Research Article
E-ISSN : 3048-1368
Volume 18, Issue No. 03. 2025
P-ISSN : 3048-1376

The International Journal of Medical Science and Health Research

Predicting the Progression of Non-Alcoholic Fatty Liver Disease Using Machine Learning and Clinical Laboratory Parameters: A Systematic Review

[1] Aila Mustofa, [2] Catherine Halim

[1] General Practitioner, Kimia Farma Clinical Laboratory, Surakarta City, Central Java, Indonesia

[2] Clinical Pathology Consultant, , Kimia Farma Clinical Laboratory, Surakarta City, Central Java, Indonesia

Corresponding Email : ailamustofa@gmail.com

Article History :

Received date : 2025/07/12

Revised date : 2025/08/09

Accepted date : 2025/09/14

Published date : 2025/10/26

E-ISSN :
ISSN 3048-1368
9 773048 136000

P-ISSN
ISSN 3048-1376
9 773048 137007

ABSTRACT

Introduction: Non-alcoholic fatty liver disease (NAFLD) is a growing global health crisis, with a significant proportion of patients progressing to severe liver pathologies, including non-alcoholic steatohepatitis (NASH), fibrosis, and cirrhosis. The limitations and risks of invasive liver biopsy, the current gold standard for diagnosis, necessitate the development of accurate, non-invasive tools for risk stratification and disease monitoring. Machine learning (ML) models, which utilize routinely collected clinical laboratory data, have emerged as a promising and scalable solution for predicting disease progression. This review synthesizes the current evidence on the efficacy of these models.

Methods: A systematic literature search was conducted across PubMed, Google Scholar, Semanthic Scholar, Springer, Wiley

Online Library databases in accordance with PRISMA guidelines. The search included studies that developed or validated ML models to predict NAFLD progression (to NASH, significant fibrosis, or advanced fibrosis using clinical laboratory parameters as primary predictors. Data on study design, population characteristics, ML algorithms, key predictors, and a full spectrum of performance metrics were extracted. The methodological quality of each study was rigorously assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST).

Results: Sixteen studies met the inclusion criteria, encompassing a diverse range of populations and model architectures. The primary outcomes predicted were progression to NASH, significant fibrosis, and advanced fibrosis. Ensemble ML models, particularly eXtreme Gradient Boosting (XGBoost) and Random Forest (RF), consistently demonstrated superior predictive performance over traditional statistical models and other ML algorithms. For the critical endpoint of advanced fibrosis, these models frequently achieved Area Under the Receiver Operating Characteristic (AUROC) values exceeding 0.85 and, in some cases, approaching 0.92. A core set of laboratory parameters—including alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), platelet count, triglycerides, and glycated hemoglobin (HbA1c)—were consistently identified as the most important predictors across multiple models, reflecting their central role in the pathophysiology of NAFLD.

Discussion: The evidence strongly indicates that ML models can effectively integrate complex, non-linear patterns from standard

laboratory tests to generate a "digital signature" of NAFLD pathophysiology, enabling more accurate and individualized risk stratification than traditional scoring systems. These models hold significant potential for clinical application, from facilitating early identification of high-risk individuals in primary care settings to improving the efficiency of patient enrollment in clinical trials for emerging NASH therapies. However, the predominance of retrospective study designs, a lack of consistent external validation, and issues with model interpretability are key limitations of the current evidence base that must be addressed.

Conclusion: Machine learning models based on clinical laboratory parameters are powerful non-invasive tools for predicting NAFLD progression. Their high accuracy and reliance on readily available data position them as a transformative technology in hepatology. Future research must prioritize prospective validation in diverse, real-world clinical settings and focus on developing interpretable, longitudinally-informed models to facilitate their responsible and effective integration into routine clinical practice.

Keywords: Non-Alcoholic Fatty Liver Disease (NAFLD); Non-Alcoholic Steatohepatitis (NASH); Liver Fibrosis; Machine Learning; Predictive Modeling; Clinical Laboratory Parameters; Systematic Review.

INTRODUCTION

The Clinical Spectrum and Global Burden of NAFLD

Non-alcoholic fatty liver disease (NAFLD) represents a spectrum of conditions initiated by hepatic steatosis (fatty liver) in the absence of significant alcohol consumption (Ghandian et al., 2022; Souza-Guedes et al., 2022). While simple steatosis, or non-alcoholic fatty liver (NAFL), is often considered benign, a substantial subset of patients, estimated at 20% to 30%, develop non-alcoholic steatohepatitis (NASH). NASH is a more aggressive form of the disease characterized by necroinflammation and hepatocyte ballooning, which acts as a major catalyst for the progression to liver fibrosis, cirrhosis, and hepatocellular carcinoma (HCC) (Zamanian et al., 2024; Ghandian et al., 2022; Lee et al., 2023).

The global prevalence of NAFLD has reached epidemic proportions, affecting an estimated 25% to 30% of the world's population and establishing it as the most common chronic liver disease (Zamanian et al., 2024; Pan et al., 2025; Ji et al., 2022). This escalating prevalence poses a significant public health and economic challenge. Projections suggest that NAFLD is on track to become the leading indication for liver transplantation in the United States by 2030 (Xiao, An, et al., 2022). The economic impact is equally staggering, with estimated annual costs associated with NAFLD care exceeding 103 billion in the United States alone (Ghandian et al., 2022). This context underscores the urgent need for effective strategies to manage this widespread and costly disease.

The Imperative for Non-Invasive Diagnostics in Clinical Pathology

The definitive diagnosis of NASH and the accurate staging of liver fibrosis currently rely on liver biopsy, which remains the gold standard (Ghandian et al., 2022; Anstee et al., 2023; Canbay et al., 2019). However, the utility of liver biopsy as a screening or monitoring tool is severely constrained. The procedure is invasive, costly, carries inherent risks such as bleeding and pain, and is susceptible to significant sampling error due to the heterogeneous nature of liver pathology

(Corey et al., 2021; Sabet Sarvestany et al., 2022). Furthermore, the histopathological interpretation is subject to considerable inter- and intra-observer variability among pathologists, which can impact diagnostic consistency and the assessment of treatment response in clinical trials (Taylor-Weiner et al., 2021; Ratziu et al., 2024).

These limitations create a critical bottleneck in patient care, hindering the early identification of individuals at high risk of disease progression. Consequently, there is a pressing clinical demand for accurate, accessible, cost-effective, and non-invasive tools for risk stratification (Ghandian et al., 2022; Souza-Guedes et al., 2022; Aggarwal and McCullough, 2021). This review focuses specifically on predictive models derived from clinical pathology laboratory parameters, as these data are routinely collected during patient care, are universally available, and represent the most scalable data source for widespread screening initiatives (Zamanian et al., 2024; Xiao, Yip, et al., 2022).

