to geographically diverse populations and resource-limited healthcare settings remains uncertain.

Similarly, Kim et al. (2023) evaluated multiple machine-learning algorithms, including Random Forest, Extreme Gradient Boosting (XGBoost), and Least Absolute Shrinkage and Selection Operator (LASSO), to predict HBV and HCV infection among approximately 15,000 individuals with diabetes using data from the U.S. National Health and Nutrition Examination Survey (NHANES). The best-performing models achieved AUC values of **0.85–0.90**, outperforming conventional risk factor-based approaches.^[13] Nevertheless, the study population was restricted to diabetic individuals from the United States, limiting the generalizability of the findings to non-diabetic populations and regions with a high burden of HBV infection, particularly low- and middle-income countries where many infections remain undiagnosed.

The Intelligen-C Exception: Real-World Deployment

Among the published applications of artificial intelligence in viral hepatitis, the Intelligen-C clinical decision support system (CDSS) is one of the few to demonstrate implementation within routine clinical practice.^[14] The system continuously analyzed electronic health record (EHR) data to identify individuals at increased risk of previously undiagnosed HCV infection, generated automated alerts for healthcare providers, and facilitated confirmatory HCV testing. Following implementation, HCV serology orders increased from 909 to 7,312, the number of HCV RNA-positive patients identified increased from 3 to 28, and the number of patients initiating antiviral treatment increased from 1 to 9, indicating substantial improvements in case identification and treatment initiation within the participating healthcare system.^[14]

Despite these encouraging findings, several important limitations should be considered. The study was conducted within a single healthcare system, and the proprietary nature of the algorithm limits independent validation and reproducibility. Furthermore, the intervention was not evaluated in a randomized controlled trial, and longer-term patient outcomes including sustained virological response (SVR), progression to cirrhosis or hepatocellular carcinoma, mortality, and health-related quality of life were not assessed. The study also did not examine algorithmic fairness across demographic subgroups or evaluate implementation in low- and middle-income countries, where healthcare infrastructure and disease epidemiology differ substantially. Consequently, although Intelligen-C provides valuable proof-of-concept for AI-assisted HCV case-finding, further multicenter prospective studies with external validation and health economic analyses are required before widespread clinical adoption can be recommended.

Table 2: Evidence Quality and Clinical Readiness of AI Applications in Viral Hepatitis

AI Application	Representative Performance	External Validation	Published RCTs	Real-World Implementation	Estimated TRL (1–9)
 1. Screening / Case-finding	Representative AUC 0.80–0.90	Limited	None	Intelligen-C (single health system)	6–7
 2. Fibrosis Assessment	Representative AUC 0.80–0.97	Limited	None	No published large-scale implementation	4–6
 3. HCC Risk Prediction	Representative AUC 0.80–0.88	Limited	None	No published large-scale implementation	4–5
 4. Decompensation Prediction	Representative AUC 0.80–0.85	Rare	None	No published large-scale implementation	4–5
 5. Treatment Response (HCV)	Variable across studies	Rare	None	No published large-scale implementation	3–4
 6. HBsAg Loss Prediction	Representative AUC ~0.98	Prospective validation reported	None	Research setting	5–6
 7. Drug Discovery	Predominantly preclinical (<i>in silico</i>)	Not applicable	Preclinical only	None	2–4

Table Legend: Representative performance values are derived from the principal studies discussed in this review and are intended to illustrate typical reported performance rather than pooled estimates. Estimated TRL values were assigned by the authors based on published evidence of technical validation, clinical evaluation, and implementation.

Abbreviations

AI – Artificial Intelligence

AUC – Area Under the Receiver Operating Characteristic Curve

C-index – Concordance Index

DL – Deep Learning

HBsAg – Hepatitis B Surface Antigen

HCC – Hepatocellular Carcinoma

HCV – Hepatitis C Virus

ML – Machine Learning

N/A – Not Applicable

TRL – Technology Readiness Level

AI FOR FIBROSIS PREDICTION AND STAGING

Serum Marker-Based Machine Learning Models

Machine-learning models based on routinely available serum biomarkers have demonstrated encouraging performance for non-invasive fibrosis assessment in chronic viral hepatitis. Wei et al. (2018) developed the LiveBoost model using gradient boosting algorithms and four routinely available clinical variables (aspartate aminotransferase, alanine aminotransferase, platelet count, and age). Across HBV and HCV cohorts, the model achieved AUC values of approximately 0.85–0.90, outperforming conventional non-invasive indices such as FIB-4.^[15] The study was strengthened by external validation across multiple cohorts and the development of a publicly accessible web-based prediction tool. However, the analysis was retrospective, and biopsy-based reference standards from selected centers may have introduced sampling and observer-related biases. Although LiveBoost demonstrated improved discriminatory performance, its effect on clinical decision-making, patient management, and long-term outcomes has not yet been established.

More recently, Liu et al. (2024) evaluated decision tree and gradient boosting algorithms for fibrosis staging in patients with chronic hepatitis B, reporting AUC values ranging from 0.89 to 0.94 across METAVIR fibrosis stages.^[16] An important observation across multiple studies is the consistent identification of similar predictive variables including platelet count, HBV DNA levels, and markers of hepatic synthetic function suggesting that these features are robust predictors of fibrosis severity. Nevertheless, most available studies remain retrospective, are conducted at single centers, and have undergone limited external validation, highlighting the need for prospective multicenter evaluation before routine clinical implementation.

An important methodological consideration is that many AI models are trained using imperfect reference standards. Liver biopsy, although regarded as the clinical reference standard, is subject to sampling error and inter-observer variability, while commonly used non-invasive indices such as APRI and FIB-4 also have recognized limitations in diagnostic accuracy.^[17] Consequently, the performance of supervised machine-learning models is inherently influenced by the quality of the reference labels used during training. This methodological constraint should be

considered when interpreting incremental improvements in model performance and their potential clinical significance.

