Received: 16 June 2022 | Revised: 20 December 2022 | Accepted: 22 December 2022

DOI: 10.1097/HEP.0000000000000364

ORIGINAL ARTICLE

OPEN

Machine learning algorithm improves the detection of NASH (NAS-based) and at-risk NASH: A development and validation study

Jenny Lee[1] | Max Westphal[2] | Yasaman Vali[1] | Jerome Boursier[3] | Salvatorre Petta[4] | Rachel Ostroff[5] | Leigh Alexander[5] | Yu Chen[6] | Celine Fournier[7] | Andreas Geier[8] | Sven Francque[9] | Kristy Wonders[10] | Dina Tiniakos[10,11] | Pierre Bedossa[10] | Mike Allison[12] | Georgios Papatheodoridis[13] | Helena Cortez-Pinto[14] | Raluca Pais[15] | Jean-Francois Dufour[16] | Diana Julie Leeming[17] | Stephen Harrison[18] | Jeremy Cobbold[18] | Adriaan G. Holleboom[19] | Hannele Yki-Järvinen[20] | Javier Crespo[21] | Mattias Ekstedt[22] | Guruprasad P. Aithal[23] | Elisabetta Bugianesi[24] | Manuel Romero-Gomez[25] | Richard Torstenson[26] | Morten Karsdal[17] | Carla Yunis[27] | Jörn M. Schattenberg[28] | Detlef Schuppan[29,30] | Vlad Ratziu[31] | Clifford Brass[32] | Kevin Duffin[6] | Koos Zwinderman[1] | Michael Pavlides[33] | Quentin M. Anstee[9,34] | Patrick M. Bossuyt[1] | on behalf of the LITMUS investigators

[1]Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, the Netherlands

[2]Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany

[3]Department of Hepatology, Angers University Hospital, Angers, France

[4]Section of Gastroenterology and Hepatology, Promozione della Salute, Materno-Infantile, di Medicina Interna e Specialistica di Eccellenza, Department, University of Palermo, Palermo, Italy

[5]SomaLogic Inc, Boulder, Colorado, USA

[6]Lilly Research Laboratories, Eli Lilly and Company Ltd (LLY), Indianapolis, Indiana, USA

[7]Echosens, 6 rue Ferrus, Paris, France

[8]Division of Hepatology, Department of Medicine II, Wurzburg University Hospital, Wurzburg, Germany

[9]Department of Gastroenterology Hepatology, Antwerp University Hospital, and Laboratory of Experimental Medicine and Paediatrics, University of Antwerp, Belgium

[10]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[11]Department of Pathology, Aretaieion Hospital, national and Kapodistrian University of Athens, Athens, Greece

[12]Liver Unit, Department of Medicine, Cambridge NIHR Biomedical Research Centre, Cambridge University NHS Foundation Trust, CB2 0QQ, Cambridge, UK

[13]Gastroenterology Department, National and Kapodistrian University of Athens, General Hospital of Athens "Laiko", Athens, Greece

[14]Clínica Universitária de Gastrenterologia, Faculdade de Medicina, Universidade de Lisboa, Portugal

[15]Assistance Publique-Hôpitaux de Paris, hôpital Pitié Salpêtrière, Sorbonne University, ICAN (Institute of Cardiometabolism and Nutrition), Paris, France

[16]Hepatology, Department of Biomedical Research, University of Bern, Bern, Switzerland

[17]Nordic Bioscience A/S, Herlev, Denmark

[18]Department of Gastroenterology and Hepatology, Oxford NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

[19]Department of Internal and Vascular Medicine, Amsterdam University Medical Centres, location AMC, Amsterdam, the Netherlands

[20]Department of Medicine, University of Helsinki and Helsinki University Hospital, Finland; Minerva Foundation Institute for Medical Research, Helsinki, Finland

[21]Department of Gastroenterology and Hepatology, University Hospital Marques de Valdecilla. Research Institute Valdecilla-IDIVAL, Santander, Spain

[22]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[23]Nottingham Digestive Diseases Centre, School of Medicine, NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and The University of Nottingham, Nottingham, UK

[24]Department of Medical Sciences, Division of Gastro-Hepatology, A.O. Città della Salute e della Scienza di Torino, University of Turin, Turin, Italy

[25]UCM Digestive Diseases, ciberehd, Virgen del Rocio University Hospital. Institute of Biomedicine of Seville (CSIC/HUVR/US), Department of Medicine, University of Seville, Seville, Spain

[26]Cardiovascular, Renal and Metabolism Regulatory Affairs, AstraZeneca, Mölndal, Sweden

[27]Internal Medicine and Hospital, Global Product Development, Pfizer, Inc, New York, New York, USA

[28]Metabolic Liver Research Program, I. Department of Medicine, University Medical Center Mainz, Mainz, Germany

[29]Institute of Translational Immunology and Research Center for Immune Therapy, University Medical Center Mainz, Mainz, Germany

[30]Division of Gastroenterology, Beth Israel Medical Center, Harvard Medical School, Boston, Massachusetts, USA

[31]Assistance Publique-Hôpitaux de Paris, hôpital Pitié Salpêtrière, Sorbonne University, ICAN (Institute of Cardiometabolism and Nutrition), Paris, France

[32]Novartis Pharmaceuticals Corporation, East Hanover, New Jersey

[33]Oxford University, Oxford, UK

[34]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK

Correspondence
Jenny Lee, Epidemiology and Data Science Amsterdam UMC, AMC Meibergdreef 9, 1105AZ Amsterdam, the Netherlands.
Email: j.a.lee@amsterdamumc.nl

Abstract

Background and Aims: Detecting NASH remains challenging, while at-risk NASH (steatohepatitis and F ≥ 2) tends to progress and is of interest for drug development and clinical application. We developed prediction models by supervised machine learning techniques, with clinical data and biomarkers to stage and grade patients with NAFLD.

Approach and Results: Learning data were collected in the Liver Investigation: Testing Marker Utility in Steatohepatitis metacohort (966 biopsy-proven NAFLD adults), staged and graded according to NASH CRN. Conditions of interest were the clinical trial definition of NASH (NAS ≥ 4;53%), at-risk NASH (NASH with F ≥ 2;35%), significant (F ≥ 2;47%), and advanced fibrosis (F ≥ 3;28%). Thirty-five predictors were included. Missing data were handled by multiple imputations. Data were randomly split into training/validation (75/25) sets. A gradient boosting machine was applied to develop 2 models for each condition: clinical versus extended (clinical and biomarkers). Two variants of the NASH and at-risk NASH models were constructed: direct and composite models. Clinical gradient boosting machine models for steatosis/inflammation/ballooning had AUCs of 0.94/0.79/0.72. There were no improvements when biomarkers were included. The direct

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; FIB-4, Fibrosis-4; GGT, gamma-glutamyl transferase; PRO-C3; Amino-terminal propeptide of procollagen type III; P3NP, Amino-terminal propeptide of type III procollagen; AUC, Area under the receiver operating characteristic curve; Carboxyterminal propeptides of procollagen type IV (PRO-C4) and VI (PRO-C6); CAP, Controlled attenuation parameter; CK-18, Cytokeratin-18; ELF, Enhanced Liver Fibrosis; FIB-4, Fibrosis-4; GBM, Gradient boosting method; HA, Hyaluronic acid; LITMUS, Liver Investigation: Testing Marker Utility in Steatohepatitis; LSM, Liver Stiffness Measurement; TIMP-1, Metalloproteinases 1; MICE, Multivariate imputation by chain equations; NAS, NAFLD activity score; NASH CRN, NASH Clinical Research Network; NITs, Noninvasive tests; VCTE, Vibration-Controlled Transient Elastography.

Quentin M. Anstee and Patrick M Bossuyt are joint senior auuthors.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.hepjournal.com.

NASH model produced AUCs (clinical/extended) of 0.61/0.65. The composite NASH model performed significantly better (0.71) for both variants. The composite at-risk NASH model had an AUC of 0.83 (clinical and extended), an improvement over the direct model. Significant fibrosis models had AUCs (clinical/extended) of 0.76/0.78. The extended advanced fibrosis model (0.86) performed significantly better than the clinical version (0.82).

Conclusions: Detection of NASH and at-risk NASH can be improved by constructing independent machine learning models for each component, using only clinical predictors. Adding biomarkers only improved the accuracy of fibrosis.

