ASIDE Internal Medicine

Review Article

Artificial Intelligence in the Evaluation of Abnormal Liver Tests and MASLD: Emerging Applications in Risk Stratification and Clinical Decision Support

Ahmed Salman[1,*]

1-Internal Medicine Department, Faculty of Medicine, Cairo University, Egypt

ARTICLE INFO

Article history:

Received 8 May 2026

Received in revised form 23 Jul. 2026

Accepted 29 Jul. 2026

Published 19 Aug. 2026

Keywords:
Artificial intelligence

Abnormal liver tests

Liver fibrosis

Machine learning

Fibrosis risk stratification

ABSTRACT

Background: Abnormal liver tests and metabolic dysfunction-associated steatotic liver disease (MASLD) are common, yet the central clinical task is identifying advanced fibrosis, cirrhosis, or an alternative liver disease that requires timely investigation or referral. **Methods:** We conducted a narrative review of PubMed, Ovid MEDLINE, Embase, Scopus, and Web of Science from database inception through April 2026, supplemented by searches of Google Scholar, reference lists, and relevant society guidelines. **Results:** Artificial intelligence (AI) can integrate clinical, laboratory, longitudinal, imaging, and elastography data to support risk stratification, identify missing investigations, and assist referral workflows. Representative studies reported promising discrimination for selected outcomes, but populations, reference standards, and validation methods were heterogeneous; calibration and external validation were often absent or incompletely reported. Clinical use should begin with standard pattern recognition, exclusion of competing etiologies and urgent red flags, and age-aware interpretation of the fibrosis-4 index before second-line testing or referral. **Conclusions:** AI is an emerging adjunct, not a replacement for clinical judgment. Evidence that it improves outcomes, reduces missed advanced fibrosis, or increases referral efficiency remains limited. Responsible adoption requires transparent models, clinician verification, external validation, calibration, workflow evaluation, and post-deployment monitoring.

1. Introduction

Abnormal liver tests are among the most common biochemical abnormalities seen in primary care, internal medicine, gastroenterology, and hepatology. Mild elevations in alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, or gamma-glutamyl transferase may indicate transient injury, alcohol or medication exposure, viral hepatitis, biliary disease, autoimmune liver disease, metabolic dysfunction-associated steatotic liver disease (MASLD), or less common infiltrative and genetic disorders. The clinical challenge is not simply identifying an abnormal value, but determining who can be reassured and followed up, who requires further investigation, and who should be referred promptly to a specialist [1, 2].

MASLD is a leading cause of abnormal aminotransferases and incidentally detected hepatic steatosis. Its prevalence has risen with obesity, type 2 diabetes, dyslipidemia, hypertension, and other cardiometabolic risk factors. Studies using the earlier term non-alcoholic fatty liver disease (NAFLD) estimate that it affects about one-third of adults worldwide and is substantially more common in people with type 2 diabetes [3, 4]. Most affected individuals do not develop progressive liver disease; fibrosis stage, rather than steatosis severity or aminotransferase level, is the main determinant of liver-related outcomes [5].

MASLD denotes steatotic liver disease associated with cardiometabolic risk and has replaced NAFLD in current nomenclature. Metabolic dysfunction-associated steatohepatitis (MASH) has similarly replaced non-alcoholic steatohepatitis (NASH). Because much of the evidence predates this transition, the original NAFLD/NASH terms are retained when describing individual studies and interpreted in the corresponding MASLD/MASH context.

This review addresses two linked clinical questions. First, abnormal liver tests require broad diagnostic assessment because hepatocellular, cholestatic, mixed, and bilirubin-predominant patterns may reflect viral, alcohol-related, drug-induced, autoimmune, biliary, infiltrative, malignant, or metabolic disease. Second, once MASLD is suspected or established, the priority shifts to fibrosis risk. Artificial intelligence (AI) may support biochemical pattern recognition, identify missing investigations or red flags, and refine fibrosis assessment and referral triage.

Current clinical pathways emphasize structured assessment over indiscriminate testing. On initial clinical assessment, the abnormality should be confirmed, the biochemical pattern of liver injury delineated, alcohol intake and medication exposure reviewed, metabolic risk assessed, and common alternative liver diseases excluded. Simple noninvasive assessment tools, including the fibrosis-4 index, may help identify patients at low risk of advanced fibrosis in suspected MASLD, while patients with indeterminate or high-risk results may need elastography, enhanced liver fibrosis testing,

* Corresponding author: Ahmed Salman, Internal Medicine Department, Faculty of Medicine, Cairo University, Egypt. Email: awea844@gmail.com

Published by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print) 3065-9671, ISSN (Online) 3065-968X.

Citation: Salman A. Artificial Intelligence in the Evaluation of Abnormal Liver Tests and MASLD: Emerging Applications in Risk Stratification and Clinical Decision Support. ASIDE Int Med. 2026;2(4):37-45, doi:10.71079/ASIDE.IM.081926834

additional imaging, or hepatology referral [1, 2, 6]. This staged approach is crucial, since under-recognition and over-investigation have consequences: missed advanced fibrosis may delay intervention, while excessive testing and referral may increase costs and burden specialist services.

