

# World Journal of *Hepatology*

*World J Hepatol* 2021 October 27; 13(10): 1203-1458



### EDITORIAL

- 1203** Transition of an acronym from nonalcoholic fatty liver disease to metabolic dysfunction-associated fatty liver disease  
*Alam S, Fahim SM*

### OPINION REVIEW

- 1208** Non-invasive real-time assessment of hepatic macrovesicular steatosis in liver donors: Hypothesis, design and proof-of-concept study  
*Rajamani AS, Rammohan A, Sai VR, Rela M*

### REVIEW

- 1215** Impact of COVID-19 pandemic on liver, liver diseases, and liver transplantation programs in intensive care units  
*Omar AS, Kaddoura R, Orabi B, Hanoura S*
- 1234** In the era of rapid mRNA-based vaccines: Why is there no effective hepatitis C virus vaccine yet?  
*Echeverría N, Comas V, Aldunate F, Perbolianachis P, Moreno P, Cristina J*
- 1269** Pediatric non-cirrhotic portal hypertension: Endoscopic outcome and perspectives from developing nations  
*Sarma MS, Seetharaman J*

### MINIREVIEWS

- 1289** Acute-on-chronic liver failure in children  
*Islek A, Tumgor G*
- 1299** Coronavirus disease 2019 in liver transplant patients: Clinical and therapeutic aspects  
*Loinaz-Seguro C, Marcacuzco-Quinto A, Fernández-Ruiz M*
- 1316** Pediatric vascular tumors of the liver: Review from the pathologist's point of view  
*Cordier F, Hoorens A, Van Dorpe J, Creyten D*
- 1328** Autoimmune hepatitis in genetic syndromes: A literature review  
*Capra AP, Chiara E, Briuglia S*
- 1341** Assessing the prognosis of cirrhotic patients in the intensive care unit: What we know and what we need to know better  
*da Silveira F, Soares PHR, Marchesan LQ, da Fonseca RSA, Nedel WL*
- 1351** Liver transplantation for pediatric inherited metabolic liver diseases  
*Vimalasvaran S, Dhawan A*

- 1367 Liver and COVID-19: From care of patients with liver diseases to liver injury

*Gaspar R, Castelo Branco C, Macedo G*

## ORIGINAL ARTICLE

### Basic Study

- 1378 Direct modulation of hepatocyte hepcidin signaling by iron

*Yu LN, Wang SJ, Chen C, Rausch V, Elshaarawy O, Mueller S*

- 1394 Serum zonulin levels in patients with liver cirrhosis: Prognostic implications

*Voulgaris TA, Karagiannakis D, Hadziyannis E, Manolakopoulos S, Karamanolis GP, Papatheodoridis G, Vlachogiannakos J*

### Retrospective Cohort Study

- 1405 Impact of biliary complications on quality of life in live-donor liver transplant recipients

*Guirguis RN, Nashaat EH, Yassin AE, Ibrahim WA, Saleh SA, Bahaa M, El-Meteini M, Fathy M, Dabbous HM, Montasser IF, Salah M, Mohamed GA*

### Retrospective Study

- 1417 Machine learning models for predicting non-alcoholic fatty liver disease in the general United States population: NHANES database

*Atsawarungruangkit A, Laoveeravat P, Promrat K*

- 1428 Acute liver failure with hemolytic anemia in children with Wilson's disease: Genotype-phenotype correlations?

*Pop TL, Grama A, Stefanescu AC, Willheim C, Ferenci P*

### Observational Study

- 1439 Clinical outcomes of patients with two small hepatocellular carcinomas

*Pham AD, Vaz K, Ardalan ZS, Sinclair M, Apostolov R, Gardner S, Majeed A, Mishra G, Kam NM, Patwala K, Kutaiba N, Arachchi N, Bell S, Dev AT, Lubel JS, Nicoll AJ, Sood S, Kemp W, Roberts SK, Fink M, Testro AG, Angus PW, Gow PJ*

## CASE REPORT

- 1450 Focal nodular hyperplasia associated with a giant hepatocellular adenoma: A case report and review of literature

*Gaspar-Figueiredo S, Kefleyesus A, Sempoux C, Uldry E, Halkic N*

**ABOUT COVER**

Editorial Board Member of *World Journal of Hepatology*, Julio Coelho, MD, PhD, Professor, Department of Surgery, Federal University of Parana, Curitiba 80240-110, Parana, Brazil. coelhojcu@yahoo.com.br

**AIMS AND SCOPE**

The primary aim of *World Journal of Hepatology* (*WJH*, *World J Hepatol*) is to provide scholars and readers from various fields of hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

*WJH* mainly publishes articles reporting research results and findings obtained in the field of hepatology and covering a wide range of topics including chronic cholestatic liver diseases, cirrhosis and its complications, clinical alcoholic liver disease, drug induced liver disease autoimmune, fatty liver disease, genetic and pediatric liver diseases, hepatocellular carcinoma, hepatic stellate cells and fibrosis, liver immunology, liver regeneration, hepatic surgery, liver transplantation, biliary tract pathophysiology, non-invasive markers of liver fibrosis, viral hepatitis.

**INDEXING/ABSTRACTING**

The *WJH* is now abstracted and indexed in PubMed, PubMed Central, Emerging Sources Citation Index (Web of Science), Scopus, China National Knowledge Infrastructure (CNKI), China Science and Technology Journal Database (CSTJ), and Superstar Journals Database. The 2021 edition of Journal Citation Reports® cites the 2020 Journal Citation Indicator (JCI) for *WJH* as 0.61. The *WJH*'s CiteScore for 2020 is 5.6 and Scopus CiteScore rank 2020: Hepatology is 24/62.

**RESPONSIBLE EDITORS FOR THIS ISSUE**

Production Editor: Xu Guo; Production Department Director: Xiang Li; Editorial Office Director: Xiang Li.