BACKGROUND

NAFLD is characterized by fat accumulation in hepatocytes. There is a need for more robust and accessible noninvasive tests (NITs), as NAFLD affects nearly 25% of the global population.[1,2] As a progressive condition, NAFLD ranges from isolated steatosis (liver fat content $\geq$ 5%) to NASH with or without fibrosis and cirrhosis.[3,4] NASH is associated with progression to liver fibrosis and HCC.[5] "At-risk" NASH (NASH with at least significant fibrosis) is an important target for drug development and the focus of health authorities, as it carries an increased risk of liver-related mortality and contributes significantly to the total burden of HCC.[6] In a prospective cohort study, the population with fibrosis stage 3 and higher had the greatest risk to develop liver endpoints, while fibrosis stage 2 and higher was linked to increased hepatic and extrahepatic morbidity.[7]

Liver biopsy remains the reference standard for a definitive NASH diagnosis; however, the procedure carries risks to the patient and has several inherent limitations, including sampling error and reader variability.[8,9] Even so, no NITs for NASH that match similar standards are available. This unmet clinical need has been the driving force for a marathon of research to develop and validate novel NITs that can distinguish patients with a greater likelihood of disease progression than those with comparable liver biopsy performance. Identifying those at higher risk is critical for risk stratification, monitoring, and expediting recruitment for NASH clinical trials.

The list of NITs for NAFLD fibrosis has rapidly grown, with the Liver Stiffness Measurement by Vibration-Controlled Transient Elastography (LSM by VCTE), Enhanced Liver Fibrosis (ELF) test, and Fibrosis-4 (FIB-4) score recommended to rule out advanced fibrosis.[10,11] However, the EASL Clinical Practice Guidelines currently do not recommend NITs for the diagnosis of NASH.[10] Extensively studied biomarkers such as caspase-cleaved cytokeratin-18 (CK-18) fragments and full-length soluble CK-18 show suboptimal performance, although combining CK-18 with synergistic markers showed some improvement.[12] Multivariable models developed using regression-based techniques, such as FIC-22,[13] the NAFLD diagnostic panel,[14] or the NASH test,[15] have either proved to be less effective in more extensive multicenter studies or have not undergone sufficient external validation. More recently, the MACK-3, FAST, and NIS4 scores were developed specifically for detecting at-risk NASH.[16–18]

While the list of NITs for NAFLD grows, few were developed based on machine learning algorithms, which are probably more suitable for handling complicated diseases with multifaceted etiology. Simple regression-based methods rely heavily on statistical assumptions, which do not always hold true for real-world data, whereas model-free machine learning algorithms adapt to data characteristics with fewer assumptions.

Machine learning uses algorithms to learn associations, identify patterns, and create predictions from complex data structures, which can provide opportunities for improving the diagnosis or prognosis of diseases. More recently, machine learning has been applied to develop diagnostic scores across multiple disciplines, offering a potential solution for developing tools for conditions that prove more difficult to detect.[19–21] Our aim was to employ machine learning to develop diagnostic models for detecting clinical trial definitions of NASH, at-risk NASH, and significant and advanced fibrosis, first by utilizing only routinely collected clinical data and second by adding biomarkers.

METHODS

This manuscript was prepared using the TRIPOD guidelines (Supplemental Table S2, http://links.lww.com/HEP/H625).[22]

Study participants (LITMUS metacohort)

We analyzed data from 966 participants in the Liver Investigation: Testing Marker Utility in Steatohepatitis

(LITMUS) metacohort.[23] These participants were recruited from 12 centers in 9 countries across Europe between 2010 and 2019 and included adults with biopsy-confirmed NAFLD with available clinical, laboratory, and biomarker data within 6 months of biopsy. Serum samples drawn within 6 months of biopsy and stored at −80°C were also available. Details of the study can be found elsewhere.[24] All participants provided informed consent before inclusion; the cohort studies were approved by the relevant ethics committees in the participating countries.

Liver biopsy

Biopsy samples were examined prospectively in each center by expert liver pathologists. NAFLD activity was graded according to the NASH Clinical Research Network (NASH CRN).[25] Liver fibrosis was graded on a 5-point scale (0 to 4), denoted as *F* in the following.

NASH is comprised of 3 components: steatosis, lobular inflammation, scored on 4-point scales (0-3), and ballooning on a 3-point scale (0-2) according to the NASH CRN classification.[25] The NAFLD activity score (NAS), the unweighted sum of steatosis, lobular inflammation, and ballooning scores thus ranges from 0 to 8.

Target conditions

This study addressed 4 target conditions:

i Significant fibrosis: Defined as $F \geq 2$;
ii Advanced fibrosis: Defined as $F \geq 3$;
iii Clinical Trial NASH: Steatohepatitis is a histopathological diagnosis based on the presence of steatosis, lobular inflammation, and hepatocyte ballooning.[26] For inclusion in therapeutic trials, the FDA and EMA mandate steatohepatitis is defined as a $NAS \geq 4$ with at least a score of 1 point for each histological component,[27] thus selecting patients with greater disease activity that are considered more likely to exhibit disease progression;[28] and
iv "At-risk" NASH: Like the above, "at-risk" NASH is defined as the presence of steatohepatitis ($NAS \geq 4$ with at least 1 point in each component) plus significant fibrosis ($F \geq 2$).[2,17,18] This defines the population commonly recruited into phase 3 trials of novel therapeutics for noncirrhotic NASH.

Predictors

Clinical assessment

Clinical data, including anthropometric, lifestyle/activity, dietary, comorbidity, pharmacotherapy, clinical biochemistry, and incident disease/events, were collected in the respective recruitment centers, with blood assays performed in local laboratories. The list of 25 clinical predictors used is shown in Supplemental Table S3, http://links.lww.com/HEP/H625.

Biomarker measurements

Additional serum samples were collected in standardized collection kits within 6 months of liver biopsy and stored at −80°C. Samples were centrally analyzed at Nordic Biosciences (Herlev, Denmark), a CLIA-certified laboratory, blinded to clinical data. The following markers were measured and included as predictors: caspase-cleaved CK-18 fragments and full-length soluble CK-18 (M30 and M65 antigens), serum peptides that represent the amino-terminal propeptide of procollagen type III (PRO-C3), and the carboxyterminal propeptides of procollagen type IV (PRO-C4) and VI (PRO-C6). We further include the components of the Siemens ELF test: tissue inhibitor of metalloproteinases 1 (TIMP-1), amino-terminal propeptide of type III procollagen (P3NP), and hyaluronic acid (HA).

LSM and controlled attenuation parameter (CAP) by VCTE (FibroScan, Echosens, Paris, France) were collected within 6 months of liver biopsy and also included as predictors. Probe sizes were selected as advised by device guidelines.

Machine learning algorithm

A variation of the gradient boosting machine (GBM) was used to develop the models. GBM is an ensemble machine learning technique for regression and classification to produce prediction models of multiple base learners (or decision trees). This algorithm involves 3 elements: optimization of a loss function, predictions made by a base learner, and an additive model to add base learners to minimize the loss function successively.

As GBM methods are known for overfitting, we applied stochastic GBM to reduce the correlation between trees in the sequence of GBM. Each iteration uses a sub-sample of the full training data set, drawn at random, used in place of the full data set to fit the base learner and compute the model update for the current iteration.[29] The randomized approach improves model robustness and reduces overfitting. Other GBM variations, such as XGBoost, were tested but were not an improvement from the stochastic method.

We explored alternative algorithms (logistic regression, k-nearest neighbors, support vector machine, and decision tree algorithms). Only the results for GBM were further evaluated as it produced the best-performing models in the preliminary analyses.

Data set preprocessing

The original metacohort data set underwent a lengthy preprocessing phase to convert raw data to the optimal structure for training and testing GBM models. As the original data set included over 200 clinical variables, we isolated those relevant to NAFLD based on clinical accessibility and established association guided by experienced hepatologists.

A pairwise Pearson correlation matrix was used to visualize the predictors' relationships and test for high intercorrelation. No variables were removed in this process, resulting in a final working data set of 35 predictors.

Missing data were handled by first assessing the degree of missingness and if data were missing at random. Variables with missing values for more than 80% of participants were excluded entirely. For the remaining variables, missing data were replaced by multiple imputations (m = 5) using the multivariate imputation by chain equations (MICE) approach.[30] As data were split prior to this step, the training and validation data were imputed separately, resulting in 5 imputed training and validation sets. By purpose, outcome variables were excluded from the predictor matrix in the validation set as we aimed to mimic a scenario where the model is used in a 'new' patient, where outcome data is obviously not available. In support of our strategy, simulation studies have shown that including outcomes when imputing the validation set leads to over-optimistic predictions.[31,32] We further excluded variables with over 60% missing from the predictor matrix.