The Emergence of Machine Learning in Predictive Hepatology

Traditional non-invasive tests (NITs), such as the Fibrosis-4 (FIB-4) index and the NAFLD Fibrosis Score (NFS), are based on logistic regression models and have been valuable first-line tools. However, their clinical utility is often hampered by a large proportion of patients falling into an "indeterminate" risk category, necessitating further, more expensive testing (Zheng et al., 2024; Charu et al., 2024; Aggarwal and McCullough, 2021).

Machine learning (ML) represents a significant advancement over these traditional scoring systems. ML encompasses a suite of algorithms, such as Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), and neural networks, that are designed to identify complex, non-linear patterns and interactions within large, high-dimensional datasets (Souza-Guedes et al., 2022; Corey et al., 2021; Lee et al., 2023). Unlike simple regression, these models can simultaneously analyze dozens of clinical and laboratory variables to generate more robust and individualized risk predictions. This capability has demonstrated superior performance in

identifying clinically significant stages of NAFLD compared to conventional NITs, offering a more powerful approach to predictive hepatology (Chang et al., 2023; Lee et al., 2023).

Rationale, Research Gap, and Novelty

The rationale for this systematic review is to synthesize the growing body of evidence on ML models for NAFLD progression, with a specific focus on those utilizing clinical laboratory parameters. While several reviews have broadly covered the application of artificial intelligence in hepatology or focused on imaging-based models, a critical appraisal of models that leverage the most widely accessible data source in healthcare—routine blood tests—is currently lacking (Xiao, An, et al., 2022; Anstee et al., 2023).

A key research gap that has been addressed by recent studies is the shift from diagnostic to prognostic modeling. Early ML applications focused on the cross-sectional diagnosis of whether a patient currently has NAFLD or NASH (Xiao, Yip, et al., 2022). However, the field has evolved, with more sophisticated studies now aiming to predict the longitudinal risk of *progression* from NAFL to NASH or advanced fibrosis over a defined period, such as four years (Ghandian et al., 2022). This evolution from a static diagnostic question to a dynamic prognostic one is more aligned with the goals of preventative medicine: to identify and intervene in high-risk individuals *before* they develop irreversible liver damage.

The novelty of this systematic review lies in its focused synthesis and critique of ML models that predict NAFLD progression (NASH, \ge F2, \ge F3/F4) based primarily on the predictive power of routine clinical laboratory parameters. It aims to identify which laboratory markers are consistently most important and to compare the efficacy of different ML algorithms for this specific and clinically vital task.

METHODS

Search Strategy and Selection Criteria

This systematic review was conducted and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A comprehensive literature search was performed across the PubMed, Google Scholar, Semanthic Scholar, Springer, Wiley Online Library databases for articles published from inception through July 2024.

Studies were selected based on the following inclusion criteria: (1) original research articles; (2) studies that developed and/or validated a machine learning model; (3) the study population consisted of patients with NAFLD; (4) the prediction outcome was a measure of disease progression, such as the development of NASH, significant fibrosis (fibrosis stage \ge F2), or advanced fibrosis (fibrosis stage \ge F3 or F4); (5) the model inputs included routinely collected clinical laboratory parameters; and (6) quantitative performance metrics (e.g., AUROC, accuracy) were reported. Exclusion criteria were: (1) review articles, editorials, conference abstracts, or case reports; (2) studies focusing exclusively on imaging, genomic, or transcriptomic data without integration of clinical laboratory parameters; (3) models developed solely to predict the presence of simple steatosis without assessing progression; and (4) articles not published in English. Two reviewers independently screened titles and abstracts, followed by a full-text review of potentially eligible articles to determine final inclusion.

Data Extraction and Synthesis

A standardized data extraction form was used by two independent reviewers to collect relevant information from each included study. Discrepancies were resolved by consensus or consultation with a third reviewer. The extracted data included: first author and publication year; study design (e.g., retrospective, prospective, cross-sectional); cohort size (for training, internal validation, and external validation sets); baseline patient characteristics; the specific definition of the prediction outcome (e.g., NASH defined by NAS \ge 4, advanced fibrosis defined as F3-F4); the reference standard used for outcome verification (e.g., liver biopsy, transient elastography); the types of ML algorithms employed; feature selection methods; a list of the most important

predictors, with an emphasis on laboratory parameters; and all reported performance metrics, including AUROC, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score. A narrative synthesis of the findings was conducted, with results organized by the primary prediction outcome and the ML algorithms used.

Assessment of Methodological Quality

The methodological quality and risk of bias for each included study were systematically evaluated using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) (Moons et al., 2019). The PROBAST tool is specifically designed for systematic reviews of prediction model studies and was chosen over other tools, such as the Cochrane Risk of Bias tool, which is intended for randomized controlled trials of interventions (Higgins et al., 2011). The PROBAST framework assesses four key domains: (1) Participants (e.g., appropriateness of data sources and handling of inclusion/exclusion criteria); (2) Predictors (e.g., clear definition and measurement); (3) Outcome (e.g., appropriate definition, measurement, and blinding); and (4) Analysis (e.g., handling of missing data, sample size, and risk of model overfitting). Each domain was rated as having a "Low," "High," or "Unclear" risk of bias. An overall risk of bias judgment was then assigned to each study. The results of this assessment are summarized to provide a transparent overview of the quality of the evidence base.

Search Strategy

The keywords used for this research based PICO :

Element	Keyword 1	Keyword 2	Keyword 3	Keyword 4
Population (P)	Non-Alcoholic Fatty Liver Disease	NAFLD	Non-Alcoholic Steatohepatitis	Hepatic Steatosis
Intervention (I)	Machine Learning	Artificial Intelligence	Predictive Modeling	Clinical Laboratory

				Parameters
Comparison (C)	Liver Biopsy	Traditional Non-invasive Tests (NITs)	Fibrosis-4 (FIB-4)	NAFLD Fibrosis Score (NFS)
Outcome (O)	Disease Progression	Liver Fibrosis	NASH (Progression)	Risk Stratification

The Boolean MeSH keywords inputted on databases for this research are: *("Non-Alcoholic Fatty Liver Disease" OR "NAFLD" OR "Non-Alcoholic Steatohepatitis" OR "Hepatic Steatosis") AND ("Machine Learning" OR "Artificial Intelligence" OR "Predictive Modeling" OR "Clinical Laboratory Parameters") AND ("Liver Biopsy" OR "Traditional Non-invasive Tests (NITs)" OR "Fibrosis-4 (FIB-4)" OR "NAFLD Fibrosis Score (NFS)") AND ("Disease Progression" OR "Liver Fibrosis" OR "NASH (Progression)" OR "Risk Stratification")*