**NAME OF JOURNAL**

*World Journal of Hepatology*

**ISSN**

ISSN 1948-5182 (online)

**LAUNCH DATE**

October 31, 2009

**FREQUENCY**

Monthly

**EDITORS-IN-CHIEF**

Nikolaos Pyrsopoulos, Ke-Qin Hu, Koo Jeong Kang

**EDITORIAL BOARD MEMBERS**

<https://www.wjnet.com/1948-5182/editorialboard.htm>

**PUBLICATION DATE**

October 27, 2021

**COPYRIGHT**

© 2021 Baishideng Publishing Group Inc

**INSTRUCTIONS TO AUTHORS**

<https://www.wjnet.com/bpg/gerinfo/204>

**GUIDELINES FOR ETHICS DOCUMENTS**

<https://www.wjnet.com/bpg/GerInfo/287>

**GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH**

<https://www.wjnet.com/bpg/gerinfo/240>

**PUBLICATION ETHICS**

<https://www.wjnet.com/bpg/GerInfo/288>

**PUBLICATION MISCONDUCT**

<https://www.wjnet.com/bpg/gerinfo/208>

**ARTICLE PROCESSING CHARGE**

<https://www.wjnet.com/bpg/gerinfo/242>

**STEPS FOR SUBMITTING MANUSCRIPTS**

<https://www.wjnet.com/bpg/GerInfo/239>

**ONLINE SUBMISSION**

<https://www.f6publishing.com>

## Retrospective Study

# Machine learning models for predicting non-alcoholic fatty liver disease in the general United States population: NHANES database

Amporn Atsawarungruangkit, Passisd Laoveeravat, Kittichai Promrat

**ORCID number:** Amporn

Atsawarungruangkit 0000-0003-0622-6839; Passisd Laoveeravat 0000-0001-6855-0437; Kittichai Promrat 0000-0002-4003-2598.

**Author contributions:**

Atsawarungruangkit A and Laoveeravat P contributed equally to this work including study design, data analysis, result interpretation, and manuscript writing; Promrat K critically revised the manuscript and provided supervision.

**Institutional review board**

**statement:** The National Health and Nutrition Examination Survey protocol was approved by the National Center for Health Statistics Research Ethics Review Board (Hyattsville, MD, United States).

**Informed consent statement:** In

NHANES III, the consent form was signed by participants in the survey.

**Conflict-of-interest statement:** No conflict of interest exists.

**Data sharing statement:** The dataset used in this manuscript is NHANES III, which is publicly available dataset.

**Open-Access:** This article is an

**Amporn Atsawarungruangkit, Kittichai Promrat,** Division of Gastroenterology, Warren Alpert Medical School, Brown University, Providence, RI 02903, United States

**Passisd Laoveeravat,** Division of Digestive Diseases and Nutrition, University of Kentucky College of Medicine, Lexington, KY 40536, United States

**Kittichai Promrat,** Division of Gastroenterology and Hepatology, Providence VA Medical Center, Providence, RI 02908, United States

**Corresponding author:** Amporn Atsawarungruangkit, MD, Academic Fellow, Instructor, Research Fellow, Division of Gastroenterology, Warren Alpert Medical School, Brown University, 593 Eddy Street, POB 240, Providence, RI 02903, United States.

[amporn\\_atsawarungruangkit@brown.edu](mailto:amporn_atsawarungruangkit@brown.edu)

## Abstract

### BACKGROUND

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease, affecting over 30% of the United States population. Early patient identification using a simple method is highly desirable.

### AIM

To create machine learning models for predicting NAFLD in the general United States population.

### METHODS

Using the NHANES 1988-1994. Thirty NAFLD-related factors were included. The dataset was divided into the training (70%) and testing (30%) datasets. Twenty-four machine learning algorithms were applied to the training dataset. The best-performing models and another interpretable model (*i.e.*, coarse trees) were tested using the testing dataset.

### RESULTS

There were 3235 participants ( $n = 3235$ ) that met the inclusion criteria. In the training phase, the ensemble of random undersampling (RUS) boosted trees had the highest F1 (0.53). In the testing phase, we compared selective machine learning models and NAFLD indices. Based on F1, the ensemble of RUS boosted trees remained the top performer (accuracy 71.1% and F1 0.56) followed by the fatty liver index (accuracy 68.8% and F1 0.52). A simple model (coarse trees) had

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

**Manuscript source:** Invited manuscript

**Specialty type:** Gastroenterology and hepatology

**Country/Territory of origin:** United States

**Peer-review report's scientific quality classification**

Grade A (Excellent): 0  
Grade B (Very good): B  
Grade C (Good): 0  
Grade D (Fair): 0  
Grade E (Poor): 0

**Received:** March 7, 2021

**Peer-review started:** March 7, 2021

**First decision:** May 2, 2021

**Revised:** May 11, 2021

**Accepted:** September 19, 2021

**Article in press:** September 19, 2021

**Published online:** October 27, 2021

**P-Reviewer:** Wu SZ

**S-Editor:** Gao CC

**L-Editor:** A

**P-Editor:** Liu JH



an accuracy of 74.9% and an F1 of 0.33.

## CONCLUSION

Not every machine learning model is complex. Using a simpler model such as coarse trees, we can create an interpretable model for predicting NAFLD with only two predictors: fasting C-peptide and waist circumference. Although the simpler model does not have the best performance, its simplicity is useful in clinical practice.

**Key Words:** Artificial intelligence; Machine learning; Non-alcoholic fatty liver disease; Fatty liver; United States population; NHANES

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core Tip:** A simple method with a good accuracy for identifying patients with non-alcoholic fatty liver disease is highly desirable. Among 24 machine learning models, the ensemble of random undersampling boosted trees was the top performer (accuracy 71.1% and F1 0.56). A simple model (coarse trees) with only two predictors (fasting C-peptide and waist circumference) had an accuracy of 74.9% and an F1 of 0.33. Not every machine learning model is complex. Using a simple model such as coarse trees, physicians can easily integrate machine learning model into their practice without any software implementation.