Continuous variables were centered and scaled to a mean of 0 and SD of 1 to improve model stability and fit.

Model development

Figure 1 provides an overview of the model training and validation workflow. The learning data were randomly split into training (75%) and validation (25%) sets. The training set (n = 742) was used to develop models using the GBM algorithm for each target condition. A grid search strategy was applied to tune hyperparameters (boosting iterations, max tree depth, shrinkage, minimum terminal node size) using 5 repeats of 10-fold cross-validation. Agreement between the model prediction and the observed outcome was inspected visually using calibration plots for each of the constructed models. Two sets of models were developed for each target condition, one using only routinely available clinical predictors (clinical model) and a second employing these same routinely available clinical predictors plus additional biomarkers (extended model).

As discussed above, NASH is established by the presence of 3 histological features (steatosis, lobular inflammation, and ballooning). To address how these may best be combined, we developed 2 variants of the NASH models: 1 directly including these 3 histological features, the other by building a composite model that aggregated the calculated probabilities from models for steatosis, lobular inflammation, and ballooning (Figure 2):

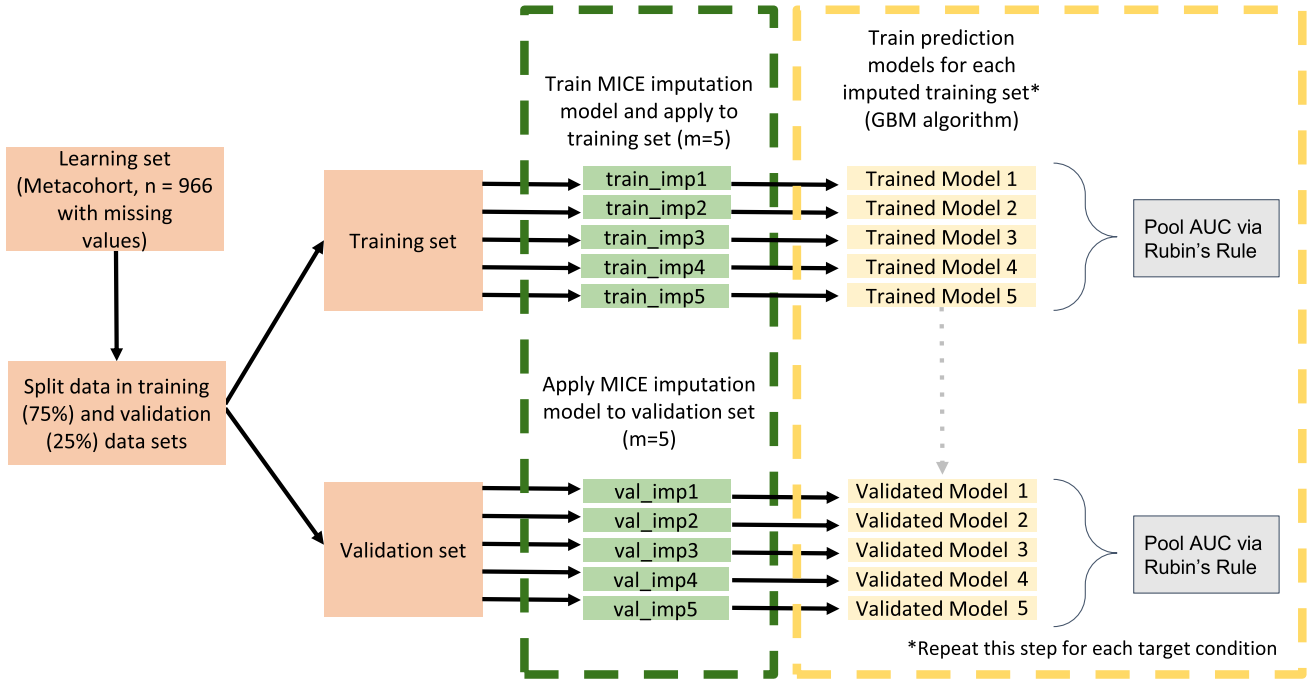

FIGURE 1 Model development workflow.

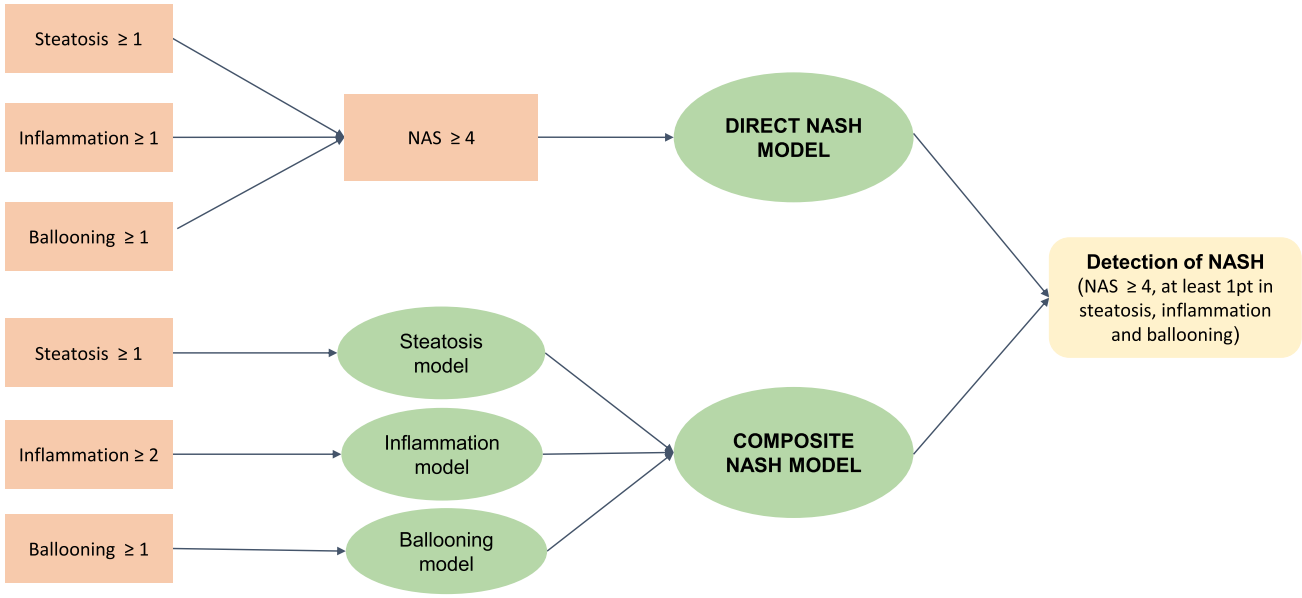

FIGURE 2 Construction of the direct and composite NASH models.

i The "direct" NASH model was trained to a NAS ≥ 4 with at least 1 point in each of the 3 components (S ≥ 1 + B ≥ 1 + LI ≥ 1, with the sum being ≥ 4).[33,34]

ii The "composite" NASH model was similarly trained to a NAS ≥ 4 but with an additional, more stringent, liver inflammation threshold of 2 points (S ≥ 1 + B ≥ 1 + LI ≥ 2, with the sum being ≥ 4). In the composite model, separate models were built for each histological feature, steatosis (0 vs. 1–3), lobular inflammation (0–1 vs. 2–3), and ballooning (0 vs. 1–2), and the respective probabilities for each component were multiplied to yield a NASH prediction index.

In the same way, 2 different "at-risk" NASH model variants were developed: a "direct" at-risk NASH model and a "composite" at-risk NASH model, the latter built by aggregating the calculated probabilities from the steatosis, lobular inflammation, ballooning, and significant fibrosis models as discussed above.

Statistical analysis

The performance of each model was evaluated in the validation data (n = 242), which were untouched and isolated from the model training process. The area under the receiver operating characteristic curve (AUC) in detecting the respective target conditions was calculated to express the accuracy of classifications against liver biopsy as the reference standard. As depicted in Figure 1, the model development and validation steps were repeated for each of the 5 imputed data sets, for each condition, and AUCs were pooled following Rubin's Rule.[35,36]

Irrespective of how the model had been trained, to ensure that the full spectrum of NAS-defined steatohepatitis was captured, the target definition of NASH (or "at-risk" NASH) used in the validation analyses was NAS of ≥ 4 with at least 1 point in each of the 3 components (S ≥ 1 + B ≥ 1 + LI ≥ 1, in any permutation where the sum is ≥ 4).