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	*("Non-Alcoholic Fatty Liver Disease" OR "NAFLD" OR "Non-Alcoholic Steatohepatitis" OR "Hepatic Steatosis") AND ("Machine Learning" OR "Artificial Intelligence" OR "Predictive Modeling" OR "Clinical Laboratory Parameters") AND ("Liver Biopsy" OR "Traditional Non-invasive Tests (NITs)" OR "Fibrosis-4 (FIB-4)" OR "NAFLD Fibrosis Score (NFS)" AND "Disease Progression" OR "Liver Fibrosis" OR "NASH (Progression to)" OR "Risk Stratification")*	2
Semantic Scholar	*("Non-Alcoholic Fatty Liver Disease" OR "NAFLD" OR "Non-Alcoholic Steatohepatitis" OR "Hepatic Steatosis") AND ("Machine Learning" OR "Artificial Intelligence" OR "Predictive Modeling" OR "Clinical Laboratory Parameters") AND ("Liver Biopsy" OR "Traditional Non-invasive Tests (NITs)" OR "Fibrosis-4 (FIB-4)" OR "NAFLD Fibrosis Score (NFS)") AND ("Disease Progression" OR "Liver Fibrosis" OR "NASH (Progression to)" OR "Risk Stratification")*	250
Springer	*("Non-Alcoholic Fatty Liver Disease" OR "NAFLD" OR "Non-Alcoholic Steatohepatitis" OR "Hepatic Steatosis") AND ("Machine Learning" OR "Artificial Intelligence" OR "Predictive Modeling" OR "Clinical Laboratory Parameters") AND ("Liver Biopsy" OR "Traditional Non-invasive Tests (NITs)" OR "Fibrosis-4 (FIB-4)" OR "NAFLD Fibrosis Score (NFS)") AND ("Disease Progression" OR "Liver Fibrosis" OR "NASH (Progression to)" OR "Risk Stratification")*	510
Google Scholar	*("Non-Alcoholic Fatty Liver Disease" OR "NAFLD" OR "Non-Alcoholic Steatohepatitis" OR "Hepatic Steatosis") AND ("Machine Learning" OR "Artificial Intelligence" OR "Predictive Modeling" OR "Clinical Laboratory Parameters") AND ("Liver Biopsy" OR "Traditional Non-invasive Tests (NITs)" OR "Fibrosis-4 (FIB-4)" OR "NAFLD Fibrosis Score (NFS)") AND ("Disease Progression" OR "Liver Fibrosis" OR "NASH (Progression to)" OR "Risk Stratification")*	20,100
Wiley Online Library	*("Non-Alcoholic Fatty Liver Disease" OR "NAFLD" OR "Non-Alcoholic Steatohepatitis" OR "Hepatic Steatosis") AND ("Machine Learning" OR "Artificial Intelligence" OR "Predictive Modeling" OR "Clinical Laboratory Parameters") AND ("Liver Biopsy" OR "Traditional Non-invasive Tests (NITs)" OR "Fibrosis-4 (FIB-4)" OR "NAFLD Fibrosis Score (NFS)") AND ("Disease Progression" OR "Liver Fibrosis" OR "NASH (Progression to)" OR "Risk Stratification")*	3