AI can combine demographics, liver biochemistry, platelet count, cardiometabolic comorbidities, medication exposure, longitudinal laboratory profiles, imaging, and electronic health record data. Potential applications include fibrosis and cirrhosis risk estimation, imaging interpretation, referral triage, and follow-up planning [7, 8]. These functions are most useful when they complement a clearly defined clinical pathway and produce outputs that clinicians can verify.

This narrative review centers on stable adult outpatient or ambulatory-care patients with persistent or incidentally detected abnormal liver tests and/or suspected MASLD. The described AI-supported pathway is not intended for patients with acute liver failure, severe acute hepatitis, pregnancy-related liver disease, pediatric liver disease, liver transplant recipients, decompensated cirrhosis, or clinically unstable inpatients. In these contexts, urgent standard clinical pathways and specialist evaluation should take precedence over any AI-supported workflow.

AI should not replace clinical judgment or established diagnostic pathways. Many models are retrospective, incompletely validated, and difficult to interpret across different populations. False reassurance could delay recognition of advanced fibrosis, cholestasis, viral or autoimmune disease, or malignancy, whereas poorly calibrated alerts could generate unnecessary testing and referral.

Accordingly, this narrative review evaluates emerging AI applications in abnormal liver tests and MASLD, with emphasis on fibrosis risk, imaging, referral support, evidence limitations, and safeguards for clinical implementation.

2. Methods

We conducted a narrative review rather than a systematic review or meta-analysis. PubMed, Ovid MEDLINE, Embase, Scopus, and Web of Science were searched from database inception through April 2026, with supplementary searches of Google Scholar. Search concepts combined terms for abnormal liver tests or liver chemistries; MASLD, NAFLD, MASH, NASH, steatosis, and liver fibrosis; the fibrosis-4 index (FIB-4) and other noninvasive fibrosis tests; and artificial intelligence, machine learning, deep learning, natural language processing, electronic health records, ultrasound, elastography, clinical decision support, and hepatology referral.

We prioritized society guidelines, primary prediction models and imaging studies, prospective workflow evaluations, and clinically relevant reviews. Reference lists of key articles and relevant society documents were also screened. Studies were selected for their relevance to diagnostic evaluation, fibrosis stratification, referral pathways, model validation, explainability, bias, reporting, safety, or implementation. Because the review was narrative, no protocol registration, duplicate screening, formal risk-of-bias tool, or quantitative synthesis was undertaken.

3. Conventional Evaluation of Abnormal Liver Tests and MASLD: The Clinical Foundation

Evaluation should begin by confirming the abnormality and classifying the biochemical pattern. Predominant alanine or aspartate aminotransferase elevation suggests a hepatocellular pattern; disproportionate alkaline phosphatase elevation, often with increased gamma-glutamyl transferase, suggests cholestasis. Mixed patterns and bilirubin-predominant abnormalities require separate diagnostic approaches. Results should be interpreted against local reference ranges, previous measurements, symptoms, and the pace of change; rapidly progressive abnormalities or impaired synthetic function warrant prompt escalation [1, 2, 9].

The initial assessment should target common, consequential, and modifiable causes. Alcohol exposure, prescribed and nonprescription medications, supplements, viral hepatitis risk, family history, and features of chronic liver disease should be reviewed. Cardiometabolic assessment should include obesity, type 2 diabetes, dyslipidemia, hypertension, obstructive sleep apnea, and cardiovascular risk. Normal or mildly elevated aminotransferases do not exclude clinically important MASLD or advanced fibrosis, particularly in patients with diabetes or multiple metabolic risk factors [10].

Alcohol quantity and pattern must be documented before applying a MASLD pathway. Metabolic dysfunction-associated steatotic liver disease with increased alcohol intake (MetALD) describes steatotic liver disease with metabolic dysfunction and alcohol exposure above MASLD thresholds but below levels generally used for predominantly alcohol-related liver disease. Viral, autoimmune, cholestatic, biliary, genetic, drug-induced, and malignant causes should also be considered. An AI-supported pathway should not assign a MASLD-specific classification when alcohol exposure or competing etiologies remain inadequately characterized [1, 2, 6, 9].

First-line testing generally includes repeat liver biochemistry and a complete blood count with platelet count; albumin and international normalized ratio are added when synthetic dysfunction is suspected. Hepatitis B and C testing, iron studies, and targeted autoimmune or metabolic investigations should follow the clinical context and biochemical pattern. Ultrasound remains appropriate when biliary obstruction, structural disease, focal lesions, or steatosis is suspected, but it is insensitive to mild steatosis and cannot reliably stage fibrosis [11].

Noninvasive fibrosis assessment is now critical to the assessment of MASLD. The fibrosis-4 index (FIB-4), calculated from age, aspartate aminotransferase, alanine aminotransferase, and platelet count, is commonly used as a first-line test in stable adult outpatients. A FIB-4 value <1.3 is generally considered low risk, 1.3–2.67 indeterminate, and >2.67 high risk. For adults aged >65 years, 2.0 rather than 1.3 is commonly used as the low-risk cutoff. FIB-4 has low accuracy in adults aged <35 years; secondary assessment should be considered when metabolic risk or persistently elevated liver chemistries increase clinical concern. FIB-4 should not be used during acute illness. Indeterminate or elevated results should prompt transient elastography, a validated serum fibrosis panel, magnetic resonance elastography where available, or specialist referral according to local pathways and clinical risk [11, 12].