The extended GBM models were compared with other tests: CK-18 and CAP by VCTE for NASH, FAST score,[17] and ADAPT[37] for at-risk NASH, and PRO-C3, LSM by VCTE, the FIB-4 score,[38] and the ELF test[39] for fibrosis, according to their original formula.

Variable importance scores were calculated for the GBM models to rank selected predictors based on their relative importance (scaled between 0 and 100) for making more accurate predictions. This was determined based on the selection of variables in the tree-building process and improvement for each boosting iteration.[40]

All statistical analysis was performed using R software version 4.0.3. Multiple imputations was applied using the MICE package;[30] GBM models were trained using the caret package.[40]

RESULTS

The study group had a mean age of 51 and a mean body mass index of 34; 58% were men, and 42% had diabetes. Based on liver biopsy, 53% had NASH, 35% had at-risk NASH, 49% had significant fibrosis, and 28% had advanced fibrosis (including 7% patients with histological cirrhosis). Details are summarized in Table 1. The flow of participants included in the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) metacohort can be seen in Supplemental Figure S1, http://links.lww.com/HEP/H625.

TABLE 1 Characteristics of the study group in the training and validation sets

	Overall study group	Training set	Validation set
n	966	724	242
Age, y	51.19 (12.97)	51.80 (12.70)	49.37 (13.61)
Male, n (%)	563 (58.3)	416 (57.5)	147 (60.7)
BMI	34.08 (8.25)	34.10 (8.36)	34.03 (7.93)
ALT, U/L	62.67 (42.54)	62.25 (42.20)	63.92 (43.63)
AST, U/L	42.88 (26.01)	43.09 (26.80)	42.25 (23.55)
GGT, U/L	110.05 (160.19)	113.97 (172.46)	98.34 (115.50)
Albumin, g/L	4.39 (0.42)	4.38 (0.42)	4.41 (0.42)
Platelet, 10^9/L	238.96 (73.45)	237.36 (73.66)	243.76 (72.74)
Glucose, mmol/L	6.50 (2.57)	6.57 (2.65)	6.30 (2.31)
Triglyceride, mg/L	2.07 (1.21)	2.10 (1.26)	1.99 (1.05)
Diabetes, n (%)	406 (42.0)	318 (43.9)	88 (36.4)
FIB-4	1.38 (1.02)	1.41 (1.04)	1.29 (0.96)
VCTE-CAP	312.85 (73.19)	314.13 (71.71)	309.04 (77.46)
VCTE-LSM	11.47 (9.29)	11.15 (8.72)	12.44 (10.78)
Steatosis grade, % (0/1/2/3)	7/33/35/24/1	8/32/35/25/1	5/35/35/23/0
Steatosis, n (%)	898 (93.0)	670 (92.5)	228 (94.2)
Inflammation grade, % (0/1/2/3)	20/57/21/2	19/57/22/2	21/58/18/3
Inflammation, n (%)	223 (23.1)	172 (23.8)	51 (21.1)
Ballooning grade, % (0/1/2)	26/50/24	26/50/25	27/51/22
Ballooning, n (%)	715 (74.0)	539 (74.4)	176 (72.7)
NASH, n (%)	512 (53.0)	385 (53.2)	127 (52.5)
At-risk NASH, n (%)	335 (34.7)	260 (35.9)	75 (31.0)
Fibrosis stage, % (0/1/2/3/4)	32/19/21/20/9	32/18/22/20/8	34/24/15/17/10
Significant fibrosis, n (%)	471 (48.8)	368 (50.8)	103 (42.6)
Advanced fibrosis, n (%)	273 (28.3)	207 (28.6)	66 (27.3)

Notes: Continuous values are shown as mean (SD).
Steatosis is defined as 0 vs. 1–3, inflammation as 0–1 vs. 2–3, ballooning as 0 vs. 1–2, NASH as NAS $\geq$ 4 (with at least 1 point in each component), significant fibrosis as F$\geq$ 2, advanced fibrosis as F$\geq$ 3, and at-risk NASH is the combination of NASH and significant fibrosis.
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CAP, controlled attenuation parameter; FIB-4, Fibrosis-4; GGT, gamma-glutamyl transferase; LSM, liver stiffness measurement; VCTE, Vibration-controlled transient elastography

Steatosis, lobular inflammation, and ballooning models

Tuning parameters and results across each imputed training and validation set can be seen in Supplemental Table S4, http://links.lww.com/HEP/H625. The list of predictors selected for the clinical and extended models can be seen in Figure 3. Model calibration can be seen in Supplemental Figure S2, http://links.lww.com/HEP/H625.

In the validation set, the steatosis model had an AUC (95% CI) of 0.94 (0.93, 0.96) and 0.94 (0.92, 0.96) for the clinical and extended version, respectively, showing no improvement with the addition of biomarkers (Table 2). The inflammation models also had similar performance for the clinical and extended versions (AUC of 0.79 (0.76, 0.81) versus 0.79 (0.76, 0.82)). The GBM model for detecting ballooning had an AUC of 0.72 (0.69, 0.76) (clinical model) and 0.74 (0.70, 0.77) (extended model).

NASH models

NASH models were constructed following 2 different approaches. The direct NASH model had an AUC of 0.61 (0.57, 0.66) using only the clinical predictors and 0.65 (0.60, 0.69) when adding biomarkers (Table 2).

The second, composite NASH model, aggregated the predicted probabilities of the steatosis, inflammation, and ballooning models. This model performed significantly better than the direct model with an AUC of 0.71 (0.67, 0.74); additional biomarkers did not improve the performance.

In comparison, CAP by VCTE and CK-18 M30 had AUCs of 0.64 (0.59, 0.70) and 0.61 (0.57, 0.65),