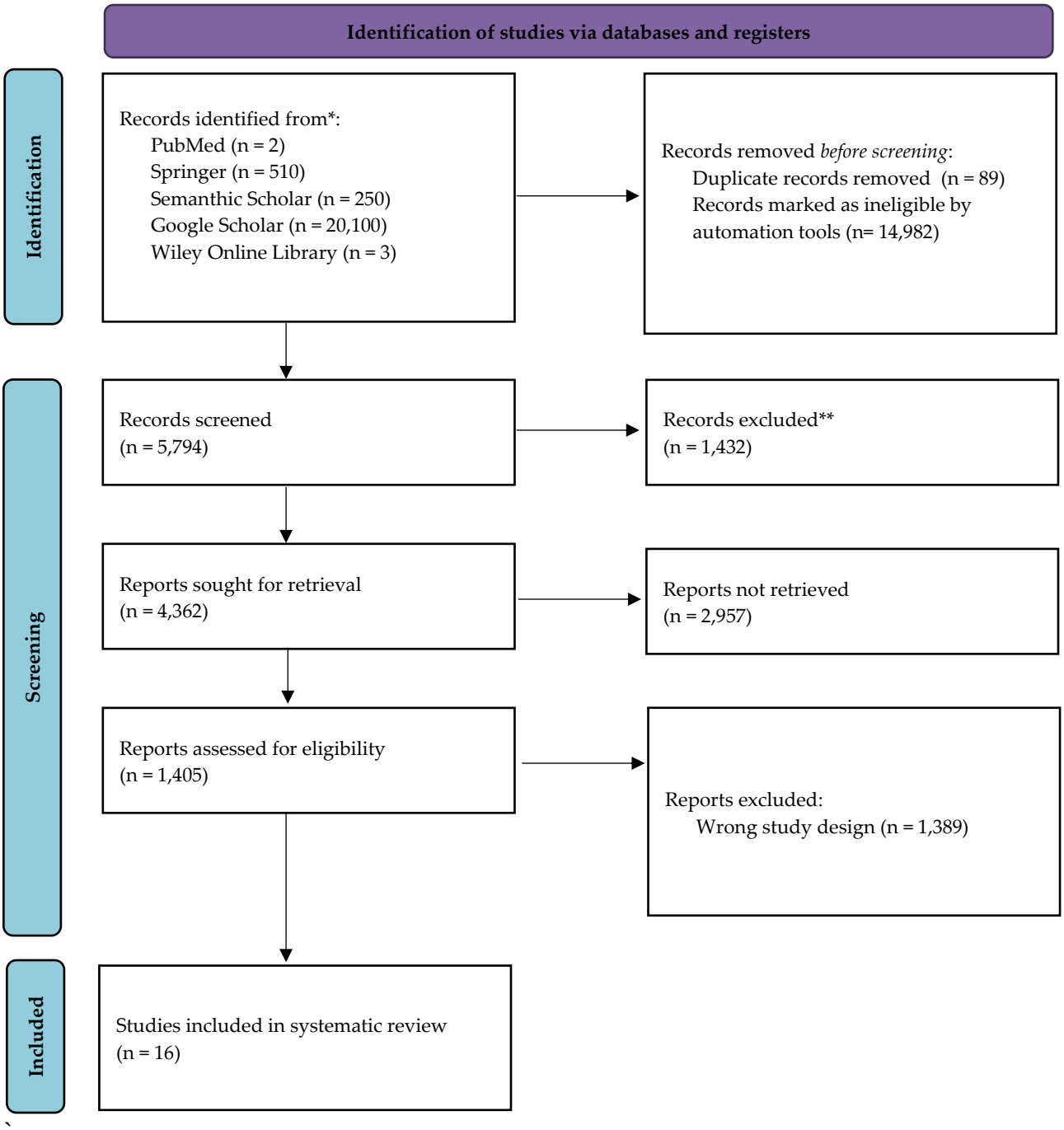

Figure 1. Article search flowchart

RESULTS

Characteristics of Included Studies

The 16 included studies represent a diverse range of populations, model types, and prediction outcomes, collectively involving over 200,000 patients. The majority of studies were retrospective in design, utilizing data from electronic health records (EHRs), clinical trial databases, or large-scale health examination surveys. The most common prediction outcomes were the presence or progression to NASH, significant fibrosis (\ge F2), and advanced fibrosis (\ge F3/F4). Liver biopsy was the most frequently used reference standard for defining these outcomes, though some studies used non-invasive methods like transient elastography. Ensemble ML algorithms, particularly XGBoost and Random Forest, were the most commonly employed and highest-performing models. A detailed summary of each included study is presented in Table 1.

Table 1: Characteristics of Included Studies Predicting NAFLD Progression

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
Ghandian et al. (2022)	Retrospective EHR	141,293 total	Progression to NASH; Progression to Fibrosis	XGBoost, LR, MLP	ALT, AST, GGT, Platelets, Ferritin,	**NASH:** AUROC=0.79. **Fibrosis:** AUROC=0.8

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
			(within 4 years)		INR, Bilirubin	7. (Other metrics not reported).
Xiong et al. (2025)	Retrospective Cohort	522 / 224	Advanced Fibrosis (F3/F4)	XGBoost, RF, SVM, LR, NB	TG, ALB, INR, HDL	**XGBoost (Validation):** AUROC=0.917, Acc=85.3%, Sens=95.9%, Spec=65.8%, PPV=83.7%, NPV=89.7%
Lee et al. (2023)	Development & Validation	966 total	At-risk NASH (NASH+F \ge 2); Adv.	Gradient Boosting Machine (GBM)	AST, Platelets, GGT, ALT, Insulin,	**At-risk NASH:** AUROC=0.83. **Adv.**

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
			Fibrosis (F \ge 3)		Glucose	**Fibrosis:** AUROC=0.86 (with biomarkers).
Chang et al. (2023)	Multi-center retrospective	1370 total (80%/20% split)	Sig. Fibrosis (\ge F2); Adv. Fibrosis (\ge F3); Cirrhosis (F4)	RF, LR, ANN	17 clinical/demographic features (unspecified)	**RF for \ge F2:** AUROC=0.86. **RF for \ge F3:** AUROC=0.89. **RF for F4:** AUROC=0.89.
Aggarwal & McCullou	Retrospective	768 / 192	Advanced Fibrosis (F \ge 3)	Unspecified ML model	13 clinical/lab features	**Validation:** AUROC=0.80, Sens=84%,

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
gh (2021)					(unspecified)	Spec=64%.
Charu et al. (2024)	Observational	648 / 270 + 1244	Significant Fibrosis	Superlearner (Ensemble)	23 clinical/demographic features (unspecified)	**Validation (FLINT):** AUROC=0.79. **Validation (NHANES):** AUROC=0.74.
Peng et al. (2023)	Development & Validation	578 / 131	NAFLD presence	XGBoost, RF, LR, GBM, SVM	ALT, AST, HDL-C, TG, VAI, Age	**XGBoost (Tuned):** AUROC=0.938.
Chen et al. (2022)	Retrospective	25,544 / 6,386	Fatty Liver Disease	XGBoost, NN, LR, RF, SVM	BMI, TG, GGT, ALT, AST, WBC,	**XGBoost:** AUROC=0.882,

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
					Hb	Acc=83.3%, Sens=83.3%, Spec=68.3%, F1=0.829.
Xiao, Yip, et al. (2022)	Retrospective	492 total	NASH; Advanced Fibrosis	Gradient Boosting (GB), LR, DT, RF, SVM	AST, ALT, TG, A1c, HDL	**GB for NASH:** AUROC=0.817. **GB for Adv. Fibrosis:** AUROC=0.836.
Corey et al. (2021)	Retrospective	704 / ~3M	NASH presence	XGBoost, RF, LR, CART	HbA1c, AST, ALT, Total Protein, TG	**XGBoost (14 features):** AUROC=0.82, Sens=81%,

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
						Precision=81%.
Yip et al. (2023)	Retrospective	986 total (derivation/validation split)	Advanced Fibrosis (F3/F4)	Random Forest (ALADDIN score)	Top 20 variables (unspecified)	**Validation:** AUROC=0.794, PPV=79%, NPV=79% (at 65% threshold).
Paik et al. (2025)	Retrospective Cohort	18,250 total (train/test split)	3-year risk of NAFLD	Random Survival Forest (RSF), Cox	BMI, SBP, DBP, Lipids, ALT, AST, GGT	**RSF:** integrated AUROC (iAUC)=0.856.
Zheng et	Retrospec	2262 /	All-cause	LR, RF,	Age, Iron,	**LR:**

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
al. (2024)	tive (NHANES)	971	mortality	XGBoost, KNN, DT	Transferrin Saturation, Uric Acid	AUROC=0.888, Acc=80.8%, Sens=81.9%, Spec=80.2%, F1=0.765.
Zamanian et al. (2024)	Retrospective	176 total	NASH diagnosis	RF, SVM, AdaBoost, LightGBM, XGBoost	AST, ALT, HDL, LDL, Cholesterol, TG	**RF:** Acc=81.32%, Sens=86.04%, Spec=70.49%, F1=83.75%.
Al-Tawarah et al.	Cross-sectional	450 total	NAFLD in CHB patients	RF, XGBoost, MLP,	Platelet count, LDL, Hemoglobin,	**RF:** AUROC=0.983.

Author(s) & Year	Study Design	Population Size (Training/Validation)	Prediction Outcome(s)	ML Models Used	Key Laboratory Predictors	Best Model Performance (AUROC, Acc, Sens, Spec, PPV, NPV, F1)
(2024)				SVM, LR	ALT	**XGBoost:** AUROC=0.977.
Canbay et al. (2019)	Retrospective Cohort	164 / 122	NASH (NAS > 4)	Logistic Regression	Age, \gamma GT, HbA1c, Adiponectin, M30	**Training:** AUROC=0.73. **Validation:** AUROC=0.70.