Table 1 summarizes the conventional pathway. In stable adults, FIB-4 <1.3 generally supports monitoring, 1.3–2.67 prompts second-line fibrosis assessment, and >2.67 supports direct specialist referral. For adults aged >65 years, 2.0 replaces 1.3 as the low-risk cutoff. Because accuracy is limited below age 35, secondary assessment should be considered when metabolic risk or persistently abnormal liver tests raise concern. Jaundice, impaired synthetic function, thrombocytopenia, unexplained cholestasis, suspected malignancy, ascites, encephalopathy, gastrointestinal bleeding, or rapidly worsening results require referral regardless of FIB-4 or AI output [11, 12].

Table 1: Conventional evaluation and escalation pathway for stable adult outpatients with abnormal liver tests or suspected MASLD.

Clinical step	Key assessment	Practical purpose	When to escalate
Confirm abnormality	Repeat liver biochemistry and review prior results.	Distinguish transient from persistent abnormality.	Persistent, progressive, or unexplained abnormalities.
Define biochemical pattern	Classify as hepatocellular, cholestatic, mixed, or bilirubin-predominant.	Narrow the differential diagnosis and guide targeted investigation.	Cholestasis, jaundice, or an unexplained mixed pattern.
Review clinical context	Assess alcohol, medications, supplements, viral hepatitis risk, family history, and symptoms or signs of chronic liver disease.	Identify common, reversible, or high-risk causes.	Red flags, high-risk exposure, or suspected chronic liver disease.
Assess metabolic risk	Assess obesity, diabetes, dyslipidemia, hypertension, obstructive sleep apnea, and cardiovascular risk.	Identify suspected MASLD and cardiometabolic risk.	Diabetes, multiple metabolic risk factors, or suspected advanced disease.
Perform first-line tests	Obtain full blood count with platelets, hepatitis B/C testing, iron studies, albumin, and international normalized ratio when indicated.	Exclude common alternative causes and assess disease severity.	Synthetic dysfunction, thrombocytopenia, or diagnostic uncertainty.
Use noninvasive fibrosis assessment.	Calculate FIB-4 from age, AST, ALT, and platelet count in stable adults; do not use during acute illness.	FIB-4 <1.3: low risk. FIB-4 1.3–2.67: indeterminate. FIB-4 >2.67: high risk.	At 1.3–2.67, arrange transient elastography or a serum fibrosis panel. At >2.67, consider direct specialist referral.
Apply age cautions	For adults aged >65 years, use 2.0 rather than 1.3 as the low-risk cutoff. FIB-4 is less accurate in adults aged <35 years.	Reduce age-related false-positive and false-negative classification.	For age >65 years, escalate at ≥2.0. For age <35 years, consider secondary assessment when metabolic risk or elevated liver chemistries are present.
Arrange second-line testing	Use transient elastography, a validated serum fibrosis panel, or advanced imaging where available.	Clarify fibrosis risk after indeterminate or high-risk first-line assessment.	Elevated liver stiffness, discordant results, or persistent clinical concern.
Identify urgent referral criteria.	Check for jaundice, impaired synthetic function, thrombocytopenia, rapidly worsening tests, unexplained cholestasis, suspected malignancy, ascites, encephalopathy, or gastrointestinal bleeding.	Prevent false reassurance from low-risk scores or AI outputs.	Urgent referral regardless of FIB-4, second-line testing, or any AI-generated estimate.

Abbreviations: AI, artificial intelligence; ALT, alanine aminotransferase; AST, aspartate aminotransferase; FIB-4, fibrosis-4 index; INR, international normalized ratio; MASLD, metabolic dysfunction-associated steatotic liver disease. Note: FIB-4 thresholds are intended for stable adult outpatients and should be interpreted with clinical context and local pathways.

4. AI-Assisted Risk Stratification and Fibrosis Prediction

The most credible near-term role for AI is risk stratification rather than autonomous diagnosis. Models can combine demographic, metabolic, laboratory, imaging, and longitudinal data to distinguish patients who need monitoring from those who need second-line testing or specialist assessment [13, 14]. This focus is appropriate in MASLD, where fibrosis risk is more clinically consequential than steatosis alone.

Machine-learning models have been developed to predict steatotic liver disease, steatohepatitis, fibrosis, cirrhosis, and liver-related outcomes. Common inputs include age, body mass index, diabetes, lipid profile, aminotransferases, platelet count, and glycemic measures. Their value depends on whether they outperform established scores, reduce indeterminate classifications, or identify high-risk patients using data already available in routine care [15, 16].

Table 2 summarizes representative primary studies of clinical, electronic health record, imaging, and referral-support applications. Reported discrimination was promising in selected settings, but populations, inputs, reference standards, and validation methods varied substantially. Calibration, independent external validation, and prospective clinical impact were frequently absent or incompletely reported. The evidence therefore supports AI as an adjunct to structured assessment, not a substitute for clinician judgment.

Electronic health record models can track serial liver biochemistry, platelet trends, metabolic risk, imaging reports, medications, and previous fibrosis assessments. This longitudinal view may reveal risk that is not apparent from a single score; for example, progressive thrombocytopenia in a patient with diabetes and obesity may appropriately trigger elastography or hepatology review [17].