Imaging-Based Deep Learning

Deep learning combined with imaging has shown excellent diagnostic performance for non-invasive fibrosis assessment. In a prospective multicenter study, Wang et al. developed a convolutional neural network (CNN)-based radiomics model using two-dimensional shear-wave elastography (2D-SWE) images for staging liver fibrosis in chronic hepatitis B. The model achieved AUC values of approximately 0.95–0.97, demonstrating excellent discrimination across fibrosis stages.^[18] The study is strengthened by its prospective design, multicenter enrollment, external validation, and rigorous methodological framework, making it one of the strongest clinical evaluations of AI-assisted fibrosis staging to date.

Despite these encouraging findings, important barriers to widespread implementation remain. Two-dimensional shear-wave elastography requires specialized imaging equipment, trained operators, and dedicated technical expertise, limiting its availability in many resource-constrained healthcare settings. Consequently, the AI approaches with the strongest diagnostic performance may currently be less accessible in regions that carry a substantial proportion of the global HBV burden. Furthermore, although imaging-based deep learning has consistently demonstrated excellent technical performance across published studies (AUC approximately 0.84–0.97), evidence based on conventional B-mode ultrasound remains relatively limited compared with elastography-based approaches, highlighting an important gap for future research.^[19]

AI FOR DISEASE PROGRESSION: HCC AND DECOMPENSATION PREDICTION

HBV Decompensation and HCC Risk Prediction

Machine-learning approaches have increasingly been explored for predicting hepatic decompensation and hepatocellular carcinoma (HCC) risk in patients with chronic hepatitis B. Akkari et al. (2025) developed machine-learning models using routinely collected electronic health record (EHR) variables to predict clinically significant decompensation events, including ascites, variceal bleeding, and jaundice, in patients with HBV-related cirrhosis. The best-performing models achieved AUC values of approximately 0.80–0.85, demonstrating good predictive performance for clinically relevant outcomes.^[20] Strengths of the study include its HBV-specific cohort, clinically meaningful endpoints, and comparison of multiple machine-learning algorithms. However, its retrospective single-center design, lack of external validation, absence of direct comparison with established prognostic scores such as MELD or Child–Pugh, and failure to assess the impact of AI-guided predictions on clinical decision-making limit the immediate applicability of these findings.

Machine-learning models have also been developed for HCC risk prediction in chronic hepatitis B, with reported AUC values generally ranging from 0.80 to 0.88.^[21,22] Nevertheless, relatively few studies have directly compared these models with established risk prediction tools such as PAGE-B and REACH-B, making it difficult to determine whether AI offers clinically meaningful improvements over current guideline-based approaches. Furthermore, a substantial

proportion of HBV-related HCC develops in patients without established cirrhosis.^[23] Consequently, AI models trained predominantly in cirrhotic populations may have reduced ability to identify individuals at risk of non-cirrhotic HCC, representing an important limitation for population-wide surveillance strategies.

Counter-argument

Despite these limitations, proponents argue that even modest improvements in predictive performance may have meaningful clinical implications. More accurate risk stratification could enable individualized surveillance strategies by identifying patients who may benefit from closer monitoring while reducing unnecessary investigations in those at lower risk. Although this rationale is biologically and clinically plausible, there is currently limited prospective evidence demonstrating that AI-guided surveillance improves early HCC detection, reduces HCC-related mortality, or is more cost-effective than existing guideline-recommended approaches. Well-designed prospective multicenter studies are therefore required to determine whether these promising technical advances translate into measurable clinical benefit.

HCV Post-SVR HCC and Liver-Related Event Risk Prediction

Although direct-acting antivirals (DAAs) achieve sustained virological response (SVR) in more than 95% of patients, individuals with advanced fibrosis remain at risk of hepatocellular carcinoma (HCC). AI-Safe-C and similar machine-learning models use random survival forest algorithms integrating liver stiffness measurement (LSM) with routinely available clinical variables to predict HCC and other liver-related events following successful DAA therapy.^[24] The AI-Safe-C model demonstrated good discriminatory performance, with a Harrell's C-index of approximately 0.87 for HCC prediction.

A major strength of these models is the integration of elastography with routinely available clinical data, allowing individualized post-SVR risk stratification beyond conventional clinical assessment. However, despite encouraging performance, prospective implementation studies, broader validation across geographically diverse populations, and formal cost-effectiveness analyses remain limited. Consequently, the incremental clinical benefit of AI-assisted post-SVR surveillance over existing guideline-based strategies has yet to be established.

AI FOR TREATMENT RESPONSE PREDICTION

HCV DAA Response Prediction

Machine-learning models have also been developed to predict treatment failure following direct-acting antiviral (DAA) therapy for chronic hepatitis C. Using data from more than 6,500 patients in the HCV-TARGET cohort, Shen et al. (2022) evaluated Elastic Net, Random Forest, and Gradient Boosting Machine algorithms, reporting modest predictive performance (AUC approximately 0.66–0.69) that exceeded conventional logistic regression.^[25] Baseline viral load and comorbidities emerged as important predictors of treatment failure. However, the study was retrospective, lacked prospective validation, and did not evaluate whether identifying high-risk patients altered treatment decisions or improved clinical outcomes.

Counter-argument

In the era of highly effective DAAs, where sustained virological response exceeds 95% in most patient populations, the incremental clinical value of predicting treatment success is relatively limited because current guidelines recommend treatment for nearly all eligible patients. AI may therefore provide greater clinical benefit by addressing more challenging questions, such as predicting antiviral resistance, relapse after treatment, or long-term complications following viral eradication areas that remain comparatively underexplored.

HBV HBsAg Loss and Relapse

Prediction of functional cure has emerged as an important application of AI in chronic hepatitis B. Kong et al. (2024) developed the GOLDEN score, a longitudinal model using serial quantitative HBsAg (qHBsAg) measurements in 6,792 patients treated with pegylated interferon- α 2b.^[26] The model demonstrated excellent predictive performance for 5-year HBsAg seroclearance (AUC ~0.98) with independent external validation, highlighting the value of longitudinal biomarker trends over static baseline measurements.