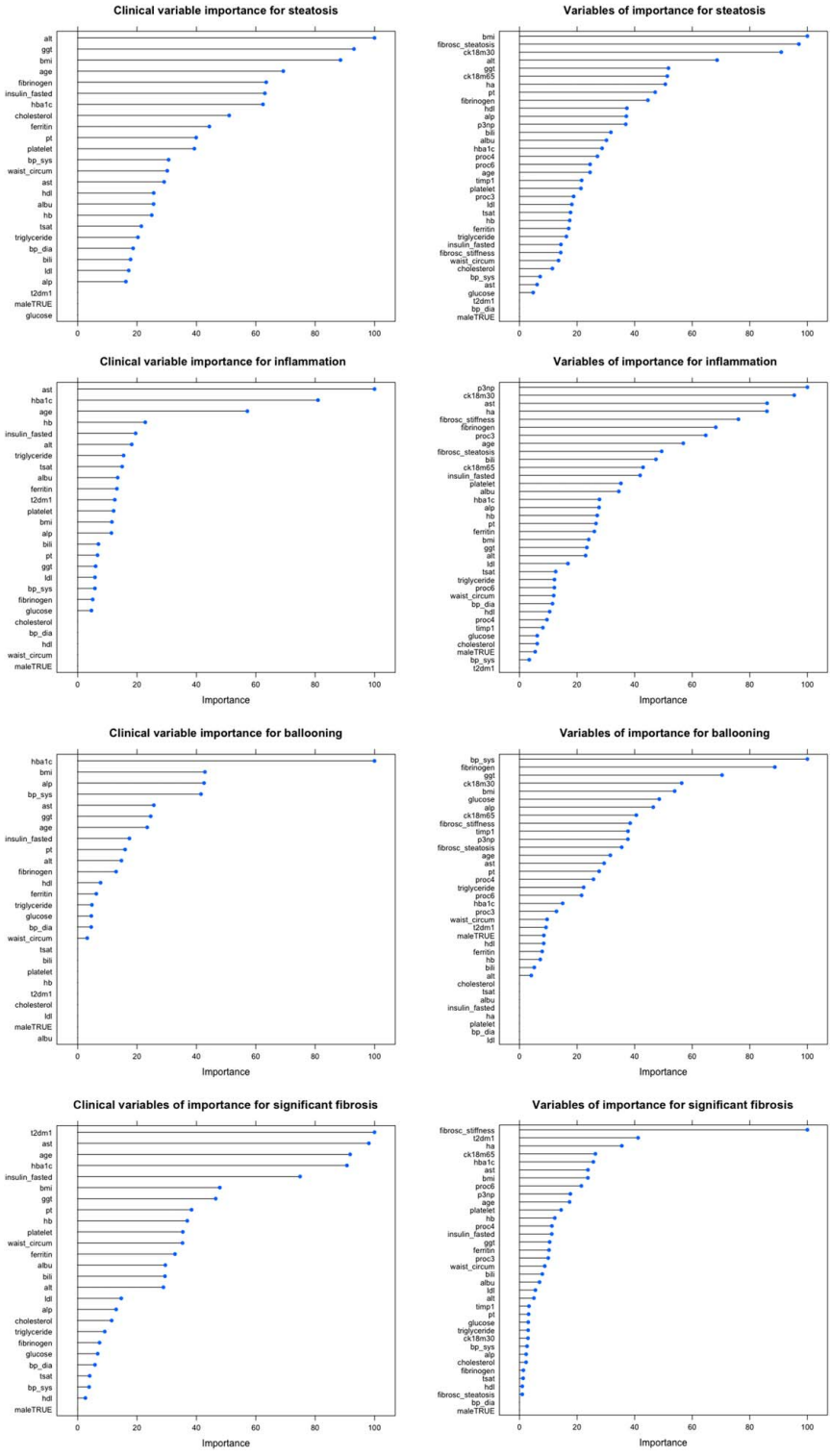

FIGURE 3 Variables of the importance for steatosis, inflammation, ballooning, and significant fibrosis for the clinical and extended GBM models systolic blood pressure (bp_sys), diastolic blood pressure (bp_dia), type 2 diabetes (t2dm), hdl, ldl, alanine aminotransferase (alt), aspartate aminotransferase (ast), gamma-glutamyl transferase (ggt), alkaline phosphatase (alp), hemoglobin (hb), transferrin saturation (tsat), albumin (albu), clotting (pt), bilirubin (bili), glycosylated hemoglobin A1c(hba1c), cytokeratin-18 (ck-18, m30 and m65 antigens), plasma propeptides of procollagen type III (pro-c3, pro-c4, pro-c6), tissue inhibitor of metalloproteinases 1 (timp-1), amino-terminal propeptide of type III procollagen (p3np), and hyaluronic acid (ha). Abbreviation: GBM, Gradient boosting method.

TABLE 2 Performance of the clinical and extended GBM models for detecting stages of NAFLD in the validation set

Outcome, model variant	Definition	Prevalence (%)	Clinical GBM model	Extended GBM model
Steatosis	0 vs. 1-3	93	0.94 (0.93, 0.96)	0.94 (0.92, 0.96)
Inflammation	0-1 vs. 2-3	23	0.79 (0.76, 0.81)	0.79 (0.76, 0.82)
Ballooning	0 vs. 1-2	74	0.72 (0.69, 0.76)	0.74 (0.70, 0.77)
NASH, composite	S * I * B	53	0.71 (0.67, 0.74)	0.71 (0.68, 0.77)
NASH, direct	NAS ≥ 4	—	0.61 (0.57, 0.66)	0.65 (0.60, 0.69)
At-risk NASH, composite	S * I * B * F	35	0.83 (0.80, 0.86)	0.83 (0.80, 0.86)
At-risk NASH, direct	NAS ≥ 4 and F ≥ 2	—	0.79 (0.76, 0.82)	0.78 (0.75, 0.82)
Significant fibrosis	F ≥ 2	47	0.76 (0.73, 0.80)	0.78 (0.75, 0.82)
Advanced fibrosis	F ≥ 3	28	0.82 (0.79, 0.84)	0.86 (0.85, 0.87)

Notes: Clinical GBM models include only clinical predictors, extended GBM models include clinical predictors and biomarkers.
Composite NASH model was constructed by aggregating the 3 NASH components: steatosis, lobular inflammation, and ballooning, which were dichotomized according to the definition as described. At-risk NASH was constructed similarly, including significant fibrosis (F ≥ 2).
Direct NASH model was constructed using the standard dichotomization of NAS score (≥ 4), with at least one point in each component of steatosis, lobular inflammation and ballooning.
Direct at-risk NASH is the combination of NAS score (≥ 4) and significant fibrosis (F ≥ 2).

respectively, for the detection of NASH (Figure 4). The composite NASH model was a significant improvement over CK-18.

At-risk NASH model

Two different at-risk NASH models were evaluated. The direct at-risk NASH model had an AUC of 0.79 (0.76, 0.82) using clinical variables and 0.78 (0.75, 0.82) for the extended model (Table 2).

For the composite at-risk NASH model, the AUCs were 0.83 (0.80, 0.86) for both the clinical and extended versions. See Figure 3 for the predictors selected for each model and aggregated to calculate the composite models.

The composite GBM models performed well compared with other multi-marker scores: FAST had an AUC of 0.77 (0.73, 0.81), and ADAPT had 0.77 (0.73, 0.80) (Figure 4).

Significant and advanced fibrosis model

The significant fibrosis models had AUCs of 0.76 (0.73, 0.80) and 0.78 (0.75, 0.82) for the clinical and extended versions, respectively. Both fibrosis model probabilities were very consistent with the observed event rates (see calibration plots in Supplemental Figure S3, http://links.

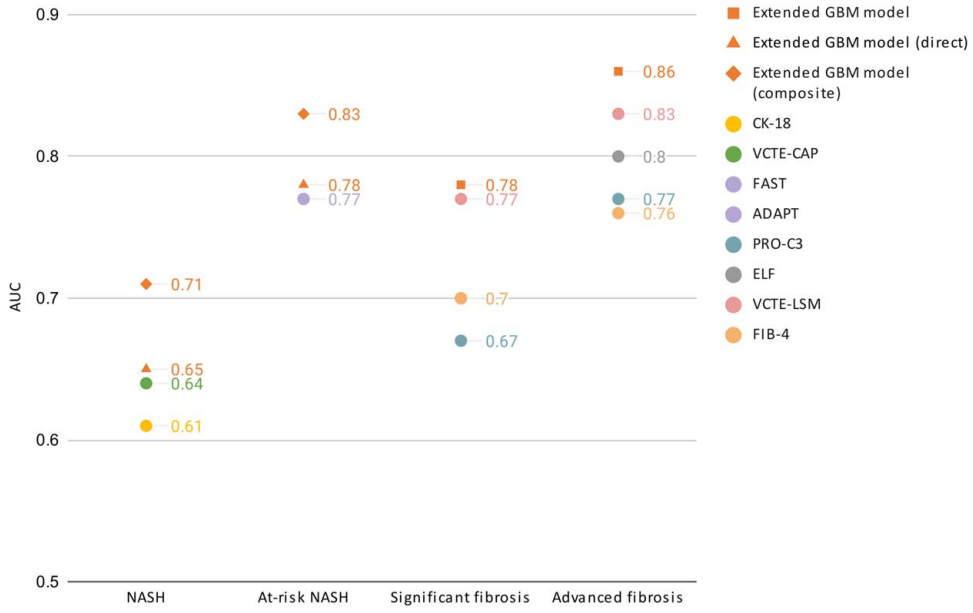

FIGURE 4 AUC of extended GBM models in the validation set compared with existing noninvasive scores for detecting NASH, at-risk NASH, and significant and advanced fibrosis. Abbreviation: GBM, Gradient boosting method.

lww.com/HEP/H625). Tuning parameters for each imputed data set are shown in Supplemental Table S4, http://links.lww.com/HEP/H625.

In comparison, PRO-C3 had an AUC of 0.67 (0.63, 0.71), LSM by VCTE had 0.77 (0.73, 0.81), FIB-4 had 0.70 (0.66, 0.73), and ELF had 0.70 (0.66, 0.73) (Figure 4).

For advanced fibrosis, the AUC for the clinical model was 0.82 (0.79, 0.84). Adding biomarkers significantly improved the detection of advanced fibrosis, with an AUC of 0.86 (0.85, 0.87).

For advanced fibrosis, PRO-C3 had an AUC of 0.77 (0.73, 0.81), LSM by VCTE had 0.83 (0.80, 0.87), FIB-4 had 0.76 (0.73, 0.80), and ELF had 0.80 (0.76, 0.83) (Figure 4).

DISCUSSION

Several diagnostic scores have been studied to identify patients with advanced stages of fibrosis. However, those for detecting active NASH (with or without fibrosis) have been less successful. The present study utilized a large histologically characterized NAFLD cohort in Europe with a rich selection of novel biomarkers to develop diagnostic models using the GBM algorithm. Two sets of models were developed for each condition, 1 using only clinical features, and a second by adding biomarkers, such as CK-18, PRO-C3/4/6, and LSM and CAP by VCTE.