**Citation:** Atsawarungruangkit A, Laoveeravat P, Promrat K. Machine learning models for predicting non-alcoholic fatty liver disease in the general United States population: NHANES database. *World J Hepatol* 2021; 13(10): 1417-1427

**URL:** <https://www.wjgnet.com/1948-5182/full/v13/i10/1417.htm>

**DOI:** <https://dx.doi.org/10.4254/wjh.v13.i10.1417>

## INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is a common chronic metabolic disease found in 25.5% of the United States population, and it is more common in patients with diabetes (55.5%), leading to a health and economic burden[1-3]. Non-alcoholic steatohepatitis (NASH) can lead to liver-related consequences, such as cirrhosis, hepatocellular carcinoma, and mortality. NASH is the second most common indication for liver transplantation in the United States and is likely to replace hepatitis C infection as the leading cause of liver transplantation in the future[4]. NAFLD is diagnosed primarily with imaging studies, transient elastography, magnetic resonance elastography, or liver biopsy[5]. Some of these diagnostic modalities are not available in every health care facility, require expert interpretation, and are invasive in case of biopsy[5,6]. To prevent adverse outcomes in these patients, early screening and detection based on risk factors are warranted. Healthcare providers and patients are aware of the risk factors of NAFLD, which include diabetes, obesity, dyslipidemia, and metabolic syndrome[5,7,8]. However, there is no well-performing tool for the early prediction of NAFLD; for example, liver enzyme levels can be normal in patients with NAFLD[9,10]. There are existing studies on the risk factors and prediction risk scores; however, their results are controversial[11-15]. Machine learning is a potential approach for the identification of the best predictive model[16].

Machine learning can be used to construct a predictive model by teaching computer algorithms to learn from data without being explicitly programmed. Applications of machine learning in gastroenterology field are steadily increasing[17]. However, there is no machine learning model for predicting NAFLD in the United States. The published models in China, Germany, and Canada focus on NAFLD prediction scores using laboratory parameters and demographic data[11,13-15]. Therefore, we aimed to evaluate the applications of machine learning in NAFLD diagnosis for easy use at clinical setting.



## MATERIALS AND METHODS

### Study population and study design

The Third National Health and Nutrition Examination Survey (NHANES III) was a nationwide probability sample of 39695 persons aged 2 mo and older, conducted from 1988-1994 by the National Center for Health Statistics (NCHS). It aimed to evaluate the health and nutritional status of the general United States population[18]. Multiple datasets were collected in this survey, including demographics, interviews, physical examinations, and laboratory testing of biologic samples. The NHANES protocol was approved by the NCHS Research Ethics Review Board.

### Definitions

Participants aged 20 years or older in NHANES III with gradable ultrasound results were included in this study. The exclusion criteria included: (1) Excessive alcohol consumption; (2) Hepatitis B or C infection; (3) Fasting period outside of 8-24 h; and (4) Incomplete or missing data on physical examination and laboratory testing. The participants were divided into two groups: The NAFLD participants and non-NAFLD participants. Since participants aged above 74 years were not eligible for ultrasonography in NHANES III, participants aged above 74 years were excluded from this study.

'NAFLD participants' was defined based on: (1) Moderate to severe hepatic steatosis on ultrasound; (2) No history of alcohol drinking more than 2 drinks per day for men or 1 drink per day for women in the last 12 mo; and (3) No history of hepatitis B or C infection.

Thirty factors associated with NAFLD were included in this study: demographic (*i.e.*, age, gender, and race/ethnicity), body measurement [*i.e.*, body mass index (BMI) and waist circumference], general biochemistry tests [*i.e.*, iron, total iron-binding capacity, transferrin saturation, ferritin, cholesterol, triglyceride, high-density lipoprotein (HDL) cholesterol, C-reactive protein, and uric acid], liver chemistry (aspartate aminotransferase, alanine aminotransferase, gamma glutamyl transferase, alkaline phosphatase, total bilirubin, total protein, albumin, and serum globulin), diabetes testing profile [*i.e.*, glycated hemoglobin, fasting plasma glucose, fasting C-peptide, and fasting insulin], and the use of diabetes medication.

### Statistical analysis

Categorical and ordinal factors are presented as frequencies (%). Continuous factors are presented as medians (interquartile ranges). The dataset was divided into the training (70%) and testing (30%) datasets using stratified sampling. Differences between the two datasets were tested using the Mann-Whitney U test. Twenty-four machine learning algorithms were applied to the training dataset. Then, we selected the best performing models determined by accuracy and the F1 score and compared the out-of-sample performance with another interpretable model (coarse trees, a decision tree model with a maximum of four splits) and three NAFLD indices on the testing dataset. The selected NAFLD indices included fatty liver index (FLI), hepatic steatosis index (HSI), and triglyceride and glucose index (TyG)[19-21]. The cut-off levels for NAFLD were  $\geq 60$  for FLI,  $> 36$  for HSI, and  $\geq 8.5$  for TyG. The performance metrics include accuracy, sensitivity or recall, specificity, precision, area under the receiver operating characteristic curve (AUC), and the F1 score. It is worth noting that the F1 score is the harmonic mean of precision and recall. All statistical analyses were performed using MATLAB R2020a (MathWorks, MA, United States).

## RESULTS

The study had 3235 participants ( $n = 3235$ ). The participant selection process is shown in Figure 1. Based on ultrasound findings, 817 (25.26%) participants had NAFLD. The data of 2265 (70%) and 970 (30%) participants made up the training and testing groups, respectively. The baseline characteristics of participants in the training and testing groups are summarized in Table 1. There were no significant differences between the datasets for all factors.

The performances of 24 machine learning algorithms that were applied to the training dataset are illustrated in Table 2. The ensemble of subspace discriminant and ensemble of random undersampling (RUS) boosted trees had the highest accuracy (78.3%) and highest F1 score (0.53), respectively; both models had an AUC of 0.76. The coarse trees, decision trees with a few leaves, had an accuracy of 76%, AUC of 0.68,