Explainability is central to safe use. Clinicians need to know whether an estimate is driven by age, diabetes, platelet count, aminotransferase pattern, body mass index, or another feature, and they need an explicit action pathway. Feature-importance methods and Shapley additive explanations can improve transparency, although they do not by themselves establish clinical benefit [18].

Table 2: Representative primary studies of artificial intelligence applications relevant to MASLD/NAFLD risk stratification, fibrosis assessment, imaging, and hepatology referral support.

Study	Sample and setting	Key inputs and model	Reference standard or comparator	Reported performance	Calibration	External validation	Main limitations
Ghandian et al., 2022 [7]	Sample size not reported in the main article; US national longitudinal EHR repository (>700 sites).	139 demographic, vital-sign, laboratory, diagnosis, medication, and NLP features; XGBoost versus logistic regression and multilayer perceptron.	New ICD-10-coded NASH or fibrosis within 4 years.	Holdout AUROC: 0.792 for NASH and 0.871 for fibrosis; site-held validation: 0.795 and 0.871.	Not reported.	Yes; held-out clinical sites.	Retrospective coded outcomes; proprietary EHR; sample size not reported in main article; no clinical-impact evaluation.
Thrift et al., 2024 [8]	n=344 primary-care patients at one US Veterans Affairs center; 34 with any fibrosis and 15 with significant fibrosis.	Age, sex, diabetes, hypertension, BMI, FIB-4, and lipid variables; logistic regression and random forest.	FibroScan liver stiffness >7 kPa (any fibrosis) or >8 kPa (significant fibrosis); compared with standard thresholds.	Any fibrosis: AUC 0.75 (95% CI, 0.67–0.84), NPV 91.5%, PPV 40%.	Not reported.	No; internal 80/20 split only.	Few positive events; single-center Veterans Affairs cohort; limited generalizability.
Mamandipoor et al., 2023 [16]	n=5,834 for steatosis and n=1,240 for fibrosis; colorectal screening cohort with temporal prospective evaluation.	Clinical and laboratory variables plus eight dietary variables; XGBoost, feed-forward neural network, and logistic regression.	Ultrasound echogenicity for steatosis; transient elastography ≥8 kPa for fibrosis; compared with FIB-4.	AUC 0.87 for steatosis and 0.75 for fibrosis; FIB-4 fibrosis AUC 0.61.	Calibration curves and Brier scores reported; 0.15 for the selected steatosis model and 0.10 for the selected fibrosis model.	Temporal prospective evaluation in the same cohort; no independent external cohort.	Screening cohort; non-histologic standards; fibrosis prevalence 7%; sex-related performance differences.
Kalka et al., 2025 [17]	2,255,580 Israeli EHR observations; 11,337 incident cirrhosis cases; prospective referral sample n=103.	Routine demographic and laboratory variables; XGBoost risk score.	Five-year coded cirrhosis in EHR development; transient elastography >12 kPa in prospective comparison; comparator FIB-4.	Temporal AUC 0.79 (development 0.81) versus FIB-4 0.71; advanced fibrosis in 21/76 (27.6%) model-selected versus 1/27 (3.7%) FIB-4-selected participants.	Not reported.	Yes; prospective single-center referral comparison.	Low attendance and selection bias; differing development and validation outcomes; no patient-outcome evaluation.
Njei et al., 2024 [18]	NHANES 2017–March 2020: n=5,281 adults with valid elastography; n=5,156 complete-case analysis.	ALT, GGT, platelets, age, and BMI; explainable XGBoost with SHAP.	FAST score ≥0.35, derived from liver stiffness, controlled attenuation parameter, and AST.	Reported AUC 0.95, sensitivity 0.82, specificity 0.91, and accuracy 0.90.	Not reported.	No; internal holdout testing only.	Cross-sectional dataset; derived non-biopsy target; possible label incorporation; no clinical impact or external validation.

continued on the next page

Table 2 (continued)

Study	Sample and setting	Key inputs and model	Reference standard or comparator	Reported performance	Calibration	External validation	Main limitations
Kosick et al., 2025 [19]	n=457 biopsy-proven MASLD patients at two Canadian centers; 29 liver-related outcomes; median follow-up 71 months.	Clinical data, laboratories, standard noninvasive tests, and B-mode ultrasound; random forest with repeated 10× fivefold cross-validation.	Clinical Research Network histology for F3–F4/fibrotic MASH; chart-adjudicated liver outcomes.	Liver-related outcome AUC 0.87 (sensitivity 0.77, specificity 0.78); decompensation AUC 0.90 (0.88, 0.78); F3–F4 AUC 0.83 (0.79, 0.70).	Not reported.	No; internal repeated cross-validation only.	Retrospective selected biopsy cohort; few events; wide uncertainty for rare outcomes; imaging added limited value.
Destrempes et al., 2022 [20]	n=82 patients with mixed chronic liver disease at two centers; 44 had NAFLD/NASH.	Quantitative ultrasound, homodyned-K parameters, attenuation, and point shear-wave elastography; random forest with bootstrap validation.	Liver biopsy for steatosis, inflammation, and fibrosis.	Best combined steatosis AUCs 0.90, 0.81, and 0.78 for ≥S1, ≥S2, and S3; fibrosis AUCs 0.72–0.77.	Not reported.	No; bootstrap internal validation only.	Small heterogeneous biopsy-selected cohort; device and acquisition dependence; no independent validation.
Shroff et al., 2026 [21]	n=50 hepatology referral records assessed by two providers at one center.	Prompt-engineered large language model extracting predefined referral and triage elements.	Original referral documents and clinician review; workflow-time comparison.	Median accuracy 94.6%; median summary length 2 versus 23 pages; triage time 37.2 versus 94.2 seconds (60% reduction).	Not applicable.	No; single-center prospective workflow feasibility study.	Small sample and two reviewers; time-saving endpoint rather than triage accuracy or patient outcomes; clinician verification required.