We explored the added value of fitting the GBM algorithm for steatosis, inflammation, and ballooning separately and creating an aggregate model combining the 3 components. The purpose was to enhance classifications, as NASH models are generally developed solely based on the NAS score and so far, none are suggested for use by clinical guidelines.[10] Our results showed that aggregating the probabilities for each component to arrive at the composite NASH score significantly improved the accuracy of detecting NASH. The same strategy also improved the detection of at-risk NASH.

The performance of the models for NASH and at-risk NASH was comparable between the clinical and extended models. The fibrosis models benefited the most from the additional biomarkers, with significant improvement in detecting advanced fibrosis.

The study presents some limitations, mostly related to the retrospective nature of the LITMUS Metacohort. Liver biopsy serves as the reference standard in our analysis as it remains the recommended technique for evaluating NASH, despite caveats such inter-reader and intra- reader variability.[8,41] We further note that our definition of NASH corresponds to the efficacy endpoint defined by health authorities and clinical trials for NAFLD drug development, which may differ from the clinical diagnosis of NASH that considers other inputs in addition to histopathologic diagnosis.[34] We relied on locally read biopsies and local lab results for standard markers, which may introduce differences across study sites. While blood-based biomarkers were centrally analyzed, they were measured in retrospectively collected samples and reserved in a biobank. They were also analyzed in batches.

Due to the limited sample volume, not all biomarkers were measured in all patients. We avoided complete case analysis by imputing missing values by multiple imputations. Five imputed data sets were produced and analyzed as simulation studies have shown a required number of repeated imputations to be as low as 3 for data sets with 20% missingness.[42] The retrospective nature of this study also meant that some samples were older than others. As the stability of these samples is largely unknown, we excluded a handful of samples collected before 2010 from the analysis.

Other machine learning models have been developed using only clinical predictors to detect NASH. Using a variant of GBM, 1 study found out-of-sample AUCs of 0.82 and 0.76, using data from the National Institute of Diabetes, Digestive and Kidney Diseases, and Optum Analytics.[21] Another study applied machine learning algorithms to predict NASH using data from Optum Analytics and found the highest AUC (0.88) using XGBoost.[43] This study, however, included healthy participants without any liver-related diseases in the model development phase. Both studies relied on data from Optum, which, in the absence of a histological diagnosis of NASH, relied on several different ICD codes for NASH or NAFLD.

Other scores have been developed using regression-based methods, such as NIS4 and MACK-3. NIS4, which includes 4 components (miR-34a-5p, alpha-2 macroglobulin, YKL-40, and glycated hemoglobin), had an AUC of 0.80 from 3 validation cohorts for detecting at-risk NASH.[18] An external validation study found that MACK-3 (fasting glucose and insulin, AST, and CK-18) also had an AUC of 0.80 for detecting at-risk NASH. More recently, the SomaSignal test developed based on elastic net produced an AUC of 0.76.[44,45] All of these multi-marker scores include more novel biomarkers, which come with the cost of additional testing. Our at-risk NASH model performed well, relying only on clinical data, highlighting a potential advantage of utilizing machine learning. This warrants further evaluation in an external cohort.

Constructing separate models for each component of NASH further allowed us to observe that different predictors were selected as most informative for each component. The most influential predictors had strong biological plausibility or an established position in the disease pathway.[12,46–48] However, the ranking of markers was variable across imputed data sets and should be interpreted with caution.

In the future, we plan to finalize the models using complete data from the ongoing prospective LITMUS

Study Cohort, focusing on the aggregated approach for constructing the models for NASH and at-risk NASH. A single model for each outcome will be converted to a user-friendly interface in the form of a Shiny app. Such tools would allow clinicians, including those in a primary care setting, to enter values of clinical parameters to detect NASH or at-risk NASH with greater ease.

Machine learning approaches are sometimes perceived as too complicated compared with classic regression-based tools. Some studies have demonstrated the superior performance of machine learning algorithms over logistic regression, such as Feng et al, who found machine learning models outperformed regression-based models for detecting significant fibrosis across different subgroups.[49] However, a large meta-analysis found no benefit of machine learning over logistic regression.[50] Given the vast selection of available algorithms, heterogeneous study designs, sparse reporting, and conflicting conclusions in the literature, more work is needed to understand which tools and study design elements are optimal for developing diagnostic models for NAFLD. This should be paired with a clear emphasis on the tool's desired clinical context of use, whether to triage patients in clinical practice or select participants most likely to benefit from therapeutic interventions in clinical trials.

Our study found promising results to explore machine learning algorithms further to improve the diagnosis of NASH and at-risk NASH, using readily available clinical data. The inherent ability to adapt to new data positions machine learning as a valuable tool for rapidly evolving health care settings and conditions with a complex etiology such as NAFLD. While the move towards machine learning to detect NAFLD is still in its infancy, concerted efforts to robust methodology, biomarker discovery, and quality data can improve the clinical management of NAFLD. Importantly, this is most needed outside expert centers, where the vast majority of patients do not have access to specialists focusing on liver disease.

AUTHOR CONTRIBUTIONS

Jenny Lee, Quentin M. Anstee, and Pierre Bedossa conceptualized and designed the study; Jerome Boursier, Salvatore Petta, Kristy Wonders, Dina Tiniakos, Pierre Bedossa, Andreas Geier, Sven Francque, Mike Allison, Georgios Papatheodoridis, Helena Cortez-Pinto, Diana Julie Leeming, Stephen Harrison, Jeremy Cobbold, Michael Pavlides, Adriaan G. Holleboom, Hannele Yki-Jarvinen, Javier Crespo, Morten Karsdal, Rachel Ostroff, Carla Yunis, Clifford Brass, Mattias Ekstedt, Guruprasad P Aithal, Jorn M. Schattenberg, Elisabetta Bugianesi, Manuel Romero-Gomez, Richard Torstenson, Vlad Ratziu, Quentin M. Anstee assisted in data collection; Jenny Lee, Yasaman Vali, Pierre Bedossa, Kristy Wonders, and Quentin M. Anstee accessed and verified the raw data; Jenny Lee performed data analyses and drafted the manuscript; Patrick M Bossuyt supervised the data analyses and Max Westphal, Koos Zwinderman, Jerome Boursier, Yu Chen, Rachel Ostroff, Quentin M. Anstee, and Vlad Ratziu commented on the data analyses; all authors had full access to all the data in the study and had final responsibility for the decision to submit for publication and critically revised the manuscript and approved of the final version.

ACKNOWLEDGMENTS

The full list of LITMUS investigators is available in Supplemental Table S1, http://links.lww.com/HEP/H625.

FUNDING INFORMATION

The LITMUS project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 777377. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. Sven Francque holds a senior clinical investigator fellowship from the Research Foundation Flanders (FWO) (1802154N).

CONFLICTS OF INTEREST

Jerome Boursier received grants from Echosens. Rachel Ostroff owns stock in and is employed by SomaLogic. Leigh Alexander owns stock in and is employed by SomaLogic. Yu Chen owns stock in Eli Lilly and Company. Celine Fournier is employed by Echosens. Sven Francque's institution has received grants from Astellas, Falk Pharma, Genfit, Gilead Sciences, GlympsBio, Janssens Pharmaceutica, Inventiva, Merck Sharp & Dome, Pfizer, and Roche. He consults for and is on the speakers' bureau for Abbvie, Allergan, Bayer, Eisai, Genfit, Gilead Sciences, Intercept, Inventiva, Merck Sharpe & Dome, Novo Nordisk, and Promethera. He also consults for Actelion, Aelin Therapeutics, Aligos Therapeutics, Astellas, AstraZeneca, Boehringer Ingelheim, Bristoll-Meyers Squibb, CSL Behring, Coherus, Echosens, Enyo, Galapagos, Galmed, Genetech, Janssens Pharmaceutica, Julius Clinical, Madrigal, Medimmune, NGMBio, Novartis, and Roche. He is on the speaker's bureau for Janssens Cilag. Pierre Bedossa is the director of LIVERPAT. Mike Allison consults for and received grants from AstraZeneca and GSK. He received grants from Takeda. George Papatheodoridis advises, is on the speakers' bureau, and received grants from Gilead Sciences. He advises and is on the speakers' bureau for GlaxoSmithKline and Novo Nordisk. Jean-Francois Dufour consults for Intercept. He advises Madrigal. He received grants from Gilead. Diana Julie Leeming owns stock for and is employed by Nordic Bioscience. Stephen A. Harrison owns stock in, consults for, advises, and is a principal investigator of grant research for Akero Therapeutics, Genfit Corp, Hepion Pharmaceuticals Inc, Metacrine Inc., NGM Biopharmaceuticals, Inc., and NorthSea Therapeutics.