Prediction of HBV relapse after treatment cessation, however, still relies largely on conventional scores such as CREATE and SCALE-B. Published machine-learning models in this area remain limited, making relapse prediction an important opportunity for future AI research.

Advantages, Limitations, and Critical Barriers

What AI Has Demonstrated?

Technical performance

Machine-learning and deep-learning models have consistently achieved strong discrimination (AUC approximately 0.80–0.95) for screening, fibrosis staging, and HCC risk prediction, often outperforming conventional scores such as APRI and FIB-4.

The Intelligen-C clinical decision support system remains the only reported hepatitis AI tool with documented real-world implementation, demonstrating substantial improvements in HCV case-finding and linkage to testing.

The GOLDEN model illustrates the potential of longitudinal AI approaches, achieving excellent prediction of HBsAg seroclearance (AUC ~0.98).

Critical Limitations

Limited external validation. Most published AI models have been developed and evaluated within single-center retrospective cohorts, with relatively few undergoing independent external validation. This raises concerns regarding overfitting and generalizability. Although internal validation is appropriate during model development, external validation is essential before clinical implementation.^[27]

Lack of prospective clinical evidence. Randomized controlled trials evaluating AI-guided hepatitis management remain lacking. Existing studies primarily report improvements in diagnostic processes or model performance rather

than demonstrating benefits in patient outcomes such as treatment response, decompensation, HCC incidence, or mortality. Geographic and population bias. Most AI studies in viral hepatitis have been conducted in East Asian populations, with comparatively limited evidence from sub-Saharan Africa and other low- and middle-income regions. Broader validation across diverse populations and HBV genotypes is necessary to ensure equitable and generalizable clinical application.^[28]

Table 3: Key Research Gaps and Future Priorities for Clinical Translation of Artificial Intelligence in Viral Hepatitis

Research Gap	Current Evidence	Why It Matters	Future Research Priority
1. Lack of prospective trials	No prospective randomized controlled trials evaluating AI interventions in hepatitis care have been published.	Clinical benefit cannot be established without randomized evidence.	Conduct multicenter prospective clinical trials to evaluate whether AI-assisted strategies improve patient outcomes, efficiency, and health system performance.
2. Limited external validation	External validation was uncommon and reported in a minority of studies.	Models are prone to overfitting and may not generalize to new settings or populations.	Ensure independent external validation in geographically and demographically diverse cohorts and report calibration and generalizability measures.
3. Limited representation of diverse HBV genotypes and populations	Most models were developed using cohorts from East Asia, predominantly genotype B and C.	Performance may be suboptimal in other genotypes and populations.	Validate and refine models across all major HBV genotypes (A–H) and diverse ethnic and geographic populations to improve equity and generalizability.
4. Underrepresentation of low- and middle-income countries (LMICs)	Evidence from LMICs remains limited, with most studies originating from high-income countries.	Limits applicability in high-burden, resource-constrained settings where the need is greatest.	Promote investigator-led research and capacity building in LMICs and validate tools in real-world, resource-limited healthcare settings.
5. Limited evidence on cost-effectiveness	Only a small number of studies report economic evaluations of AI tools.	Health systems need economic evidence to justify investment and scale-up.	Conduct comprehensive cost-effectiveness and budget impact analyses to guide policy and implementation decisions.
6. Insufficient implementation science	Few studies address adoption, user acceptance, workflow integration, or sustainability.	Without implementation evidence, real-world uptake and scalability remain uncertain.	Undertake implementation research to evaluate feasibility, acceptability, integration into clinical workflows, and sustainability at scale.
7. Limited focus on pediatric and special populations	Very limited evidence in children, HDV infection, and HBV/HIV or HCV/HIV coinfection.	Findings from adult populations may not apply to these groups.	Develop and validate AI models specifically for pediatric populations, HDV infection, and viral hepatitis in the context of HIV coinfection.

Abbreviations

AI – Artificial Intelligence

RCTs – Randomized Controlled Trials

LMICs – Low- and Middle-Income Countries

WHO – World Health Organization

NIH – National Institutes of Health

Bill & Melinda Gates Foundation – Global philanthropic organization supporting health research

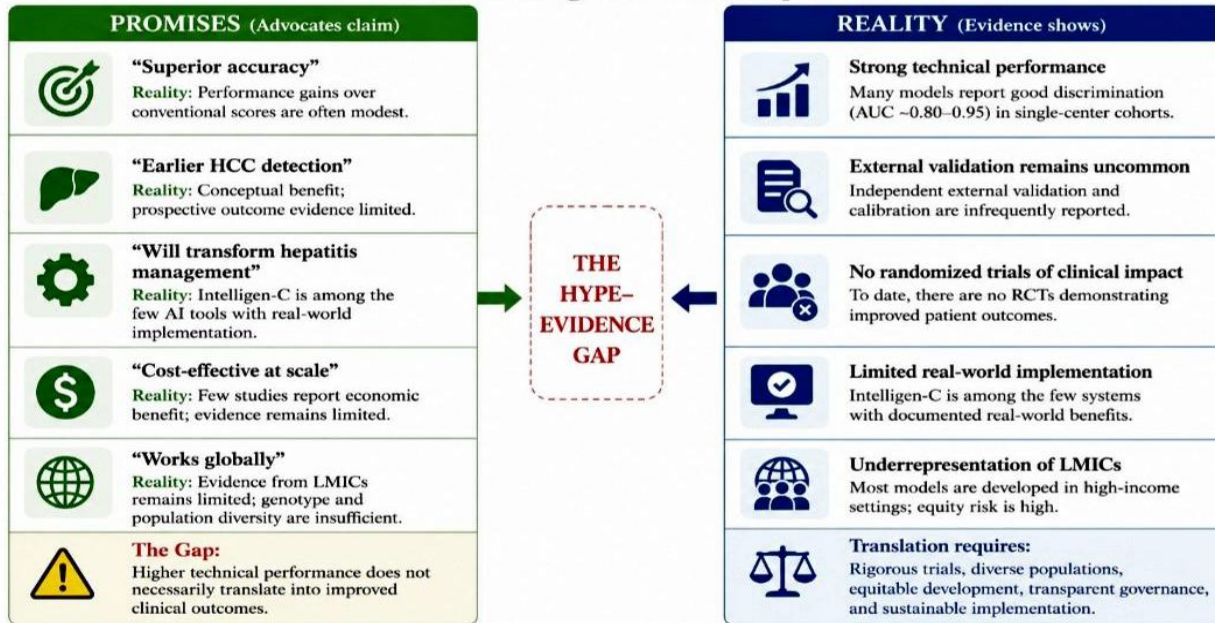
Global Fund – The Global Fund to Fight AIDS, Tuberculosis and Malaria