**Table 1** Baseline characteristics of participants in training and testing data

	Training data (n = 2265)	Testing data (n = 970)	P value
Demographic			
Age (yr)	43 (29)	43.5 (28)	0.328
Gender (male) (%)	944 (41.68)	428 (44.12)	0.197
Race/ethnicity			
White (non-Hispanic) (%)	959 (42.34)	392 (40.41)	0.308
Black (non-Hispanic) (%)	627 (27.68)	271 (27.94)	0.882
Mexican American (%)	576 (25.43)	254 (26.19)	0.652
Others (%)	103 (4.55)	53 (5.46)	0.265
Body measurement			
Body mass index (kg/m <sup>2</sup> )	26.4 (7.2)	26.7 (7.4)	0.120
Waist circumference (cm)	93 (20.5)	93.5 (20.8)	0.182
Biochemistry tests			
Iron (ug/dL)	73 (39)	74 (39)	0.098
Total iron-binding capacity (ug/dL)	355 (72)	356 (72)	0.450
Transferrin saturation (%)	20.5 (11.1)	20.8 (11.8)	0.329
Ferritin (ng/mL)	87 (125)	84.5 (124)	0.508
Cholesterol (mg/dL)	201 (57)	204 (59)	0.155
Triglyceride (mg/dL)	120 (100.25)	122.5 (102)	0.562
HDL cholesterol (mg/dL)	48 (18)	48.5 (18)	0.585
C-reactive protein (mg/dL)	0.21 (0.29)	0.21 (0.23)	0.686
Uric acid (mg/dL)	5 (1.9)	5.1 (2)	0.427
Liver chemistry			
Aspartate aminotransferase (U/L)	19 (8)	19 (7)	0.908
Alanine aminotransferase (U/L)	14 (10)	14 (10)	0.581
Gamma glutamyl transferase (U/L)	21 (18)	21 (18)	0.787
Alkaline phosphatase (U/L)	83 (33)	81 (32)	0.524
Total bilirubin (mg/dL)	0.5 (0.2)	0.5 (0.2)	0.855
Total protein (g/dL)	7.4 (0.6)	7.4 (0.6)	0.559
Albumin (g/dL)	4.1 (0.5)	4.1 (0.4)	0.543
Serum globulin (g/dL)	3.3 (0.6)	3.3 (0.7)	0.941
Diabetes testing profile			
Glycated hemoglobin (%)	5.4 (0.8)	5.4 (0.7)	0.075
Fasting plasma glucose (mg/dL)	91.6 (12.52)	92.05 (12.2)	0.726
Fasting C-peptide (pmol/mL)	0.65 (0.68)	0.66 (0.69)	0.746
Fasting insulin (uU/mL)	9.36 (9.51)	9.73 (10.04)	0.378
Diabetes medication	165 (7.28%)	68 (7.01%)	0.782

and F1 score of 0.36.

As shown in the first half of [Table 3](#), the ensemble of subspace discriminant, coarse trees, and ensemble of RUS-boosted trees models were selected for testing the process on the testing data. When tested on the testing data, ensemble of subspace discriminant and ensemble of RUS-boosted trees still had a high accuracy (77.7%) and high F1 (0.56), respectively. The coarse tree had an accuracy of 74.9% and an F1 of 0.33. All the machine learning models and datasets are available for public access in the File



**Table 2** The performance comparison of machine learning models on training data

No.	Description	Accuracy (%)	AUC	PPV/precision (%)	NPV (%)	Sensitivity/recall (%)	Specificity (%)	F1
1	Fine tree	71.6	0.64	42.9	79.8	37.8	83.0	0.40
2	Medium tree	74.4	0.70	48.9	79.1	30.1	89.4	0.37
3	Coarse tree	76.0	0.68	55.1	78.9	26.4	92.7	0.36
4	Linear discriminant	78.0	0.75	61.1	80.9	35.5	92.4	0.45
5	Logistic regression	78.1	0.75	62.2	80.6	33.9	93.0	0.44
6	Gaussian naïve Bayes	75.1	0.74	50.8	81.1	40.2	86.8	0.45
7	Kernel naïve Bayes	72.7	0.73	46.8	85.1	60.1	76.9	0.53
8	Linear SVM	77.0	0.74	64.4	78.1	19.9	96.3	0.30
9	Quadratic SVM	77.4	0.70	59.9	80.1	31.8	92.8	0.42
10	Cubic SVM	72.8	0.64	45.1	79.6	35.3	85.5	0.40
11	Fine Gaussian SVM	74.7	0.67		74.7		100.0	
12	Medium Gaussian SVM	77.5	0.74	63.9	79.0	25.3	95.2	0.36
13	Coarse Gaussian SVM	75.7	0.74	66.2	76.0	7.9	98.6	0.14
14	Fine KNN	68.9	0.58	38.0	78.9	36.9	79.7	0.37
15	Medium KNN	76.5	0.71	59.7	78.1	21.0	95.2	0.31
16	Coarse KNN	76.6	0.75	78.1	76.5	10.0	99.1	0.18
17	Cosine KNN	76.6	0.72	57.9	79.2	27.6	93.2	0.37
18	Cubic KNN	77.0	0.72	62.0	78.5	22.6	95.3	0.33
19	Weighted KNN	76.5	0.71	56.7	79.4	28.8	92.6	0.38
20	Ensemble of boosted trees	76.9	0.74	57.3	80.3	33.6	91.6	0.42
21	Ensemble of bagged trees	77.2	0.74	58.9	80.2	32.5	92.3	0.42
22	Ensemble of subspace discriminant	78.3	0.76	66.7	79.7	28.3	95.2	0.40
23	Ensemble of subspace KNN	75.5	0.69	54.7	77.2	16.4	95.4	0.25
24	Ensemble of RUS boosted trees	70.4	0.76	44.2	86.3	66.4	71.7	0.53

AUC: Area under the curve; KNN: K-nearest neighbors; NPV: Negative predictive value; PPV: Positive predictive value; RUS: Random undersampling; SVM: Support vector machine.

Exchange portal of the MATLAB Central File Exchange[22]. The performance of three NAFLD on the testing data are also displayed in the second half of Table 3. FLI was the best performer among the NAFLD indices with the accuracy of 68.6% and F1 score of 0.52. However, the ensemble of RUS boosted trees was superior to FLI in all metrics.