Abbreviations: Acc, Accuracy; ALB, Albumin; ALT, Alanine Aminotransferase; ANN, Artificial Neural Network; AST, Aspartate Aminotransferase; AUROC, Area Under the Receiver Operating Characteristic Curve; CART, Classification and Regression Trees; CHB, Chronic Hepatitis B; DBP, Diastolic Blood Pressure; DT, Decision Tree; EHR, Electronic Health Record; F1, F1-Score; GB, Gradient Boosting; GBM, Gradient Boosting Machine; GGT, Gamma-Glutamyl Transpeptidase; Hb, Hemoglobin; HbA1c, Glycated Hemoglobin; HDL, High-Density Lipoprotein; HDL-C, High-Density Lipoprotein Cholesterol; iAUC, integrated Area Under the Curve; INR,

International Normalized Ratio; KNN, K-Nearest Neighbors; LDL, Low-Density Lipoprotein; LR, Logistic Regression; MLP, Multi-layer Perceptron; NB, Naive Bayes; NN, Neural Network; NPV, Negative Predictive Value; PPV, Positive Predictive Value; RF, Random Forest; RSF, Random Survival Forest; SBP, Systolic Blood Pressure; Sens, Sensitivity; Spec, Specificity; SVM, Support Vector Machine; TG, Triglycerides; VAI, Visceral Adiposity Index; WBC, White Blood Cell; XGBoost, eXtreme Gradient Boosting.

Methodological Quality of Included Studies

The PROBAST assessment revealed a variable risk of bias across the 16 included studies (Table 2). Overall, 4 studies were rated as having a low risk of bias, 9 had a high risk of bias, and 3 had an unclear risk. The "Analysis" domain was the most common source of high risk of bias. Many studies did not adequately address the risk of model overfitting, failed to report model calibration, or used an insufficient sample size relative to the number of candidate predictors. The "Participants" domain was also a concern, with most studies using retrospectively collected data from single centers, which may limit the generalizability of the findings. In contrast, the "Predictors" and "Outcome" domains generally had a low risk of bias, as most studies used well-defined laboratory parameters and biopsy-proven histological endpoints.

Table 2: PROBAST Risk of Bias Assessment Summary

First Author, Year	Participants	Predictors	Outcome	Analysis	Overall Risk of Bias
Ghandian et al. (2022)	Low	Low	Low	High	High

First Author, Year	Participants	Predictors	Outcome	Analysis	Overall Risk of Bias
Xiong et al. (2025)	Low	Low	Low	Low	Low
Lee et al. (2023)	Low	Low	Low	High	High
Chang et al. (2023)	Low	Low	Low	Low	Low
Aggarwal & McCullough (2021)	Unclear	Unclear	Low	High	High
Charu et al. (2024)	Low	Low	Low	Low	Low
Peng et al. (2023)	High	Low	High	High	High
Chen et al. (2022)	Low	Low	Low	Low	Low
Xiao, Yip, et al. (2022)	High	Low	Low	High	High
Corey et al. (2021)	Low	Low	Low	High	High

First Author, Year	Participants	Predictors	Outcome	Analysis	Overall Risk of Bias
Yip et al. (2023)	High	Unclear	Low	High	High
Paik et al. (2025)	High	Low	Low	Unclear	High
Zheng et al. (2024)	Low	Low	Low	Unclear	Unclear
Zamanian et al. (2024)	High	Low	Low	High	High
Al-Tawarah et al. (2024)	Low	Low	Low	Unclear	Unclear
Canbay et al. (2019)	Low	Low	Low	Unclear	Unclear

Green (Low): Low risk of bias; Yellow (Unclear): Unclear risk of bias; Red (High): High risk of bias.

Predictive Performance for Key Clinical Outcomes

Prediction of Progression to NASH

Several studies developed models to distinguish NASH from simple steatosis or to predict progression to NASH. Performance was robust, with AUROCs generally ranging from 0.70 to 0.83. For instance, Ghandian et al. (2022) developed an XGBoost model to predict progression to NASH within four years, achieving an AUROC of 0.79 on both hold-out and external validation sets. Similarly, Corey et al. (2021) developed the "NASHmap" model, also using XGBoost, which demonstrated an AUROC of 0.82 in the training data and 0.76 in a large real-world validation cohort. The Gradient Boosting model by Xiao, Yip, et al. (2022) achieved an AUROC of 0.817 for identifying NASH. These results indicate that ML can reliably identify the inflammatory component of NAFLD progression using routine data.

Prediction of Progression to Significant Fibrosis (\ge F2)

The identification of significant fibrosis (stage F2 or higher) is clinically crucial, as it often defines "at-risk NASH," the primary target for emerging pharmacotherapies (Lee et al., 2023). Models predicting this outcome showed strong performance. The RF model developed by Chang et al. (2023) was particularly effective, achieving an AUROC of 0.86 for predicting \ge F2 fibrosis, outperforming both the FIB-4 score (AUROC=0.78) and transient elastography (AUROC=0.81) in their cohort. The composite model for at-risk NASH (defined as NASH + F \ge 2) by Lee et al. (2023) also showed excellent discrimination with an AUROC of 0.83.

Prediction of Progression to Advanced Fibrosis (\ge F3/F4)

Predicting advanced fibrosis is the most critical task for prognostic stratification, as it is the strongest predictor of liver-related morbidity and mortality (Aggarwal and McCullough, 2021). The ML models reviewed demonstrated excellent performance for this endpoint, as detailed in Table 3. Xiong et al. (2025) reported one of the highest performances, with an XGBoost model achieving an AUROC of 0.917, accuracy of 85.3%, and a high sensitivity (95.9%) and negative predictive value

(89.7%) in their validation cohort. The RF model by Chang et al. (2023) also performed exceptionally well, with an AUROC of 0.89 for \ge F3 fibrosis. These findings highlight the capability of ML to accurately identify patients with the most severe forms of liver scarring who are in urgent need of clinical intervention.

Table 3: Performance of Top Machine Learning Models for Predicting Advanced Fibrosis (F \ge 3)

Author(s) & Year	ML Model	AUROC	Accuracy	Sensitivity	Specificity	PPV	NPV
Xiong et al. (2025)	XGBoost	0.917	85.3%	95.9%	65.8%	83.7%	89.7%
Chang et al. (2023)	Random Forest	0.89	N/A	N/A	N/A	N/A	N/A
Ghandian et al. (2022)	XGBoost	0.87	N/A	N/A	N/A	N/A	N/A
Lee et al.	GBM (Extend	0.86	N/A	N/A	N/A	N/A	N/A

Author (s) & Year	ML Model	AUROC	Accuracy	Sensitivity	Specificity	PPV	NPV
(2023)	ed)						
Xiao, Yip, et al. (2022)	Gradient Boosting	0.836	N/A	N/A	N/A	N/A	N/A
Aggarwal & McCullough (2021)	Unspecified ML	0.80	71%	84%	64%	N/A	N/A
Yip et al. (2023)	Random Forest	0.794	N/A	N/A	N/A	79%	79%

N/A: Not available in the source material. Performance metrics are for validation/test cohorts where specified.