Cost-effectiveness – Assessment of whether a health intervention provides sufficient value relative to its cost

Implementation science – The scientific study of methods to promote the adoption and integration of evidence-based interventions into routine healthcare practice

Genotype – Genetic variant or strain of a virus (e.g., hepatitis B virus genotypes A–E)

Figure 1. Promise versus Clinical Reality: Translational Challenges of Artificial Intelligence in Viral Hepatitis



Abbreviations: AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; APRI, aspartate aminotransferase-to-platelet ratio index; FIB-4, fibrosis-4 index; HCC, hepatocellular carcinoma; LMIC, low- and middle-income country; RCT, randomized controlled trial.

Figure 1: Promise vs. Reality in Hepatitis AI The Hype-Evidence Gap

Abbreviations

AI – Artificial Intelligence

AUC – Area Under the Receiver Operating Characteristic Curve

APRI – Aspartate Aminotransferase-to-Platelet Ratio Index

FIB-4 – Fibrosis-4 Index

HCC – Hepatocellular Carcinoma

LMIC – Low- and Middle-Income Countries

RCT – Randomized Controlled Trial

Intelligen-C – Artificial Intelligence based Clinical Decision Support System (CDSS) for Hepatitis C Virus screening and case-finding

CDSS – Clinical Decision Support System

NOTE: The figure contrasts the reported promises of artificial intelligence with the current level of clinical evidence in viral hepatitis. Although many artificial intelligence models demonstrate high technical performance, most lack prospective validation, randomized clinical trial evidence, and broad implementation across diverse healthcare settings, particularly in low- and middle-income countries.

Additional Challenges to Clinical Translation

LMIC Applicability

Many AI models depend on technologies that are not routinely available in resource-constrained settings. Imaging-based deep learning often requires specialized modalities such as 2D shear-wave elastography or advanced MRI, while many machine-learning models rely on well-developed electronic health record (EHR) systems that remain unavailable in numerous low- and middle-income countries.^[29] In addition, evidence on the cost-effectiveness of AI-assisted hepatitis care remains limited, making large-scale implementation uncertain.

Label Noise

Many AI models are trained using imperfect reference standards, including liver biopsy, which is subject to inter-observer variability, or surrogate fibrosis scores such as APRI and FIB-4 with known diagnostic limitations. Consequently, modest improvements in model performance may reflect reductions in measurement error rather than genuine improvements in diagnostic accuracy.

Methodological Gaps

Although many studies report excellent discrimination, relatively few evaluate model calibration, clinical utility, or decision-curve analysis, limiting assessment of real-world performance. Greater adherence to standardized reporting frameworks such as TRIPOD-AI is essential to improve transparency, reproducibility, and clinical interpretability.^[30,31]

Counter-argument

Although exploratory machine-learning studies may reasonably prioritize algorithm development over comprehensive reporting, models intended for clinical implementation require rigorous external validation, appropriate calibration, transparent reporting, and prospective evaluation before they can be incorporated into routine hepatitis care.^[31]

Equity and Fairness Concerns

Most published AI studies in viral hepatitis have been developed using datasets from high-income or East Asian populations, with limited representation from low- and middle-income countries, sub-Saharan Africa, and other underserved populations. Few studies evaluate model performance across demographic subgroups such as sex, ethnicity, or socioeconomic status, raising concerns about the generalizability of these models.

In addition, fairness metrics such as demographic parity and equalized odds are rarely reported, making it difficult to determine whether AI models perform equitably across diverse patient populations.^[32] Without local validation, deploying AI models developed in well resourced settings to resource limited healthcare systems may introduce algorithmic bias, reduce clinical trust, and inadvertently widen existing health disparities.

RESEARCH GAPS AND FUTURE DIRECTIONS

Despite encouraging technical performance, several critical gaps must be addressed before AI can be integrated into routine hepatitis care. Priority areas include prospective multicenter studies and randomized controlled trials comparing AI-assisted strategies with current guideline-based management. Future research should emphasize external validation across geographically diverse populations, HBV genotypes, and healthcare settings, with federated learning offering a promising approach for collaborative model development while preserving data privacy.

Economic evaluations remain scarce, and future studies should assess the cost-effectiveness of AI-assisted screening, surveillance, and treatment strategies to inform healthcare policy and resource allocation. Greater adherence to standardized reporting frameworks such as TRIPOD+AI will improve transparency, reproducibility, and clinical interpretability of prediction models.^[33]

Future research should also prioritize investigator-led collaborations in low- and middle-income countries and focus on underrepresented populations, including children, individuals with hepatitis D virus (HDV) infection, patients with HBV/HIV or HCV/HIV coinfection, and pregnant women. Expanding evidence in these populations will be essential to ensure equitable implementation of AI and support global hepatitis elimination efforts.^[34]

FIGURE 2: AI Across the Hepatitis Care Continuum Current Applications and Gaps

Note: This figure summarizes the role of artificial intelligence across the viral hepatitis care continuum, highlighting current applications, the strength of available evidence, and key research gaps. It demonstrates that while AI shows promising performance in several stages of care, most applications require further validation, real-world implementation, and evaluation of clinical impact before routine adoption.