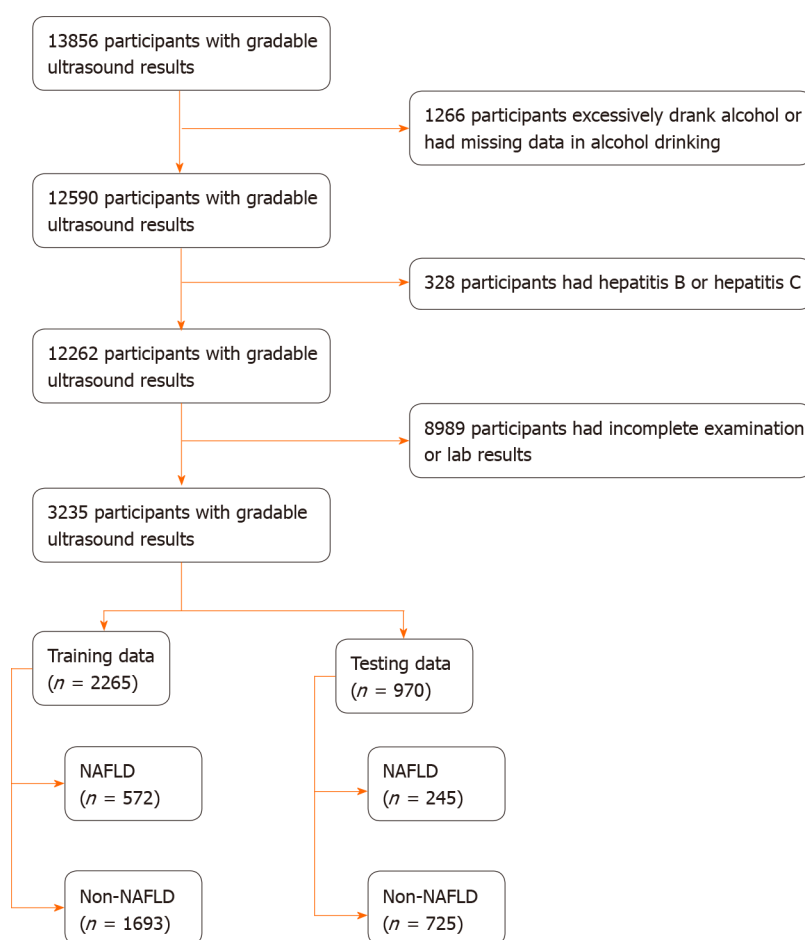
## DISCUSSION

Our study compared 24 different machine learning techniques to determine the optimal clinical predictive model for NAFLD. The accuracy of these models on the training data did not show much variation (range 9.4%), with an average of 75.5% (Table 2). The top two models were ensemble of subspace discriminant and ensemble of RUS boosted trees. The ensemble of subspace discriminant model had a higher accuracy while the ensemble of RUS boosted trees model had a better performance in classifying positive NAFLD, as indicated by the F1 score. Both models were ensemble type, which use multiple diverse models in combination to produce an optimal prediction. They are more complex machine learning models that apparently yield better predictions. Compared to accuracy, the F1 score is regarded as a superior performance metric for a class imbalance problem (often a large number of actual

**Table 3** The performance of machine learning models and other non-alcoholic fatty liver disease indices on testing data

No.	Description	Accuracy (%)	AUC	PPV/precision (%)	NPV (%)	Sensitivity/recall (%)	Specificity (%)	F1
<b>Machine learning models</b>								
1	Ensemble of subspace discriminant	77.7	0.78	66.7	78.8	23.7	96	0.35
2	Coarse trees	74.9	0.72	50.8	78.3	24.5	92	0.33
3	Ensemble of RUS boosted trees	71.1	0.79	45.5	88.4	72.7	70.6	0.56
<b>NAFLD indices</b>								
4	Fatty liver index	68.6	0.74	42.4	86.6	68.6	68.6	0.52
5	Hepatic steatosis index	65.1	0.70	37.9	83.3	60.4	66.6	0.47
6	Triglyceride and glucose index	56.9	0.69	34.8	88.3	80.8	48.8	0.49

AUC: Area under the curve; NAFLD: Non-alcoholic fatty liver disease; NPV: Negative predictive value; PPV: Positive predictive value; RUS: Random undersampling.


**Figure 1** Study design and data partitioning flow chart. NAFLD: Non-alcoholic fatty liver disease.

negatives). In our opinion, the ensemble of RUS boosted trees model was the best performing machine learning model in this study.

Technically, the final prediction of the ensemble method was derived from a combination of multiple predictions from different algorithms. In our case, the predicted outcome of the ensemble of RUS boosted trees model was derived from a weighted average outcome of 30 RUS boosted trees; the sample visualization of these RUS boosted trees can be found in the file uploaded to the MATLAB Central File Exchange[22].

On the other hand, we compared the performance of the previous model with the coarse trees model, simple decision trees with several leaves and splits (Figure 2). The decision logic of the coarse trees model consisted of only two factors: Waist circumference and serum C-peptide. In terms of testing performance, it had a reasonable accuracy (AUC, 0.72; accuracy, 74.9%; and F1 score, 0.33). Since it is simple-to-use and easily interpretable, the coarse trees model can be more practically used in clinical practice.

Waist circumference is directly associated with obesity and metabolic syndrome[23, 24]. They are also the established risk factors of NAFLD. The cut-off of 109.35 cm seems to be slightly higher than the general cut off value for metabolic syndrome (men, 102 cm and women, 80 cm)[25]. It is used to calculate the visceral adiposity index, which provides a good predictive capability[26]. The advantage of incorporating waist circumference into the model is its retrieval ability.

Our results are similar to those of a previous study identifying the risk factors of NAFLD[27]. C-peptide is an indicator of insulin resistance[28,29]. Serum C-peptide is associated with NAFLD, NASH, and fibrosis progression[28-30]. Additionally, serum C-peptide levels increase with NAFLD severity[29,31,32]. In our study, serum C-peptide is more significantly associated with NAFLD prediction than liver function test. This can be explained by the fact that liver enzymes are possibly not specific to NAFLD. They can also be elevated in other liver diseases. On the contrary, serum C-peptide is related to metabolic alterations, which play a direct role in NAFLD development.

We compared the performance of three NAFLD indices (FLI, HSI, and TyG) on the testing data. Among these three NAFLD indices, FLI had the highest performance in terms of accuracy (68.6%) and F1 (0.52). However, performance-wise, the ensemble of RUS boosted trees was superior to FLI in all aspects. In terms of simplicity, FLI is not complex, but it might be impossible for physicians to use it without spreadsheets or computers because it involves many mathematical operations, such as multiplication, logarithm function, and exponential function. Therefore, coarse trees remained the simplest model.