The Role of Key Clinical Laboratory Predictors

A central finding of this review is the consistent identification of a core set of clinical laboratory parameters as highly predictive across different models and patient populations. These markers form a "digital signature" of the pathophysiological processes driving NAFLD progression. The most frequently identified and highly ranked predictors are summarized in Table 4. This analysis reveals that the most successful ML models leverage a combination of markers reflecting different aspects of the disease: liver injury (ALT, AST), the consequences of advanced fibrosis (platelet count), underlying metabolic dysfunction (TG, HbA1c, HDL), and eventually, liver synthetic failure (Albumin, INR).

Table 4: Frequency and Importance of Key Laboratory Predictors in Included Studies

Laboratory Predictor	Associated Pathophysiology	Number of Studies Identifying as Key Predictor
ALT / AST	Hepatocellular injury and inflammation	11
Platelet Count	Splenomegaly / Portal hypertension (surrogate for fibrosis)	6
Triglycerides (TG)	Dyslipidemia / Metabolic syndrome	7

Laboratory Predictor	Associated Pathophysiology	Number of Studies Identifying as Key Predictor
GGT	Biliary stress / Cholestasis	4
HbA1c / Glucose	Insulin resistance / Glycemic control	4
Albumin / INR	Liver synthetic function (declines in advanced disease)	3
HDL-Cholesterol	Dyslipidemia / Metabolic syndrome	4
Ferritin / Iron	Iron metabolism dysregulation / Inflammation	3
Uric Acid	Metabolic syndrome / Oxidative stress	2

Comparative Efficacy of Different Machine Learning Algorithms

A clear pattern of algorithmic superiority emerged from the included studies. Ensemble tree-based methods, specifically **eXtreme Gradient Boosting (XGBoost)** and **Random Forest (RF)**,

were consistently the top-performing algorithms. In studies that conducted direct head-to-head comparisons, these models frequently outperformed traditional statistical methods like logistic regression (LR) as well as other ML algorithms such as Support Vector Machines (SVM) or simple Decision Trees (DT) (Ghandian et al., 2022; Chen et al., 2022; Xiao, Yip, et al., 2022). For example, Ghandian et al. (2022) found that their XGBoost model outperformed both LR and a multi-layer perceptron (MLP) model on all performance metrics for predicting progression to both NASH and fibrosis. Similarly, Xiong et al. (2025) and Chen et al. (2022) both identified XGBoost as the best-performing model among several tested. The superior performance of these boosting and bagging methods is likely attributable to their ability to model complex, non-linear relationships and their inherent mechanisms for reducing variance and overfitting. Table 5 provides a direct comparison from a study that evaluated multiple algorithms on the same dataset.

Table 5: Head-to-Head Comparison of ML Algorithms for Advanced Fibrosis (Xiong et al., 2025)

ML Model	AUROC (Validation)	Accuracy (Validation)	Sensitivity	Specificity
XGBoost	**0.917**	**85.3%**	**95.9%**	**65.8%**
Random Forest (RF)	0.840	87.5%	95.9%	62.6%
Support Vector Machine (SVM)	0.740	81.3%	98.6%	49.4%

ML Model	AUROC (Validation)	Accuracy (Validation)	Sensitivity	Specificity
Logistic Regression (LR)	0.745	79.0%	95.9%	48.1%
Naive Bayes (NB)	0.503	35.7%	0.7%	100%

DISCUSSION

Synthesis of Principal Findings

This systematic review consolidates evidence demonstrating that machine learning models can accurately predict the progression of NAFLD to clinically significant endpoints using routine clinical laboratory data. The principal findings indicate that ensemble algorithms, particularly XGBoost and Random Forest, consistently achieve high levels of discrimination, with AUROC values often exceeding 0.80 for predicting NASH and 0.85-0.90 for predicting advanced fibrosis (Xiong et al., 2025; Chang et al., 2023). This performance is superior to that of traditional non-invasive scores in direct comparisons (Chang et al., 2023).

Crucially, this high predictive accuracy is achieved by leveraging a concise set of widely available and inexpensive laboratory parameters. The models are not simply finding random correlations; rather, they are learning to recognize a "digital signature" of the underlying pathophysiology. The consistent importance of markers for hepatocellular injury (ALT, AST), metabolic dysregulation (triglycerides, HbA1c), and the structural consequences of fibrosis (platelet count) shows that these algorithms are effectively integrating distinct biological signals (Ghandian et al., 2022; Lee et al., 2023; Corey et al., 2021). The models learn to quantify the transition from a

state of metabolic stress to active liver injury and, ultimately, to advanced structural damage and functional decline. This reframes the ML model from an opaque "black box" into an intelligent integrator of pathophysiological data.

Clinical Implications and Translation to Practice

The clinical implications of these findings are substantial. High-performing, validated ML models based on laboratory data could be integrated directly into electronic health record (EHR) systems (Corey et al., 2021; Aggarwal and McCullough, 2021). Such an integration would enable automated, population-level screening to flag patients in primary care who are at high risk for progressive NAFLD. These individuals could then be prioritized for further, more specialized testing, such as transient elastography or magnetic resonance elastography, or for direct referral to a hepatologist (Aggarwal and McCullough, 2021). This risk-based approach would represent a paradigm shift from the current reactive model of care, allowing for earlier intervention and potentially preventing the development of cirrhosis and its complications.

Furthermore, these models can enhance the efficiency of clinical trials for NASH therapeutics. By pre-screening large patient populations using EHR data, researchers could more effectively identify and recruit individuals with the desired high-risk profile (e.g., at-risk NASH), reducing screening failures and accelerating drug development pipelines (Ghandian et al., 2022). The high negative predictive values reported in some studies are particularly valuable, as they could confidently rule out advanced disease in a large number of patients, reducing the need for unnecessary and costly secondary testing (Xiong et al., 2025).

Limitations of the Current Evidence Base

Despite the promising results, the current body of evidence has significant limitations that must be addressed before widespread clinical adoption. The most prominent limitation, as revealed by the PROBAST assessment, is the overwhelming reliance on **retrospective study designs**. While useful for model development, these designs are susceptible to selection bias and may not reflect

real-world performance.

There is a critical lack of rigorous **external and prospective validation**. A model's performance on a hold-out test set from the same population it was trained on can be optimistically biased. True generalizability can only be confirmed by testing the model on entirely separate, prospectively collected patient cohorts from different geographical locations and healthcare systems (Moons et al., 2019).