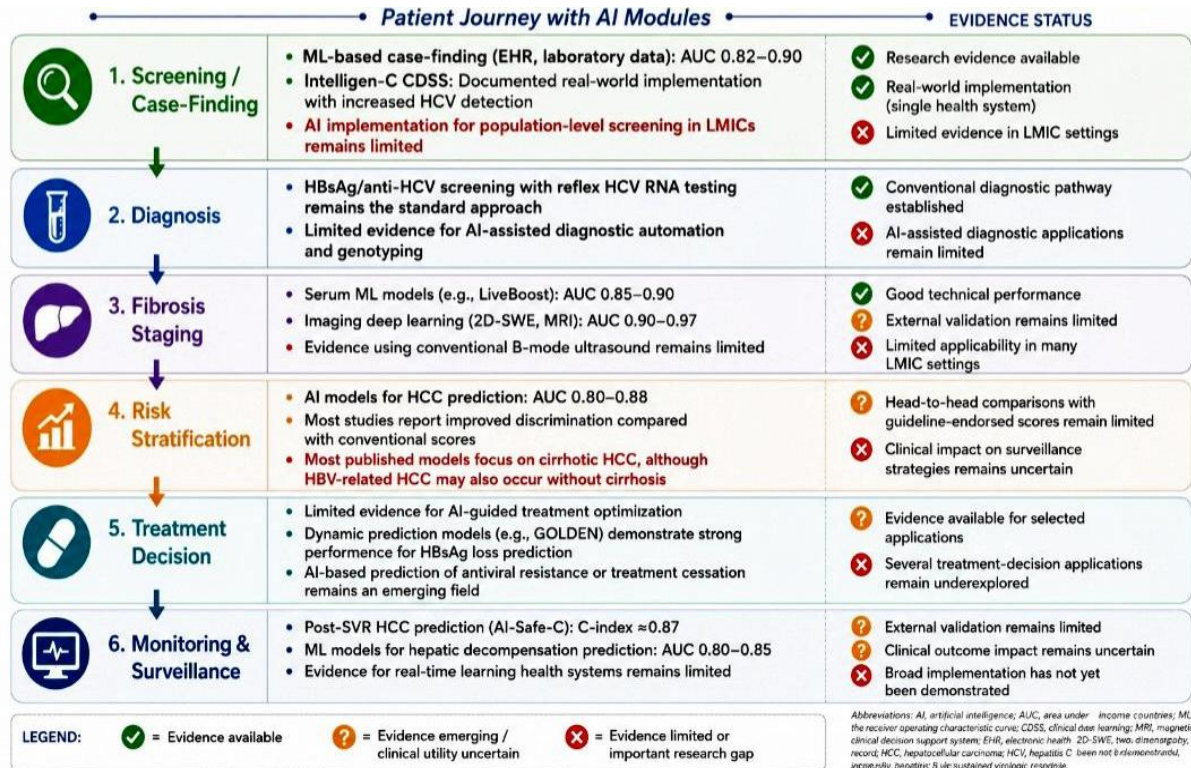


Figure 2: Artificial Intelligence across the viral Hepatitis care continuum: Current evidence with research gaps

AI – Artificial Intelligence

AUC – Area Under the Receiver Operating Characteristic Curve

HBV – Hepatitis B Virus

HCV – Hepatitis C Virus

HCC – Hepatocellular Carcinoma

CDSS – Clinical Decision Support System

RNA – Ribonucleic Acid

ML – Machine Learning

AUC – Area Under the Receiver Operating Characteristic Curve

LMICs – Low- and Middle-Income Countries

WHO-Coordinated Global Hepatitis AI Consortium (Proposed)

Structure:

Coordinating center (WHO or university consortium with LMIC co-leadership)

Working groups: screening, diagnostics, prognosis, treatment, drug discovery, surveillance, equity/implementation

Federated data governance (privacy-preserving; LMIC data sovereignty)

★ Year 1–2 milestones: Governance agreements, prospective multicenter cohorts (3+ genotypes, 2+ continents), RCT initiation, LMIC training

★ Year 3–5 milestones: Federated models trained, external validation published, LMIC-led research emerging

★ Year 5–10 milestones: CDSS deployed in 10+ countries, learning health systems operational, first AI-guided vaccines in trials, measurable cascade-of-care improvements