Previously developed machine learning models for NAFLD prediction have used more complex parameters, including laboratory and noninvasive scores. A population-based study in Italy developed a score for NAFLD diagnosis with a moderate accuracy of 68% in the model development phase, but extremely high performance in the testing (prediction) phase using the small sample size of 50. The predictors used in the model were of abdominal volume index, glucose, gamma glutamyl transferase, age, and sex[33]. A Chinese study incorporated three demographic factors and 15 Laboratory tests as predictors for Bayesian network model[8]. The inclusion of simple constituents, liver enzymes, lipid panels, and complete blood count resulting in an accuracy of up to 80% in a 10-fold cross validation; there was no separate data set for external validation or testing. A Taiwanese study revealed that waist circumference was the most influential factor in the model resulting in a high performance with an AUC of 0.925[13]. Similarly, such performance was based on a 10-fold cross validation, not on a separate data set for external validation or testing. In addition, the ethnic Chinese population generally has a lower alcohol consumption; it might not be generalized to other ethnic groups[12,15]. A Canadian study revealed that HDL, BMI, sex, plasma glucose, blood pressure, and age were factors used in the decision criteria of decision trees with an AUC of 0.73[14]. These reports showed different significant factors in their models. This might be explained by the different populations in terms of ethnicity, alcohol consumption, and obesity prevalence. Compared to prior reports, our study involved a general population of the United States, which has less selection bias and contains diverse races. Therefore, the derived models in this study can be applied to diverse ethnic and racial backgrounds. A detailed comparison of the proposed machine learning models in prior reports is summarized in Table 4.

The application of machine learning in regarding NAFLD has evolved from the diagnosis with the noninvasive screening methods to liver biopsy. The new score achieves the reasonable performance with AUC of 0.70, in terms of differentiating between NAFL and NASH[11]. Deep learning model was evaluated for diagnosis NAFLD based on ultrasound images and had a good predictive ability (AUC > 0.7) [34]. Given the advancement in this field, it can also be used to quantify steatosis, inflammation, ballooning, and fibrosis in biopsy histology of patients with NAFLD having excellent results[35].

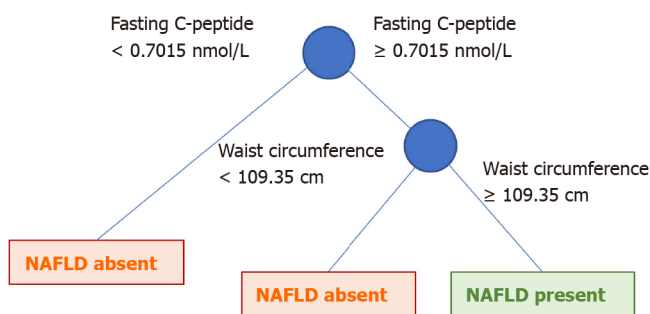
This study had strengths. First, this is the first United States population-based study with more than 3000 individuals from NHANES III. Secondly, we aimed to propose the simple model with a reasonable predictive power for NAFLD. This model will be potentially applied in clinical practice, especially by primary care providers, prior to

**Table 4 The performance comparison of published machine learning models on non-alcoholic fatty liver disease prediction**

Ref.	Type of data/country or territory of data	Number of train/external testing data	Model	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	F1
Sorino <i>et al</i> [33], 2020	Population/Italy	2920/50	Support vector machine	68 <sup>1</sup>	N/A	98.5	100	N/A
Wu <i>et al</i> [13], 2019	Hospital/Taiwan	577/NA	Random forest	86.5 <sup>1</sup>	0.925 <sup>1</sup>	87.2 <sup>1</sup>	85.9 <sup>1</sup>	N/A
Islam <i>et al</i> [36], 2018	Hospital/Taiwan	994/NA	Logistic regression	70 <sup>1</sup>	0.763 <sup>1</sup>	74.1 <sup>1</sup>	64.9 <sup>1</sup>	N/A
Ma <i>et al</i> [12], 2018	Hospital/China	10508/NA	Bayesian network	82.92 <sup>1</sup>	N/A	67.5 <sup>1</sup>	87.8 <sup>1</sup>	0.655 <sup>1</sup>
Perveen <i>et al</i> [14], 2018	Primary care network/Canada	64%/34% of 40637	Decision trees	N/A	0.73	73	N/A	0.67
Yip <i>et al</i> [15], 2017	Hospital/Hong Kong	500/442	Ridge regression	87	0.87	92	90	N/A
Birjandi <i>et al</i> [37], 2016	Hospital/Iran	359/1241	Decision trees	75	0.75	73	77	N/A
Our study	Population based/United States	2265/970	Ensemble of RUS boosted trees	71.1	0.79	72.7	70.6	0.56
			Coarse trees	74.9%	0.72	24.5%	92%	0.33

<sup>1</sup>Cross-validation performance (no separate dataset designated for testing the performance).

RUS: Random undersampling; AUC: area under the receiver operating characteristic curve; N/A: Not applicable; NA: Not available.


**Figure 2 The decision logic of coarse trees.** NAFLD: Non-alcoholic fatty liver disease.

referring patients to hepatologists. This study had some limitations. (1) Missing data were inherited from the nature of population dataset from NHANES III; (2) NAFLD was diagnosed with ultrasonography, which is not the gold standard; however, it is the primary imaging modality for NAFLD diagnosis in population-based studies and available in primary care medical facilities; (3) At the time of writing this article, there was no external dataset available that like that of NHANES III for validating the models; and (4) It may be impossible to completely reproduce the machine learning algorithms in this study since randomization was used in the modeling process, such as data partitioning, cross validation, and creation of some machine learning models. This explains why we made the trained models available to the public so that anyone can use the models directly and/or validate our results.

## CONCLUSION

Machine learning algorithms can summarize a large dataset into predictive models. The best performing model measured by the F1 score from our study is the ensemble of RUS boosted trees, which is a complex model that uses all 30 factors and behaves more like a black box to physicians. In contrast, the coarse trees model, which is composed of serum C-peptide and waist circumference, can generate a reasonable predictive performance, and most importantly is the simplest to use. To facilitate

clinical decision-making, complex models should be incorporated into the electronic medical record system. This will lead to proper investigation and treatment selection for specific individuals at risk, helping to maximize healthcare resource utilization. If software deployment is not achievable, a simple model be used directly by physicians. Therefore, the model choice depends on the user objectives and resources. Therefore, the more complex model required more resources and was likely to outperform. The less complex model may not be the most accurate model but can be easily implemented and interpreted in clinical practice.