Abbreviations: AI, artificial intelligence; ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUC/AUROC, area under the receiver operating characteristic curve; BMI, body mass index; EHR, electronic health record; FAST, FibroScan–AST; FIB-4, fibrosis-4 index; GGT, gamma-glutamyl transferase; MASLD, metabolic dysfunction-associated steatotic liver disease; MASH, metabolic dysfunction-associated steatohepatitis; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; NLP, natural language processing; NPV, negative predictive value; PPV, positive predictive value; SHAP, Shapley additive explanations. “Not reported” indicates that the item was not presented in the article or accessible in the main report.

At the health-system level, AI could calculate basic fibrosis indices, detect missing first-line investigations, identify warning signs, and prompt appropriate second-line testing. Such automation may improve consistency across primary care, internal medicine, gastroenterology, and hepatology, but its effect on missed disease, referral volume, workload, and outcomes requires prospective evaluation [17, 18].

Most fibrosis models remain developmental. Retrospective cohorts, selected populations, imperfect labels, and internal validation can inflate apparent performance. Before clinical use, a model should demonstrate calibration, external validity, interpretability, workflow compatibility, and net clinical benefit [13, 17, 18].

5. AI in Imaging, Referral Triage, and Clinical Decision Support

AI applications in abnormal liver tests and MASLD extend beyond blood-based risk prediction. Imaging represents a major area of progress, as ultrasound, elastography, computed tomography, and magnetic resonance imaging are commonly used to evaluate steatosis, fibrosis, cirrhosis, portal hypertension, and focal liver lesions. Conventional imaging interpretation can be limited by operator dependence, subjective estimation, inconsistent reporting language, and challenges in the early detection of fibrosis. AI-based image analysis could assist by extracting quantitative characteristics, standardizing interpretation, and highlighting subtle patterns that may not be consistently recognized by visual assessment alone [22, 23].

Ultrasound is especially important because it is widely available and commonly used as an initial investigation in patients with abnormal liver tests or suspected steatotic liver disease. Deep-learning and machine-learning approaches have been investigated for the automated detection and grading of hepatic steatosis from ultrasound images, aiming to reduce subjectivity and improve reproducibility. Some models integrate B-mode ultrasound features with clinical and biochemical parameters to improve prediction of advanced fibrosis or liver-related prognostic outcomes. However, ultrasound-based AI should be interpreted with caution, as performance may be affected by image quality, obesity, equipment type, acquisition protocol, and disease spectrum [19, 24].

AI integration may also be useful in elastography and magnetic resonance-based methods. Transient elastography, shear-wave elastography, magnetic resonance elastography, and MRI-proton density fat fraction provide more quantitative assessment than conventional ultrasound, but findings still require clinical interpretation. AI may potentially help by combining stiffness values, fat quantification, laboratory markers, metabolic risk factors, and long-term trends into a more individualized risk estimate. In the future, this may help differentiate patients needing routine follow-up from those requiring hepatology review, cirrhosis surveillance, or assessment for competing causes of liver disease [19, 20].

Natural language processing and large language models (LLMs) may support referral triage by extracting laboratory trends, imaging findings, medication exposure, comorbidities, and missing investigations from referral documents. Structured summaries could reduce review time, but every output should remain source-linked, auditable, and clinician-verified because omissions, chronology errors, hallucinations, and unsupported conclusions can alter triage or follow-up [21].

The clinical role of AI varies with context. For abnormal liver tests, AI may assist with biochemical pattern recognition and flag missing first-line investigations, urgent features, or non-MASLD causes, including viral hepatitis, alcohol-related liver disease, drug-induced liver injury, autoimmune or cholestatic disease, biliary obstruction, infiltrative disease, and malignancy. In suspected or established MASLD, it may support fibrosis-risk stratification by computing or verifying FIB-4, integrating elastography and imaging results, and suggesting monitoring, second-line assessment, or referral. AI must not override jaundice, synthetic dysfunction, thrombocytopenia, unexplained cholestasis, suspected malignancy, ascites, encephalopathy, gastrointestinal bleeding, or other urgent features; these require escalation regardless of an AI-generated estimate. **Figure 1** presents an AI-supported pathway for abnormal liver tests and suspected MASLD [21, 22].