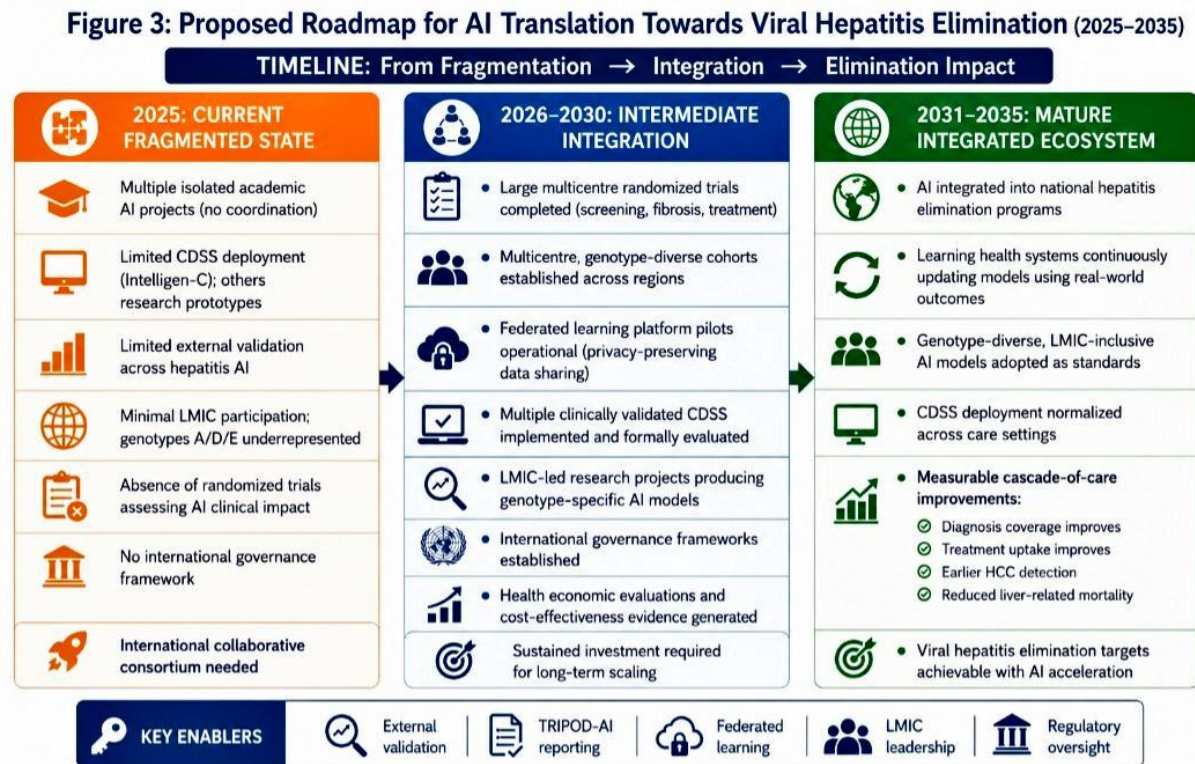


Figure 3: WHO-Coordinated Hepatitis AI Consortium Proposed Roadmap (2025–2035)

Note: This figure presents a proposed roadmap for the implementation of artificial intelligence in viral hepatitis through a World Health Organization (WHO)-coordinated consortium. It illustrates the transition from the current fragmented research landscape (2025) to intermediate integration (2028–2030) and finally to a mature, globally integrated ecosystem (2035), emphasizing coordinated research, multi centre validation, equitable implementation, and measurable improvements in hepatitis elimination outcomes.

Abbreviations

WHO – World Health Organization

AI – Artificial Intelligence

RCT – Randomized Controlled Trial

CDSS – Clinical Decision Support System

LMIC – Low- and Middle-Income Countries

DECIDE-AI – Developmental and Exploratory Clinical Investigation of Decision Support Systems Driven by Artificial Intelligence (reporting guideline for evaluating AI-based clinical decision support systems)

SVR – Sustained Virological Response

HBsAg – Hepatitis B Surface Antigen

HCC – Hepatocellular Carcinoma

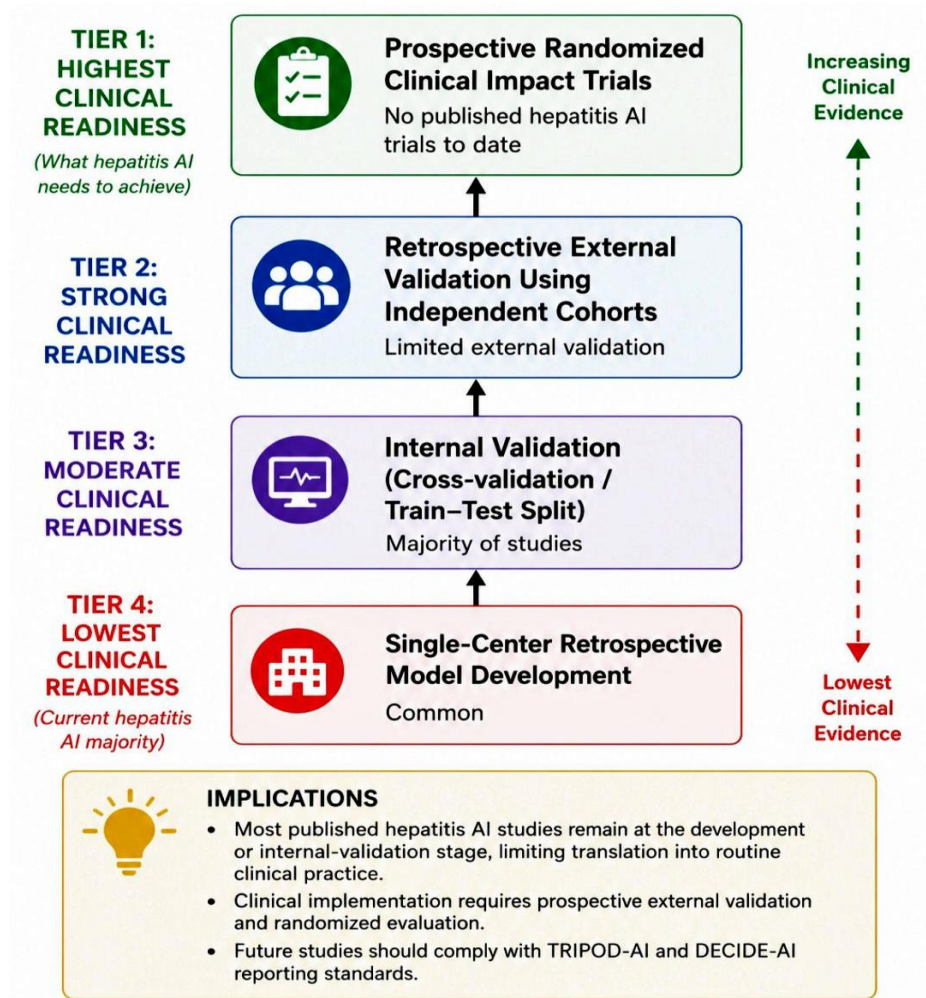


Figure 4: The Methodological Quality Pyramid Evidence Hierarchy for Hepatitis AI

NOTE: The methodological quality pyramid illustrates the hierarchy of evidence supporting artificial intelligence (AI) applications in viral hepatitis, ranging from retrospective single-center studies to randomized controlled trials (RCTs). The figure highlights that most published AI studies remain at lower evidence levels, with limited external validation and clinical applicability. Progression toward higher-quality evidence requires prospective multicenter validation, standardized reporting, and RCTs demonstrating real-world clinical benefit before widespread implementation in hepatitis care.