He owns stock in, consults for, and advises Galectin Therapeutics, Inc, HistoIndex PTE LTD, and Sonic Incytes Medical Corp. He consults for, advises, and is a principal investigator of grant research for Axcella Health Inc., Cymabay Therapeutics Inc., Hightide Therapeutics, Inc., Intercept Pharmaceuticals Inc., Madrigal Pharmaceuticals, Novo Nordisk, Poxel, and Sagimet Biosciences. He owns stock and advises Chronwell Inc. He owns stock in and is a principal investigator of grant research for Cirius Therapeutics, Inc. He consults for and advises Altimmune, Echosens North America Inc., Forsite Labs, LLC, HistoIndex PTE LTD, Medpace Inc., Prometic Pharma SMT LTD, Ridgeline, and Terns. He consults for and is a principal investigator of grant research for Enyo Pharma S.A. and Viking Therapeutics, Inc. He advises and is a principal investigator of grant research for Galmed Research & Dev. LTD., Gilead Sciences, CiVi Biopharma Inc., and Novartis Pharmaceuticals. He consults for AgomAB, Alentis Therapeutics AG, Altimentiv, Inc., Boston Pharmaceuticals, B Riley FBR Inc., BVF Partners LP, Cohbar, Inc., Canfite, Corcept Therapeutics, Inc., Fibronostics, Fortress Biotech, Inc., GNS, Inipharm, Ionis, Kowa Research Institute, Microba, Nutrasource, Perspectum Diagnostics, and Piper Sandler. He advises 89Bio, Arrowhead, Indalo, PathAI, and Theratechnologies. Javier Crespo consults for, advises, and received grants from Gilead, AbbVie, and Janssen. Mattias Ekstedt advises ARMA Medical. Guruprasad P. Aithal consults and advises for Clinipace, Amryth, AstraZeneca, GSK, DNDi, Albireo Pharma Inc, and Benevolent AI Bio. Elisabetta Bugianesi consults for and advises Novo Nordisk. She consults for and is on the speakers' bureau for MSD. She received grants from Gilead. Richard Torstenson is employed by and owns stock in AstraZeneca. Morten Karsdal owns stock in and is employed by Nordic Bioscience. Carla Yunis owns stock in and is employed by Pfizer, Inc. She owns stock in Inventiva and Amgen. Jorn M. Schattenberg consults for, is on the speakers' bureau for, and received grants from Boehringer Ingelheim. He consults for and is on the speakers' bureau for Gilead Sciences. He is on the speakers' bureau for and received grants from Novo Nordisk and Madrigal. He consults for Apollo Endosurgery, Albireo Pharma Inc, Bayer, GSK, Intercept Pharmaceuticals, Ipsen, Inventiva Pharma, MSD, Northsea Therapeutics, Novartis, Pfizer, Roche, Sanofi, and Siemens Healthineers. He is on the speakers' bureau for Siemens Health care GmbH. He received grants from Echosens, and MedPublico GmbH. Vlad Ratziu consults for Novo Nordisk, Terns, Madrigal, Enyo, Poxel, BMS, Intercept, and NGM. Clifford A. Brass owns stock in Novartis. Michael Pavlides owns stock in Perspectum LTD. He is on the speakers' bureau for and advises Intercept, Orphalan, Novo Nordisk, Roche Portugal and EISAI. Quentin M. Anstee, on behalf of Newcastle University, consults for Alimentiv, Akero, AstraZeneca, Axcella, 89Bio, Boehringer Ingelheim, Bristol Myers Squibb, Galmed, Genfit, Genentech, Gilead, GlaxoSmithKline, Hanmi, HistoIndex, Intercept, Inventiva, Ionis, IQVIA, Janssen, Madrigal, Medpace, Merck, NGM Bio, Novartis, Novo Nordisk, PathAI, Pfizer, Prosciento, Poxel, Resolution Therapeutics, Roche, Ridgeline Therapeutics, RTI, Shionogi, and Terns. He is on the speakers' bureau for Fishawack, Integritas Communications, Kenes, Novo Nordisk, Madrigal, Medscape, and Springer Health care. He received grants from AbbVie, AstraZeneca, Boehringer Ingelheim, Glympse Bio, Intercept Novartis, and Pfizer. He received royalties from Elsevier Ltd. He, on behalf of Newcastle University, serves on the DSMB for Medpace (NorthSea Therapeutics). The remaining authors have nothing to report.

ORCID

Jenny Lee https://orcid.org/0000–0003–4024–0933
Max Westphal https://orcid.org/0000–0002–8488–758X
Yasaman Vali https://orcid.org/0000–0001–7002–118X
Jerome Boursier https://orcid.org/0000–0002–7282–1436
Salvatorre Petta https://orcid.org/0000–0002–0822–9673
Rachel Ostroff https://orcid.org/000–0002–6867–5464
Leigh Alexander https://orcid.org/0000–0003–3469–1308
Celine Fournier https://orcid.org/0000–0003–1934–4965
Andreas Geier https://orcid.org/0000–0002–9626–5083
Sven Francque https://orcid.org/0000–0002–7527–4714
Kristy Wonders https://orcid.org/0000–0001–6407–9721
Dina Tiniakos https://orcid.org/0000–0003–4657–7780
Pierre Bedossa https://orcid.org/0000–0002–8487–7322
Georgios Papatheodoridis https://orcid.org/0000–0002–3518–4060
Helena Cortez-Pinto https://orcid.org/0000–0002–8537–8744
Raluca Pais https://orcid.org/0000–0001–8070–0733
Jean-Francois Dufour https://orcid.org/0000–0002–8062–1346
Diana Julie Leeming https://orcid.org/0000–0002–4256–140X
Stephen Harrison https://orcid.org/0000–0001–6399–5744
Adriaan G. Holleboom https://orcid.org/0000–0002–2911–2917
Hannele Yki-Järvinen https://orcid.org/0000–0001–6766–1549

Javier Crespo https://orcid.org/0000–0001–8248–0172
Mattias Ekstedt https://orcid.org/0000–0002–5590–8601
Guruprasad P. Aithal https://orcid.org/0000–0003–3924–4830
Elisabetta Bugianesi https://orcid.org/0000–0002–0502–4381
Richard Torstenson https://orcid.org/*
Morten Karsdal https://orcid.org/0000–0001–5026–8740
Jörn M. Schattenberg https://orcid.org/0000–0002–4224–4703
Detlef Schuppan https://orcid.org/0000–0002–4972–1293
Vlad Ratziu https://orcid.org/0000–0002–6865–3791
Koos Zwinderman https://orcid.org/0000–0003–0361–3139
Michael Pavlides https://orcid.org/0000–0001–9882–8874
Quentin M. Anstee https://orcid.org/0000–0002–9518–0088
Patrick M. Bossuyt https://orcid.org/0000–0003–4427–0128