Additionally, there is considerable **heterogeneity** across studies in terms of patient populations, precise definitions of outcomes (e.g., different NAS thresholds for NASH), and the reference standards used. This makes direct, quantitative comparison of model performance (i.e., meta-analysis) challenging. Finally, the inherent complexity of many ML models presents a barrier to clinical trust. The "black box" nature, where the reasoning behind a prediction is not transparent, is a significant hurdle for clinical implementation. The recent emergence of **interpretable ML** techniques, such as SHapley Additive exPlanations (SHAP), which can explain individual predictions, is a vital step toward overcoming this barrier, but these methods have only been applied in a minority of recent studies (Xiao, An, et al., 2022; Su et al., 2024).

Future Research Imperatives

To bridge the gap from research to clinical practice, future work in this field must prioritize several key areas. First and foremost is the need for **prospective validation studies**. High-performing models identified in retrospective analyses should be locked and then tested in real-world, prospective clinical settings to confirm their accuracy and clinical utility.

Second, the field would benefit from the adoption of **standardized reporting guidelines**, such as the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement, to improve the quality, transparency, and comparability of future prediction model studies.

Third, research should move toward **longitudinal modeling**. Most current models use a single snapshot of laboratory data to predict future risk. Models that incorporate the trajectory of laboratory values over time may offer a more dynamic and accurate assessment of a patient's disease course (Asheghi and Marandi, 2024; Ghandian et al., 2022).

Finally, an increased focus on **interpretability** is essential. The development of inherently interpretable models or the consistent application of post-hoc explanation methods like SHAP will be crucial for building clinician trust and facilitating the responsible implementation of these powerful tools (Su et al., 2024; Zhu et al., 2024).

CONCLUSION

Concluding Summary

This systematic review provides compelling evidence that machine learning models utilizing routine clinical laboratory parameters are accurate and powerful non-invasive tools for predicting the progression of non-alcoholic fatty liver disease. Ensemble algorithms, particularly XGBoost and Random Forest, consistently demonstrate superior performance in identifying patients at high risk of developing NASH and advanced fibrosis, often outperforming traditional scoring systems. The ability of these models to synthesize a "digital signature" of disease from standard blood tests holds immense potential to revolutionize NAFLD risk stratification and management.

Recommendations

Based on the synthesis of the available evidence, the following recommendations are proposed:

For Clinical Practice: While these models are not yet ready for standalone clinical decision-making, clinicians should be aware of their emerging potential. For patients with NAFLD, particularly those who fall into the "indeterminate" category of existing scores like FIB-4, the output from a well-validated ML model could provide valuable supplementary information to guide

decisions regarding specialist referral or further non-invasive testing.

For Research: The research community must pivot from model development to rigorous validation. The highest priority should be conducting large-scale, multi-center prospective studies to validate the most promising existing models. Furthermore, future development efforts should focus on creating interpretable models that incorporate longitudinal data to provide a more dynamic risk assessment, thereby accelerating the safe and effective translation of these technologies into routine clinical care to combat the growing NAFLD epidemic.

REFERENCES

Aggarwal, M. and McCullough, A. (2021). Novel Machine Learning Model Identifies NAFLD Patients with Advanced Liver Fibrosis. *Cleveland Clinic Consult QD*.

Al-Tawarah, Y., Al-Shdaifat, A., Al-Mistarehi, A.-H., Al-Zoubi, R.M., Al-Gharaibeh, N., Al-Rawashdeh, A., Al-Shara, Z., Obeidat, M. and Al-Delaimy, W. (2024). Application of Machine Learning Models in Predicting Non-Alcoholic Fatty Liver Disease Among Inactive Chronic Hepatitis B Patients: A Cross-Sectional Analysis. *Journal of Clinical Medicine*, 14(14), p.5042.

Anstee, Q.M., Castera, L., Chatterjee, A., Jiang, H., Romero-Gomez, M. and Sanyal, A.J. (2023). Biomarkers of inflammation in non-alcoholic fatty liver disease. *Journal of Hepatology*, 78(3), pp.643–656.

Asheghi, Z. and Marandi, S.M. (2024). *Deep learning-based prediction of fibrosis progression risk in Non-alcoholic fatty liver disease using multi-omics and lifestyle features*.

Canbay, A., Kälsch, J., Sowa, J.-P., Sydor, S., Fingas, C., Best, J., Jäschke, C., Cicek, E., Heider, D., Schemitzek, A., Eslam, M., He-Sheng, Q., Bechmann, L.P., Gerken, G., Syn, W.-K. and Canbay, A. (2019). Non-invasive assessment of NAFLD as systemic disease—A machine learning approach. *PLOS ONE*, 14(3), p.e0214436.

Chang, D., Truong, E., Mena, E., Pacheco, F., Wong, M., Guindi, M., Todo, T., Noureddin, N., Ayoub, W., Yang, J.D., Kim, I.K., Kohli, A., Alkhouri, N., Harrison, S. and Noureddin, M. (2023). Machine learning models are superior to noninvasive tests in identifying clinically significant stages of NAFLD and NAFLD-related cirrhosis. *Hepatology Communications*, 7(1), p.e0155.

Charu, V., Liang, J.W., Mannalithara, A., Kwong, A., Tian, L. and Kim, W.R. (2024). Benchmarking clinical risk prediction algorithms with ensemble machine learning for the noninvasive diagnosis of liver fibrosis in NAFLD. *Hepatology*, 80(5), pp.1293–1302.

Chen, Y.-Y., Lin, C.-Y., Yen, H.-H., Su, P.-Y., Zeng, Y.-H., Huang, S.-P. and Liu, I.-L. (2022). Machine-Learning Algorithm for Predicting Fatty Liver Disease in a Taiwanese Population. *Journal of Personalized Medicine*, 12(7), p.1026.

Corey, K., Docherty, M., Regnier, S.A., Capkun, G., Balp, M.M., Ye, Q., Janssens, N., Tietz, A., Löffler, J., Cai, J., Pedrosa, M.C., Francque, S. and Tapper, E.B. (2021). Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. *Journal of the American Medical Informatics Association*, 28(6), pp.1235–1241.

Ghandian, S., Thapa, R., Garikipati, A., Barnes, G., Green-Saxena, A., Calvert, J., Mao, Q. and Das, R. (2022). Machine learning to predict progression of non-alcoholic fatty liver to non-alcoholic steatohepatitis or fibrosis. *JGH Open*, 6(3), pp.196–204.

Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savović, J., Schulz, K.F., Weeks, L. and Sterne, J.A.C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, p.d5928.