## ARTICLE HIGHLIGHTS

### Research background

Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease that can progress to more severe liver disease.

### Research motivation

Early patient identification using a simple method is highly desirable for preventing the progression of NAFLD.

### Research objectives

To create machine learning models for predicting NAFLD in the general United States population.

### Research methods

This study was designed as a retrospective cohort by using the NHANES 1988-1994. Adults (20 years and above in age) with gradable ultrasound results were included in this study.

### Research results

Based on F1, the ensemble of ensemble of random undersampling boosted trees was the top performer (accuracy 71.1% and F1 0.56) while a simple model (coarse trees) had an accuracy of 74.9% and an F1 of 0.33.

### Research conclusions

Although a simpler model such as coarse trees was not the top performer, it consisted of only two predictors: fasting C-peptide and waist circumference. Its simplicity is useful in clinical practice.

### Research perspectives

The findings from this study can facilitate clinical decision-making for clinicians and also allow researchers to investigate the developed machine learning models. This will lead to proper investigation and treatment selection for specific individuals at risk, helping to maximize healthcare resource utilization.

## REFERENCES

- 1 **Fazel Y**, Koenig AB, Sayiner M, Goodman ZD, Younossi ZM. Epidemiology and natural history of non-alcoholic fatty liver disease. *Metabolism* 2016; **65**: 1017-1025 [PMID: 26997539 DOI: 10.1016/j.metabol.2016.01.012]
- 2 **Younossi ZM**, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, Racila A, Hunt S, Beckerman R. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatology* 2016; **64**: 1577-1586 [PMID: 27543837 DOI: 10.1002/hep.28785]
- 3 **Younossi ZM**, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016; **64**: 73-84 [PMID: 26707365 DOI: 10.1002/hep.28431]
- 4 **Younossi Z**, Stepanova M, Ong JP, Jacobson IM, Bugianesi E, Duseja A, Eguchi Y, Wong VW, Negro F, Yilmaz Y, Romero-Gomez M, George J, Ahmed A, Wong R, Younossi I, Ziayee M, Afendy A; Global Nonalcoholic Steatohepatitis Council. Nonalcoholic Steatohepatitis Is the Fastest Growing Cause of Hepatocellular Carcinoma in Liver Transplant Candidates. *Clin Gastroenterol Hepatol* 2019; **17**: 748-755.e3 [PMID: 29908364 DOI: 10.1016/j.cgh.2018.05.057]
- 5 **Younossi ZM**, Loomba R, Anstee QM, Rinella ME, Bugianesi E, Marchesini G, Neuschwander-Tetri BA, Serfaty L, Negro F, Caldwell SH, Ratziu V, Corey KE, Friedman SL, Abdelmalek MF, Harrison