REFERENCES

1. Younossi Z, Henry L. The Burden of NAFLD Worldwide. Non-Alcoholic Fatty Liver Disease: Springer; 2020:15–24.
2. Noureddin M, Truong E, Gornbein JA, Saouaf R, Guindi M, Todo T, et al. MRI-based (MAST) score accurately identifies patients with NASH and significant fibrosis. J Hepatol. 2022;76:781–7.
3. Satapathy SK, Sanyal AJ. Epidemiology and natural history of nonalcoholic fatty liver disease Seminars in liver disease. Thieme Medical Publishers; 2015.
4. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: Trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol. 2018;15:11–20.
5. Younossi Z, Stepanova M, Ong JP, Jacobson IM, Bugianesi E, Duseja A, et al. Nonalcoholic steatohepatitis is the fastest growing cause of hepatocellular carcinoma in liver transplant candidates. Clin Gastroenterol Hepatol. 2019;17:748–55. e3.
6. Ascha MS, Hanouneh IA, Lopez R, Tamimi TAR, Feldstein AF, Zein NN. The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis. Hepatology. 2010;51:1972–8.
7. Sanyal AJ, Van Natta ML, Clark J, Neuschwander-Tetri BA, Diehl A, Dasarathy S, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N Engl J Med. 2021;385:1559–69.
8. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology. 2005;128:1898–906.
9. Brunt EM, Clouston AD, Goodman Z, Guy C, Kleiner DE, Lackner C, et al. Complexity of ballooned hepatocyte feature recognition: Defining a training atlas for artificial intelligence-based imaging in NAFLD. J Hepatol. 2022;76:1030–41.
10. Berzigotti A, Tsochatzis E, Boursier J, Castera L, Cazzagon N, Friedrich-Rust M, et al. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis–2021 update. J Hepatol. 2021;75:659–89.
11. Mózes FE, Lee JA, Selvaraj EA, Jayaswal ANA, Trauner M, Boursier J, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: An individual patient data meta-analysis. Gut. 2022;71:1006–9.
12. Lee J, Vali Y, Boursier J, Duffin K, Verheij J, Brosnan MJ, et al. Accuracy of cytokeratin 18 (M30 and M65) in detecting non-alcoholic steatohepatitis and fibrosis: A systematic review and meta-analysis. PLoS ONE. 2020;15:e0238717.
13. Tada T, Kumada T, Toyoda H, Saibara T, Ono M, Kage M. New scoring system combining the FIB-4 index and cytokeratin-18 fragments for predicting steatohepatitis and liver fibrosis in patients with nonalcoholic fatty liver disease. Biomarkers. 2018;23:328–4.
14. Younossi ZM, Page S, Rafiq N, Birerdinc A, Stepanova M, Hossain N, et al. A biomarker panel for non-alcoholic steatohepatitis (NASH) and NASH-related fibrosis. Obes Surg. 2011;21:431–9.
15. Anty R, Iannelli A, Patouraux S, Bonnafous S, Lavallard V, Senni-Buratti M, et al. A new composite model including metabolic syndrome, alanine aminotransferase and cytokeratin-18 for the diagnosis of non-alcoholic steatohepatitis in morbidly obese patients. Aliment Pharmacol Ther. 2010;32(11-12):1315–22.
16. Boursier J, Anty R, Vonghia L, Moal V, Vanwolleghem T, Canivet C, et al. Screening for therapeutic trials and treatment indication in clinical practice: MACK-3, a new blood test for the diagnosis of fibrotic NASH. Aliment Pharmacol Ther. 2018;47:1387–96.
17. Newsome PN, Sasso M, Deeks JJ, Paredes A, Boursier J, Chan W-K, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: A prospective derivation and global validation study. Lancet Gastroenterol Hepatol. 2020;5:362–73.
18. Harrison SA, Ratziu V, Boursier J, Francque S, Bedossa P, Majd Z, et al. A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis: A prospective derivation and global validation study. Lancet Gastroenterol Hepatol. 2020;5:970–85.
19. Adamichou C, Genitsaridi I, Nikolopoulos D, Nikoloudaki M, Repa A, Bortoluzzi A, et al. Lupus or not? SLE Risk Probability Index (SLERPI): A simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus. Ann Rheum Dis. 2021;80:758–66.
20. Karaglani M, Gourlia K, Tsamardinos I, Chatzaki E. Accurate blood-based diagnostic biosignatures for Alzheimer's disease via automated machine learning. J Clin Med. 2020;9:3016.
21. Docherty M, Regnier SA, Capkun G, Balp M-M, Ye Q, Janssens N, et al. Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. J Am Med Inform Assoc. 2021;28:1235–41.
22. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162:W1–73.
23. Vali Y, Lee J, Boursier J, Petta S, Wonders K, Tiniakos D, et al. Liver Investigation: Testing Marker Utility in Steatohepatitis (LITMUS) consortium investigators. Biomarkers for staging fibrosis and non-alcoholic steatohepatitis in non-alcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. Lancet Gastroenterol Hepatol. doi:10.1016/S2468-1253(23)00017-1. Epub ahead of print. PMID: 36958367.
24. Hardy T, Wonders K, Younes R, Aithal GP, Aller R, Allison M, et al. The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. Contemp Clin Trials. 2020;98:106175.
25. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological

scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41:1313–21.
26. Bedossa P. Diagnosis of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis: Why liver biopsy is essential. Liver Inter. 2018;38(S1):64–6.
27. Anania FA, Dimick-Santos L, Mehta R, Toerner J, Beitz J. Nonalcoholic steatohepatitis: Current thinking from the Division of Hepatology and Nutrition at the Food and Drug Administration. Hepatology. 2021;73:2023–7.
28. Ratziu V, Harrison SA, Francque S, Bedossa P, Lehert P, Serfaty L, et al. Elafibranor, an agonist of the peroxisome proliferator−activated receptor−α and −δ, induces resolution of nonalcoholic steatohepatitis without fibrosis worsening. Gastroenterology. 2016;150:1147–59.e5.
29. Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal. 2002;38:367–78.
30. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45:1–67.
31. Hoogland J, van Barreveld M, Debray TPA, Reitsma JB, Verstraelen TE, Dijkgraaf MGW, et al. Handling missing predictor values when validating and applying a prediction model to new patients. Stat Med. 2020;39:3591–607.
32. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. BMC Med Res Methodol. 2014;14:1–13.
33. Sanyal AJ, Brunt EM, Kleiner DE, Kowdley KV, Chalasani N, Lavine JE, et al. Endpoints and clinical trial design for non-alcoholic steatohepatitis. Hepatology. 2011;54:344–53.
34. Noncirrhotic nonalcoholic steatohepatitis with liver fibrosis: Developing drugs for treatment. In: Services DoHaH, (CDER) CfDEaR, editors.
35. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. BMC Med Res Methodol. 2009;9:57.
36. Heymans M, Eekhout I. Applied missing data analysis with SPSS and (R) Studio. Amsterdam, The Netherlands: Heymans and Eekhout; 2019:20; Available online: https://bookdownorg/mwheymans/bookmi/[accessed 23 May 2020].
37. Daniels SJ, Leeming DJ, Eslam M, Hashem AM, Nielsen MJ, Krag A, et al. ADAPT: An Algorithm Incorporating PRO-C3 Accurately Identifies Patients With NAFLD and Advanced Fibrosis. Hepatology. 2019;69:1075–86.
38. Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. Hepatology. 2007;46:32–6.
39. Day JW, Rosenberg WM. The enhanced liver fibrosis (ELF) test in diagnosis and management of liver fibrosis. Br J Hosp Med (Lond). 2018;79:694–9.
40. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28:1–26.
41. Davison BA, Harrison SA, Cotter G, Alkhouri N, Sanyal A, Edwards C, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol. 2020;73:1322–32.
42. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med. 1999;18:681–94.
43. Fialoke S, Malarstig A, Miller MR, Dumitriu A. Application of machine learning methods to predict non-alcoholic steatohepatitis (NASH) in non-alcoholic fatty liver (NAFL) patients. AMIA Annu Symp Proc. 2018;2018:430–9.
44. Vali Y, Lee J, Schattenberg J, Gomez MR, Tiniakos D, Bedossa P, et al. Comparative diagnostic accuracy of blood-based biomarkers for diagnosing NASH: Phase 1 results of the LITMUS project. International Liver Congress; 2021.
45. Sanyal AJ, Williams YA, Lavine JE, Neuschwander-Tetri BA, Alexander L, Ostroff R, et al. Defining the serum proteomic signature of hepatic steatosis, inflammation, ballooning and fibrosis in non-alcoholic fatty liver disease. J Hepatol. 2023;78:693–703.
46. Tanwar S, Trembling PM, Guha IN, Parkes J, Kaye P, Burt AD, et al. Validation of terminal peptide of procollagen III for the detection and assessment of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease. Hepatology. 2013;57:103–11.
47. Darweesh SK, AbdElAziz RA, Abd-ElFatah DS, AbdElazim NA, Fathi SA, Attia D, et al. Serum cytokeratin-18 and its relation to liver fibrosis and steatosis diagnosed by FibroScan and controlled attenuation parameter in nonalcoholic fatty liver disease and hepatitis C virus patients. Eur J Gastroenterol Hepatol. 2019;31:633–41.
48. Feldstein AE, Alkhouri N, De Vito R, Alisi A, Lopez R, Nobili V. Serum cytokeratin-18 fragment levels are useful biomarkers for nonalcoholic steatohepatitis in children. Am J Gastroenterol. 2013;108:1526–31.
49. Feng G, Zheng KI, Li Y-Y, Rios RS, Zhu P-W, Pan X-Y, et al. Machine learning algorithm outperforms fibrosis markers in predicting significant fibrosis in biopsy-confirmed NAFLD. J Hepatobiliary Pancreat Sci. 2021;28:593–603.
50. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12–22.

How to cite this article: Lee J, Westphal M, Vali Y, Boursier J, Ostroff R, Alexander L, et al. Machine learning algorithm improves the detection of NASH (NAS-based) and at-risk NASH: a development and validation study. Hepatology. 2023;78:258–271. https://doi.org/10.1097/HEP.0000000000000364