Abbreviations

- AI – Artificial Intelligence
- RCTs – Randomized Controlled Trials
- TRIPOD-AI – Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis– Artificial Intelligence
- Cross-validation – A statistical technique used to evaluate model performance by repeatedly splitting the dataset into training and validation subsets
- Train–test split – Division of a dataset into separate training and testing sets for model development and evaluation
- Internal validation – Evaluation of a model using the same dataset on which it was developed
- External validation – Evaluation of a model using an independent dataset from a different population or setting
- Prospective study – A study in which participants are followed forward in time after enrollment
- Retrospective study – A study based on analysis of previously collected data
- Multicenter study – A study conducted across multiple hospitals or research institutions
- Cohort – A group of individuals followed or analyzed as part of a research study.

CONCLUSION

Artificial intelligence has demonstrated considerable technical potential in viral hepatitis across screening, diagnosis, prognosis, and treatment support. The Intelligen-C clinical decision support system and the GOLDEN model illustrate how AI can improve case-finding and individualized risk prediction. However, evidence supporting routine clinical implementation remains limited. Real-world deployment is scarce, prospective randomized trials are lacking, external validation remains uncommon, and studies from low- and middle-income countries and diverse HBV genotypes are underrepresented.

The major barriers to clinical translation extend beyond algorithm development. Fragmented datasets, limited external validation, inadequate evaluation of patient-centered outcomes, and evolving regulatory and implementation frameworks continue to hinder widespread adoption.

The central message of this review is that technical performance does not necessarily translate into clinical benefit. While many AI models report excellent discrimination, relatively few demonstrate improvements in clinical decision-making, patient outcomes, or healthcare efficiency. Future progress will depend not only on developing more accurate algorithms but also on rigorous prospective validation, equitable implementation across diverse populations, and integration into evidence-based clinical workflows. Only through these steps can AI meaningfully contribute to the global elimination of viral hepatitis.

To realize the full potential of AI in achieving global hepatitis elimination goals, future efforts should focus on:

1. Strengthening clinical evidence: Conduct prospective multicenter studies, randomized controlled trials where appropriate, independent external validation, model calibration, and transparent reporting in accordance with TRIPOD+AI principles.

2. Promoting equitable implementation: Expand investigator-led research in low- and middle-income countries, validate models across diverse HBV genotypes and populations, routinely assess algorithmic fairness, and evaluate cost-effectiveness in resource-constrained settings.
3. Building robust governance: Foster international collaboration, establish clear regulatory pathways, implement post-deployment monitoring, and promote transparent accountability for AI-assisted clinical decision-making.
4. Supporting sustainable research: Encourage long-term multicenter collaborations with sustained investment to facilitate rigorous validation, implementation, and continuous model improvement.

From Promise To Practice

The potential of AI in viral hepatitis is substantial. Emerging evidence suggests that AI can improve case-finding (as demonstrated by the Intelligen-C clinical decision support system), enhance risk stratification for personalized management, and support future therapeutic discovery. However, technical innovation alone is insufficient. Responsible implementation requires rigorous validation, equitable evaluation across diverse populations, transparent reporting, and integration within evidence-based clinical practice.

The next phase of hepatitis AI research should move beyond prototype development toward prospective validation, broader collaboration with low- and middle-income countries, and integration into national hepatitis elimination strategies. Achieving this will require coordinated efforts from researchers, funding agencies, healthcare systems, policymakers, and international public health organizations. With sustained scientific rigor and equitable implementation, AI has the potential to become a meaningful contributor to global hepatitis elimination.

REFERENCES

1. Polaris Observatory Collaborators. Global prevalence, cascade of care, and prophylaxis coverage of hepatitis B in 2022: a modelling study. *Lancet Gastroenterology & Hepatology*. 2023;8(10):879–907. doi:10.1016/S2468-1253(23)00197-8
2. Polaris Observatory HCV Collaborators. Global change in hepatitis C virus prevalence and cascade of care between 2015 and 2020: a modelling study. *Lancet Gastroenterol Hepatol*. 2022;7(5):396–415. doi:10.1016/S2468-1253(21)00472-6
3. Alberts CJ, Clifford GM, Georges D, et al. Worldwide prevalence of hepatitis B virus and hepatitis C virus among patients with cirrhosis at country, region, and global levels: a systematic review. *Lancet Gastroenterol Hepatol*. 2022;7(8):724–735. doi:10.1016/S2468-1253(22)00050-4
4. World Health Organization. Global Hepatitis Report 2024: Action for Access in Low- and Middle-Income Countries. WHO; 2024. <https://www.who.int/publications/i/item/9789240091672>
5. Liu W, Liu X, Peng M, et al. Artificial intelligence for hepatitis evaluation. *World J Gastroenterol*. 2021;27(34):5715–5726. doi:10.3748/wjg.v27.i34.5715
6. Yip TC-F, Liu S, Wong VW-S. AI in hepatology: revolutionizing the diagnosis and management of liver disease. *J Clin Med*. 2024;13(24):7833. doi:10.3390/jcm13247833