6. Practical Limitations, Safety Concerns, and Implementation Challenges

Clinical utility depends on performance outside the development dataset. Liver-disease models are often retrospective, single-system, or derived from selected populations; transportability may be affected by disease prevalence, metabolic risk, alcohol and viral hepatitis burden, test availability, and referral thresholds. Discrimination alone is insufficient. Calibration, false-negative and false-positive rates, downstream testing, referral burden, clinician adherence, safety events, patient outcomes, and cost-effectiveness should be evaluated prospectively [25, 26].

Bias is a major concern. Training data may underrepresent some groups or encode unequal access to laboratory testing, imaging, elastography, biopsy, or specialist care. In MASLD, body composition, diabetes prevalence, socioeconomic conditions, ethnicity, coding practices, and longitudinal data availability may all affect performance. Models should therefore report subgroup performance and should not be assumed to generalize from tertiary cohorts to primary care [26, 27].

Data quality and interoperability are equally important. Alcohol exposure, medications and supplements, comorbidities, fibrosis testing, imaging findings, laboratory units, reference ranges, and diagnostic codes may be missing or inconsistent across health systems. Data harmonization, unit standardization, coding validation, and transparent handling of missing values are prerequisites for reliable electronic health record models [8, 17, 25, 26].

Outputs should be interpretable and linked to a defined action. A prediction driven by thrombocytopenia, rising bilirubin, or low albumin carries different implications from one driven mainly by age or mild aminotransferase elevation. Unsupported alerts risk alert fatigue, false reassurance, and unnecessary testing; early clinical evaluation should therefore examine usability, human factors, failure modes, and safety [25, 28].

Validation should match the intended use. Population-level prediction models require calibration across health systems and subgroups [25, 26]; imaging algorithms require testing across devices, acquisition protocols, operators, and image quality [20, 22, 23]; and language tools require verification against source records [21]. Generative AI requires additional safeguards against omissions, hallucinations, and unsupported recommendations [29].

The reference standard also matters. Liver biopsy is invasive, affected by sampling variability, and uncommon in routine MASLD care. Transient elastography, serum fibrosis panels, and magnetic resonance elastography are more scalable but have gray zones and technical limitations. Studies should therefore state the reference