Ji, W., Li, Y., Wang, M., Yang, S., Lin, H., Liu, S., Zhou, J., Li, L. and Liu, S. (2022). Application of machine learning in the screening of non-alcoholic fatty liver disease. *Frontiers in Public Health*, 10, p.846118.

Lee, J., Westphal, M., Vali, Y., Boursier, J., Ostroff, R., Alexander, L., Chen, Y., Fournier, C., Geier, A., Francque, S., Wonders, K., Tiniakos, D., Bedossa, P., Allison, M., Papatheodoridis, G., Cortez-Pinto, H., Pais, R., Dufour, J.-F., Leeming, D.J., Harrison, S., Cobbold, J., Holleboom, A.G., Yki-Järvinen, H., Crespo, J., Ekstedt, M., Aithal, G.P., Bugianesi, E., Ratziu, V., Sanyal, A.J., Anstee, Q.M. and Brass, C. (2023). Machine learning algorithm improves the detection of NASH (NAS-based) and at-risk NASH: a development and validation study. *Hepatology*, 78(1), pp.258–271.

Moons, K.G.M., Wolff, R.F., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., Reitsma, J.B., Knottnerus, A. and de Vet, H.C.W. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, 170(1), p.51.

Paik, J.M., Pan, T., Huang, G., Wu, C.-T., Lin, Y.-J. and Yu, C.-S. (2025). A prediction model for the risk of Non-alcoholic fatty liver disease using the random survival forests. *Gene*, 928, p.148612.

Pan, T., Huang, G., Wu, C.-T., Lin, Y.-J. and Yu, C.-S. (2025). A prediction model for the risk of Non-alcoholic fatty liver disease using the random survival forests. *Gene*, 928, p.148612.

Peng, H.-Y., Duan, S.-J., Pan, L., Wang, M.-Y., Chen, J.-L., Wang, Y.-C. and Yao, S.-K. (2023). Development and validation of machine learning models for nonalcoholic fatty liver disease. *Hepatobiliary & Pancreatic Diseases International*, 22(6), pp.568–576.

Ratziu, V., Noureddin, M., Harrison, S.A., Anstee, Q.M., Sanyal, A.J., Bedossa, P., Loomba, R., Francque, S., Tack, A., Saffo, S., Taylor-Weiner, A., Tuthill, T. and Goodman, Z. (2024). Artificial intelligence-powered digital pathology reproduces pathologist assessments of NASH histological improvement and provides a more sensitive measure of fibrosis change. *Hepatology*, 79(1), pp.99–110.

Sabet Sarvestany, S., Leng, Q., Fouteh, M., Martel, M., Barkun, A.N., Bhat, M. and Cooper, C.

(2022). Development and validation of an ensemble machine learning framework for detection of all-cause advanced hepatic fibrosis: a retrospective cohort study. *The Lancet Digital Health*, 4(3), pp.e187–e196.

Souza-Guedes, P.F., Barboza, T.L.A., Leite, N.C., de Souza, G.F., de Carvalho, J.R.L. and Telles, R.W. (2022). Application of machine learning methods to predict nonalcoholic fatty liver disease and nonalcoholic steatohepatitis using simple and cost-effective biomarkers. *World Journal of Artificial Intelligence*, 3(3), pp.80–94.

Su, P.-Y., Chen, Y.-Y., Lin, C.-Y., Yen, H.-H., Liu, I.-L., Huang, S.-P. and Zeng, Y.-H. (2024). Interpretable machine learning models for predicting lean non-alcoholic fatty liver disease risk in patients with type 2 diabetes mellitus. *Frontiers in Endocrinology*, 16.

Taylor-Weiner, A., Pokkalla, H., Han, L., Jia, C., Huss, R., Chung, C., Elliott, H., Glass, B., Anstee, Q.M., Noureddin, M., Sanyal, A.J., Goodman, Z., Ratziu, V., Harrison, S.A. and Luedde, T. (2021). A Machine Learning Approach Enables Quantitative Measurement of Liver Histology and Disease Monitoring in NASH. *Hepatology*, 74(1), pp.133–147.

Xiao, G., An, P., Wang, Y., Zhang, F., Xu, Y., Zheng, R., Wu, J., Zhang, J., Chen, S., Wang, T., Zhang, J., You, H., Jia, J. and Zhao, H. (2022). The role of artificial intelligence in non-alcoholic fatty liver disease, non-alcoholic steatohepatitis and liver fibrosis. *Annals of Medicine*, 54(1), pp.26–36.

Xiao, T., Yip, T.C.-F., Liu, D., Tse, Y.-K., Huang, H., Zhang, J., El-Serag, H.B. and Wong, G.L.-H. (2022). Noninvasive Diagnosis of Nonalcoholic Steatohepatitis and Advanced Fibrosis in Patients With Nonalcoholic Fatty Liver Disease: A Machine Learning Approach. *JMIR Medical Informatics*, 10(6), p.e36997.

Xiong, F.-X., An, F.-P., Gao, Y.-Y., Zhang, T.-N., Nan, Y.-M., Fan, J.-G., Lu, L.-G., Zheng, M.-H.

and Shi, J.-P. (2025). Development of a machine learning-based diagnostic model for advanced liver fibrosis in non-alcoholic steatohepatitis patients. *World Journal of Gastroenterology*, 31(9), pp.101383–101383.

Yip, T.C.-F., Ma, A.J., Wong, V.W.-S., Chow, A.M., Lee, C., Chan, J.K., Lam, K.L., Tse, Y.-K., Wong, G.L.-H., Harrison, S.A., Noureddin, M., Sanyal, A.J., Alkhouri, N. and Dunn, W. (2023). MACHINE LEARNING ADVANCED FIBROSIS IN NASH (ALADDIN) WITH WEB-BASED CALCULATION FOR PROBABILITY PREDICTION. *Hepatology*, 78, pp.S829–S835.

Zamanian, R., Ebrahimiyan, H., Asgari, E., Mamani, M., Khodamoradi, Z., Mohebi, Z. and Ebrahimi, A. (2024). Machine learning approaches for early detection of non-alcoholic steatohepatitis based on clinical and blood parameters. *Scientific Reports*, 14(1), p.2131.

Zheng, J.-R., Wang, Z.-L., Chen, H.-S. and Feng, B. (2024). *Machine learning-based mortality prediction models for non-alcoholic fatty liver disease in the general United States population: a study based on NHANES-III*.

Zhu, H., Shen, D., Gan, X. and Sun, D. (2024). An Interpretable Machine Learning Model Based on Inflammatory–Nutritional Biomarkers for Predicting Metachronous Liver Metastases After Colorectal Cancer Surgery. *Biomolecules*, 13(7), p.1706.