7. EASL Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol.* 2025;82(S1):S1–S58. doi:10.1016/S0168-8278(25)00174-6
8. Ghany MG, Pan CQ, Lok AS, et al. AASLD-IDSA Practice Guideline on treatment of chronic hepatitis B. *Hepatology.* 2026;83(4):974–997. doi:10.1097/HEP.0000000000001549
9. AASLD-IDSA HCV Guidance Panel. Hepatitis C guidance 2023 update: AASLD-IDSA recommendations for testing, managing, and treating hepatitis C virus infection. *Clin Infect Dis.* 2023;77(Suppl 2):S1–S66. doi:10.1093/cid/ciad319
10. Sterling RK, Garcia-Tsao G, Landis CS, et al. AASLD Practice Guideline on imaging-based noninvasive liver disease assessment of hepatic fibrosis and steatosis. *Hepatology.* 2025;81(5):1377–1415. doi:10.1002/hep.33123
11. Papatheodoridis GV, Chan HL, Hansen BE, et al. Page-B and REACH-B score predicting HCC and other serious complications in treated and untreated HBsAg-positive patients: a systematic review. *J Hepatol.* 2021;74(3):500–519. doi:10.1016/j.jhep.2020.09.011
12. Rigg JD, Flannery BA, Kramer JR, et al. Finding undiagnosed patients with hepatitis C virus. *BMJ Health Care Inform.* 2023;30(1):e100651. doi:10.1136/bmjhci-2023-100651
13. Kim SH, Park MK, Lim MJ, et al. Machine learning for predicting hepatitis B or C virus infection in diabetic patients. *Sci Rep.* 2023;13(1):19723. doi:10.1038/s41598-023-45298-1
14. Castro Urda JL, Álvarez M, Cantero H, et al. The efficiency of artificial intelligence for management and clinical decision-making in the identification of patients with hidden HCV infection (Intelligen-C strategy). *Gastroenterol Hepatol.* 2025;48(6):online ahead of print. doi:10.1016/j.gastrohep.2025.02.010
15. Wei R, Wang J, Wang X, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. *EBioMedicine.* 2018;35:124–132. doi:10.1016/j.ebiom.2018.07.041
16. Zhang C, Shu Z, Chen S, et al. A machine learning-based model analysis for serum markers of liver fibrosis in chronic hepatitis B patients. *Sci Rep.* 2024;14:12081. doi:10.1038/s41598-024-63095-8
17. Tapper EB, Lok AS. Use of liver imaging and biopsy in clinical practice. *Hepatology.* 2017;65(3):1028–1037. doi:10.1002/hep.28890
18. Wang K, Teng W, Qin X, 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. *Hepatology.* 2019;70(5):1289–1302. doi:10.1002/hep.30635
19. Cui YY, Zhu RS, Gong XB, et al. Artificial intelligence and machine learning in liver disease and transplantation. *Nat Rev Gastroenterol Hepatol.* 2023;20(12):783–804. doi:10.1038/s41575-023-00805-z
20. Akkari J, Chen A, Garcia-Tsao G, et al. Machine learning-based prediction of decompensation in hepatitis B virus-related cirrhosis. *J Clin Med.* 2025;14(1):181. doi:10.3390/jcm14010181
21. Kaya E, Aksoz M, Sengul Akar S. Comparison of PAGE-B, REACH-B and APRI scores in predicting hepatocellular carcinoma development in chronic hepatitis B patients. *Eur J Gastroenterol Hepatol.* 2023;35(6):653–659. doi:10.1097/MEG.0000000000002575

22. Omata M, Kanda T, Nakamura M, et al. APASL clinical practice guidelines on the management of hepatocellular carcinoma. *Hepatol Int.* 2022;16(2):213–260. doi:10.1007/s12072-022-10293-1
23. Nishwa A, Smith C, O'Brien T, et al. Progression of chronic liver disease to hepatocellular carcinoma: implications for surveillance and management. *BJC Rep.* 2024;2(2):e50. doi:10.1038/s44276-024-00050-0
24. Lin H, Yip TC-F, Lee HW, Meng X, Lai JC-T, Ahn SH, et al. AI-Safe-C score: Assessing liver-related event risks in patients without cirrhosis after successful direct-acting antiviral treatment. *J Hepatol.* 2025;82(3):456–463. doi:10.1016/j.jhep.2024.09.020.
25. Shen H, Albhaisi S, Chiang DJ, et al. Machine learning algorithms for predicting direct-acting antiviral treatment failure in chronic hepatitis C: an HCV-TARGET analysis. *Hepatology.* 2022;76(4):993–1003. doi:10.1002/hep.32347
26. Kong DH, Sarin SK, Kumar M, et al. GOLDEN score: a longitudinal model for predicting hepatitis B surface antigen seroclearance in chronic hepatitis B patients treated with pegylated interferon alpha-2b. *Gut.* 2024;73(10):1725–1734. doi:10.1136/gutjnl-2023-331547
27. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56. doi:10.1038/s41591-018-0300-9
28. Razavi-Shearer D, Gamkrelidze I, Pan Y, et al. Global burden of hepatitis B in 2019: a systematic analysis of the disease epidemiology. *Lancet Gastroenterol Hepatol.* 2021;6(12):963–975. doi:10.1016/S2468-1253(21)00320-5
29. Beamish AJ, Lauffenburger JC, Osborn G, et al. Electronic health record use and clinical decision support systems in low- and middle-income countries. *Lancet Digit Health.* 2023;5(2):e85–e95. doi:10.1016/S2589-7500(22)00246-7
30. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur Urol.* 2021;79(4):495–501. doi:10.1016/j.eururo.2020.12.038
31. Cohen JF, Bossuyt PM. Test evaluation in the era of precision medicine: from performance to patient outcome. *Eur J Clin Invest.* 2021;51(2):e13429. doi:10.1111/eci.13429
32. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *JAMA.* 2019;322(17):1694–1697. doi:10.1001/jama.2019.14635
33. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378.
34. Fonseca SN, Ribeiro PC, Oliveira-Reyes FC, et al. Hepatitis B and hepatitis C virus coinfection in HIV-positive individuals: prevalence, risk factors, and clinical outcomes. *AIDS Rev.* 2023;25(1):13–24. doi:10.24875/AIDSRev.23000007