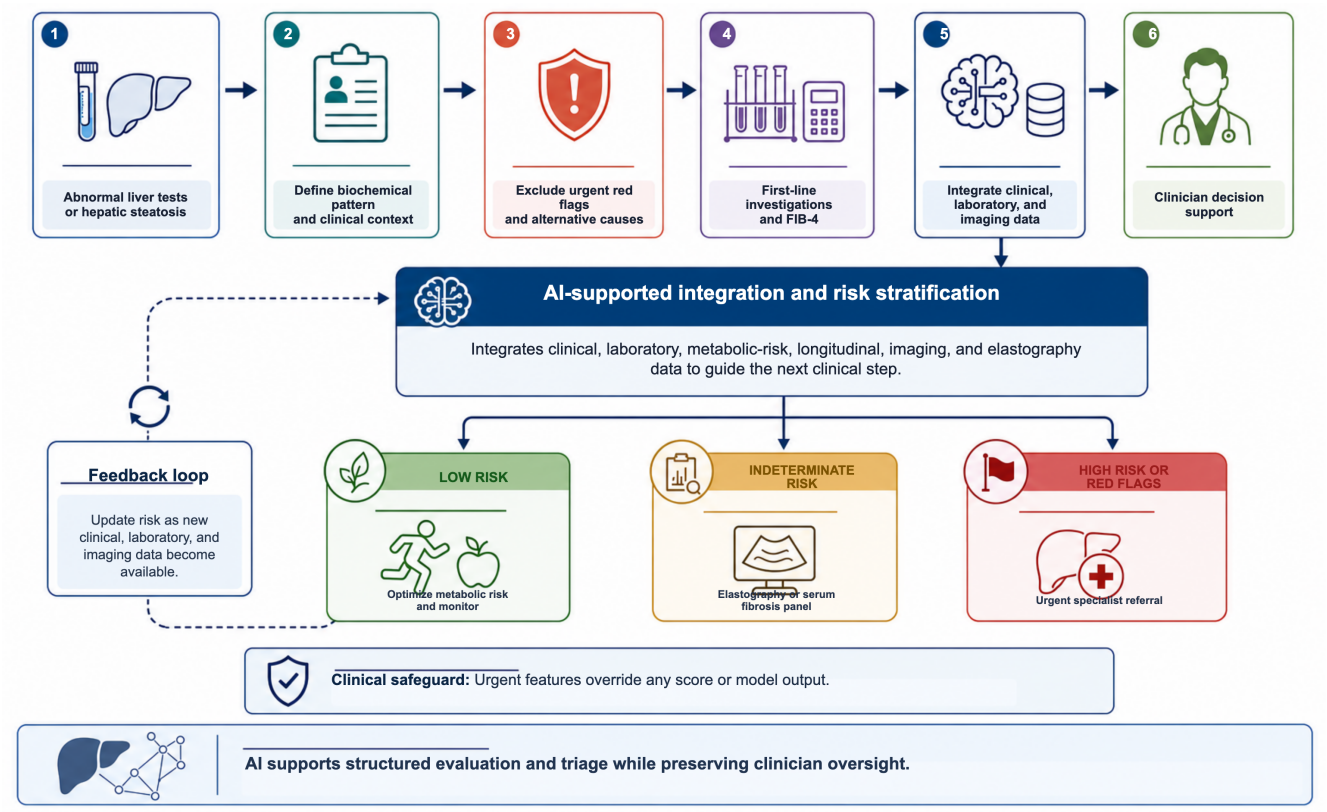

Figure 1: Proposed AI-supported clinical pathway for abnormal liver tests and suspected metabolic dysfunction-associated steatotic liver disease (MASLD). The pathway applies to stable adult outpatients. Conventional clinical assessment defines the biochemical pattern, excludes urgent or competing diagnoses, and establishes first-line fibrosis risk. AI may then integrate clinical, laboratory, metabolic, longitudinal, imaging, and elastography data to support monitoring, second-line assessment, or specialist referral. Urgent features override any score or model output.

standard, outcome definition, follow-up interval, and intended clinical use [30].

Implementation must also address privacy, security, cost, medicolegal responsibility, and workflow design. Tools should appear at a meaningful decision point, present verifiable evidence, and preserve clinician responsibility for the final judgment. Human oversight, transparency, accountability, and patient autonomy remain essential for generative and predictive systems alike [29].

Post-deployment monitoring is critical, as AI performance can differ after implementation. Even externally validated models can decay over time due to variations in patient mix, laboratory assays, reference ranges, electronic health record documentation, imaging protocols, coding, referral behavior, or treatment pathways. Implementation may therefore involve model versioning, drift checking, recalibration as needed, periodic audit of performance and safety results, procedures for reporting errors and unexpected outputs, and clear assignment of clinical accountability. AI tools must also be reviewed whenever the input data pipeline, clinical workflow, or intended use changes because these changes may affect model performance and patient safety [25, 26, 28].

Reporting should follow the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis plus Artificial Intelligence (TRIPOD+AI) statement. Early-stage clinical evaluations should address usability, workflow, human factors, failure modes, and safety, and prospective trials should report the model version, inputs, intended use, human-model interaction, error handling, and monitoring [25, 28, 31].

7. Conclusion

AI may strengthen the evaluation of abnormal liver tests and suspected MASLD by supporting fibrosis risk stratification, imaging interpretation, referral triage, and clinical decision support. Its most useful role is likely to be identifying patients who need second-line fibrosis assessment or specialist review while helping clinicians recognize missing investigations, longitudinal change, and urgent features.

Current evidence is heterogeneous and is dominated by retrospective development, internal validation, selected cohorts, and early workflow studies. Few studies have shown better patient outcomes, fewer missed cases of advanced fibrosis, more efficient referral, or cost-effectiveness. AI should therefore augment, not replace, structured assessment and established fibrosis testing. Routine adoption requires external validation, calibration, explainability, bias assessment, secure workflow integration, and prospective evidence of clinical benefit.

Conflicts of Interest

The author declares no conflicts of interest relevant to this manuscript.

Funding Source

No specific funding was received for this work.

Acknowledgments

None.

Institutional Review Board (IRB)

Not applicable.

Large Language Model

OpenAI Codex was used solely for language editing and grammatical correction during manuscript preparation. The author reviewed and verified the final text and takes full responsibility for the manuscript.

Author Contributions

AS conceived the review, conducted the literature search and interpretation, and drafted and critically revised the manuscript. The author reviewed and approved the final manuscript and accepts accountability for the work.

Data Availability

No new datasets were generated or analyzed for this narrative review. All data discussed are derived from previously published studies cited in the manuscript.

References

1. Newsome PN, Cramb R, Davison SM, et al. Guidelines on the management of abnormal liver blood tests. Gut. 2018;67(1):6-19. [PMID: 29122851, doi:10.1136/gutjnl-2017-314924].
2. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology. 2023;77(5):1797-835. [PMID: 36727674, doi:10.1097/HEP.0000000000000323].
3. Younossi ZM, Golabi P, Paik JM, et al. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. Hepatology. 2023;77(4):1335-47. [PMID: 36626630, doi:10.1097/HEP.0000000000000004].
4. Younossi ZM, Golabi P, Price JK, et al. The Global Epidemiology of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis Among Patients With Type 2 Diabetes. Clin Gastroenterol Hepatol. 2024;22(10):1999-2010. [PMID: 38521116, doi:10.1016/j.cgh.2024.03.006].
5. Dulai PS, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology. 2017;65(5):1557-65. [PMID: 28130788, doi:10.1002/hep.29085].
6. European Association for the Study of the Liver, European Association for the Study of Diabetes, European Association for the Study of Obesity. EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD): Executive Summary. Diabetologia. 2024;67(11):2375-92. [PMID: 38869512, doi:10.1007/s00125-024-06196-3].
7. Ghandian S, Thapa R, Garikipati A, et al. Machine learning to predict progression of non-alcoholic fatty liver to non-alcoholic steatohepatitis or fibrosis. JGH Open. 2022;6(3):196-204. [PMID: 35355667, doi:10.1002/jgh3.12716].
8. Thrift AP, Nguyen Wenker TH, Godwin K, et al. An Electronic Health Record Model for Predicting Risk of Hepatic Fibrosis in Primary Care Patients. Dig Dis Sci. 2024;69(7):2430-6. [PMID: 38700632, doi:10.1007/s10620-024-08437-2].
9. Kwo PY, Cohen SM, Lim JK. ACG Clinical Guideline: Evaluation of Abnormal Liver Chemistries. Am J Gastroenterol. 2017;112(1):18-35. [PMID: 27995906, doi:10.1038/ajg.2016.517].
10. Long MT, Noureddin M, Lim JK. AGA Clinical Practice Update: Diagnosis and Management of Nonalcoholic Fatty Liver Disease in Lean Individuals: Expert Review. Gastroenterology. 2022;163(3):764-74. [PMID: 35842345, doi:10.1053/j.gastro.2022.06.023].
11. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol. 2021;75(3):659-89. [PMID: 34166721, doi:10.1016/j.jhep.2021.05.025].
12. Wattacheril JJ, Abdelmalek MF, Lim JK, et al. AGA Clinical Practice Update on the Role of Noninvasive Biomarkers in the Evaluation and Management of Nonalcoholic Fatty Liver Disease: Expert Review. Gastroenterology. 2023;165(4):1080-8. [PMID: 37542503, doi:10.1053/j.gastro.2023.06.013].
13. Wong GL, Yuen PC, Ma AJ, et al. Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. J Gastroenterol Hepatol. 2021;36(3):543-50. [PMID: 33709607, doi:10.1111/jgh.15385].
14. Aggarwal P, Alkhouri N. Artificial Intelligence in Nonalcoholic Fatty Liver Disease: A New Frontier in Diagnosis and Treatment. Clin Liver Dis (Hoboken). 2021;17(6):392-7. [PMID: 34386201, PMCID: PMC8340349, doi:10.1002/cld.1071].
15. Goldman O, Ben-Assuli O, Rogowski O, et al. Non-alcoholic Fatty Liver and Liver Fibrosis Predictive Analytics: Risk Prediction and Machine Learning Techniques for Improved Preventive Medicine. J Med Syst. 2021;45(2):22. [PMID: 33426569, doi:10.1007/s10916-020-01693-5].
16. Mamandipoor B, Wernly S, Semmler G, et al. Machine learning models predict liver steatosis but not liver fibrosis in a prospective cohort study. Clin Res Hepatol Gastroenterol. 2023;47(7):102181. [PMID: 37467893, doi:10.1016/j.clinre.2023.102181].
17. Kalka IN, Hazzan R, Yacovzada NS, et al. Fibro: a machine learning risk score for advanced liver fibrosis in the general population using Israeli electronic health records. Sci Rep. 2025;15(1):32035. [PMID: 40887472, doi:10.1038/s41598-025-17534-9].
18. Njei B, Osta E, Njei N, et al. An explainable machine learning model for prediction of high-risk nonalcoholic steatohepatitis. Sci Rep. 2024;14(1):8589. [PMID: 38615137, doi:10.1038/s41598-024-59183-4].
19. Kosick HM, McIntosh C, Bera C, et al. Machine learning models using non-invasive tests & B-mode ultrasound to predict liver-related outcomes in metabolic dysfunction-associated steatotic liver disease. Sci Rep. 2025;15(1):24579. [PMID: 40628914, doi:10.1038/s41598-025-09288-1].
20. Destrempes F, Gesnik M, Chayer B, et al. Quantitative ultrasound, elastography, and machine learning for assessment of steatosis, inflammation, and fibrosis in chronic liver disease. PLoS One. 2022;17(1):e0262291. [PMID: 35085294, doi:10.1371/journal.pone.0262291].
21. Shroff H, Shankar A, Baron A, et al. A Large Language Model Assistant for Summarizing Hepatology Referral Documents. Am J Gastroenterol. 2026;121(4):925-31. [PMID: 41504326, doi:10.14309/ajg.0000000000003905].
22. Nam D, Chapiro J, Paradis V, et al. Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction. JHEP Rep. 2022;4(4):100443. [PMID: 35243281, doi:10.1016/j.jhepr.2022.100443].
23. Li J, Chen J, Zeng X, et al. Update of machine learning for ultrasound diagnosis of metabolic dysfunction-associated steatotic liver disease: a bright future for deep learning. PeerJ. 2025;13:e19645. [PMID: 40611943, doi:10.7717/peerj.19645].
24. Ahn JC, Connell A, Simonetto DA, et al. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. Hepatology. 2021;73(6):2546-63. [PMID: 33098140, doi:10.1002/hep.31603].
25. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. [PMID: 38626948, doi:10.1136/bmj-2023-078378].
26. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med.

2019;17(1):195. [PMID: 31665002, doi:10.1186/s12916-019-1426-2].
27. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. [PMID: 31649194, doi:10.1126/science.aax2342].
28. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904. [PMID: 35584845, doi:10.1136/bmj-2022-070904].
29. Ning Y, Teixayavong S, Shang Y, et al. Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. Lancet Digit Health. 2024;6(11):e848-56. [PMID: 39294061, doi:10.1016/S2589-7500(24)00143-2].
30. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology. 2005;128(7):1898-906. [PMID: 15940625, doi:10.1053/j.gastro.2005.03.084].
31. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364-74. [PMID: 32908283, doi:10.1038/s41591-020-1034-x].