- SA, Sanyal AJ, Lavine JE, Mathurin P, Charlton MR, Goodman ZD, Chalasani NP, Kowdley KV, George J, Lindor K. Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis. *Hepatology* 2018; **68**: 349-360 [PMID: [29222917](#) DOI: [10.1002/hep.29721](#)]
- 6 Arab JP, Dirchwolf M, Álvares-da-Silva MR, Barrera F, Benítez C, Castellanos-Fernandez M, Castro-Narro G, Chavez-Tapia N, Chiodi D, Cotrim H, Cusi K, de Oliveira CPMS, Díaz J, Fassio E, Gerona S, Giralda M, Hernandez N, Marciano S, Masson W, Méndez-Sánchez N, Leite N, Lozano A, Padilla M, Panduro A, Paraná R, Parise E, Perez M, Poniachik J, Restrepo JC, Ruf A, Silva M, Tagle M, Tapias M, Torres K, Vilar-Gomez E, Costa Gil JE, Gadano A, Arrese M. Latin American Association for the study of the liver (ALEH) practice guidance for the diagnosis and treatment of non-alcoholic fatty liver disease. *Ann Hepatol* 2020; **19**: 674-690 [PMID: [33031970](#) DOI: [10.1016/j.aohp.2020.09.006](#)]
- 7 Portillo-Sanchez P, Bril F, Maximos M, Lomonaco R, Biernacki D, Orsak B, Subbarayan S, Webb A, Hecht J, Cusi K. High Prevalence of Nonalcoholic Fatty Liver Disease in Patients With Type 2 Diabetes Mellitus and Normal Plasma Aminotransferase Levels. *J Clin Endocrinol Metab* 2015; **100**: 2231-2238 [PMID: [25885947](#) DOI: [10.1210/jc.2015-1966](#)]
- 8 Ma J, Hwang SJ, Pedley A, Massaro JM, Hoffmann U, Chung RT, Benjamin EJ, Levy D, Fox CS, Long MT. Bi-directional analysis between fatty liver and cardiovascular disease risk factors. *J Hepatol* 2017; **66**: 390-397 [PMID: [27729222](#) DOI: [10.1016/j.jhep.2016.09.022](#)]
- 9 Siddiqui MS, Sterling RK, Luketic VA, Puri P, Stravitz RT, Bouneva I, Boyett S, Fuchs M, Sargeant C, Warnick GR, Grami S, Sanyal AJ. Association between high-normal levels of alanine aminotransferase and risk factors for atherogenesis. *Gastroenterology* 2013; **145**: 1271-9.e1 [PMID: [23973920](#) DOI: [10.1053/j.gastro.2013.08.036](#)]
- 10 Blais P, Husain N, Kramer JR, Kowalkowski M, El-Serag H, Kanwal F. Nonalcoholic fatty liver disease is underrecognized in the primary care setting. *Am J Gastroenterol* 2015; **110**: 10-14 [PMID: [24890441](#) DOI: [10.1038/ajg.2014.134](#)]
- 11 Canbay A, Kälisch J, Neumann U, Rau M, Hohenester S, Baba HA, Rust C, Geier A, Heider D, Sowa JP. Non-invasive assessment of NAFLD as systemic disease-A machine learning perspective. *PLoS One* 2019; **14**: e0214436 [PMID: [30913263](#) DOI: [10.1371/journal.pone.0214436](#)]
- 12 Ma H, Xu CF, Shen Z, Yu CH, Li YM. Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China. *Biomed Res Int* 2018; **2018**: 4304376 [PMID: [30402478](#) DOI: [10.1155/2018/4304376](#)]
- 13 Wu CC, Yeh WC, Hsu WD, Islam MM, Nguyen PAA, Poly TN, Wang YC, Yang HC, Jack Li YC. Prediction of fatty liver disease using machine learning algorithms. *Comput Methods Programs Biomed* 2019; **170**: 23-29 [PMID: [30712601](#) DOI: [10.1016/j.cmpb.2018.12.032](#)]
- 14 Perveen S, Shahbaz M, Keshavjee K, Guergachi A. A Systematic Machine Learning Based Approach for the Diagnosis of Non-Alcoholic Fatty Liver Disease Risk and Progression. *Sci Rep* 2018; **8**: 2112 [PMID: [29391513](#) DOI: [10.1038/s41598-018-20166-x](#)]
- 15 Yip TC, Ma AJ, Wong VW, Tse YK, Chan HL, Yuen PC, Wong GL. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment Pharmacol Ther* 2017; **46**: 447-456 [PMID: [28585725](#) DOI: [10.1111/apt.14172](#)]
- 16 Ahn JC, Connell A, Simonetto DA, Hughes C, Shah VH. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. *Hepatology* 2021; **73**: 2546-2563 [PMID: [33098140](#) DOI: [10.1002/hep.31603](#)]
- 17 Spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, Bhat M. Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review. *Hepatology* 2020; **71**: 1093-1105 [PMID: [31907954](#) DOI: [10.1002/hep.31103](#)]
- 18 National Center for Health Statistics. Third National Health and Nutrition Examination Survey Data (NHANES III). [cited 15 Dec 2020]. In: National Center for Health Statistics [Internet]. Available from: <https://www.cdc.gov/nchs/nhanes/nhanes3/>
- 19 Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, Tiribelli C. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol* 2006; **6**: 33 [PMID: [17081293](#) DOI: [10.1186/1471-230X-6-33](#)]
- 20 Lee JH, Kim D, Kim HJ, Lee CH, Yang JI, Kim W, Kim YJ, Yoon JH, Cho SH, Sung MW, Lee HS. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis* 2010; **42**: 503-508 [PMID: [19766548](#) DOI: [10.1016/j.dld.2009.08.002](#)]
- 21 Zhang S, Du T, Zhang J, Lu H, Lin X, Xie J, Yang Y, Yu X. The triglyceride and glucose index (TyG) is an effective biomarker to identify nonalcoholic fatty liver disease. *Lipids Health Dis* 2017; **16**: 15 [PMID: [28103934](#) DOI: [10.1186/s12944-017-0409-6](#)]
- 22 Atsawarungruangkit A. Machine learning models for predicting NAFLD. MATLAB Central File Exchange. [cited 15 Dec 2020]. In: MathWorks [Internet]. Available from: <https://www.mathworks.com/matlabcentral/fileexchange/83953-machine-learning-models-for-predicting-naflD>
- 23 Bitew ZW, Alemu A, Ayele EG, Tenaw Z, Alebel A, Worku T. Metabolic syndrome among children and adolescents in low and middle income countries: a systematic review and meta-analysis. *Diabetol Metab Syndr* 2020; **12**: 93 [PMID: [33117455](#) DOI: [10.1186/s13098-020-00601-8](#)]
- 24 Staynor JMD, Smith MK, Donnelly CJ, Sallam AE, Ackland TR. DXA reference values and anthropometric screening for visceral obesity in Western Australian adults. *Sci Rep* 2020; **10**: 18731 [PMID: [33127940](#) DOI: [10.1038/s41598-020-73631-x](#)]



- 25 **Eyvazlou M**, Hosseinpouri M, Mokarami H, Gharibi V, Jahangiri M, Cousins R, Nikbakht HA, Barkhordari A. Prediction of metabolic syndrome based on sleep and work-related risk factors using an artificial neural network. *BMC Endocr Disord* 2020; **20**: 169 [PMID: [33183282](#) DOI: [10.1186/s12902-020-00645-x](#)]
- 26 **Vural Keskinler M**, Mutlu HH, Sirin A, Erkalma Senates B, Colak Y, Tuncer I, Oguz A. Visceral Adiposity Index As a Practical Tool in Patients with Biopsy-Proven Nonalcoholic Fatty Liver Disease/Nonalcoholic Steatohepatitis. *Metab Syndr Relat Disord* 2021; **19**: 26-31 [PMID: [32898457](#) DOI: [10.1089/met.2020.0054](#)]
- 27 **Atsawarungruangkit A**, Chenbhanich J, Dickstein G. C-peptide as a key risk factor for non-alcoholic fatty liver disease in the United States population. *World J Gastroenterol* 2018; **24**: 3663-3670 [PMID: [30166861](#) DOI: [10.3748/wjg.v24.i32.3663](#)]
- 28 **Yesilova Z**, Ozata M, Oktenli C, Bagci S, Ozcan A, Sanisoglu SY, Uygur A, Yaman H, Karaeren N, Dagalp K. Increased acylation stimulating protein concentrations in nonalcoholic fatty liver disease are associated with insulin resistance. *Am J Gastroenterol* 2005; **100**: 842-849 [PMID: [15784030](#) DOI: [10.1111/j.1572-0241.2005.40838.x](#)]
- 29 **Chitturi S**, Abeygunasekera S, Farrell GC, Holmes-Walker J, Hui JM, Fung C, Karim R, Lin R, Samarasinghe D, Liddle C, Weltman M, George J. NASH and insulin resistance: Insulin hypersecretion and specific association with the insulin resistance syndrome. *Hepatology* 2002; **35**: 373-379 [PMID: [11826411](#) DOI: [10.1053/jhep.2002.30692](#)]
- 30 **Chalasani N**, Deeg MA, Persohn S, Crabb DW. Metabolic and anthropometric evaluation of insulin resistance in nondiabetic patients with nonalcoholic steatohepatitis. *Am J Gastroenterol* 2003; **98**: 1849-1855 [PMID: [12907343](#) DOI: [10.1111/j.1572-0241.2003.07619.x](#)]
- 31 **Francque SM**, Verrijken A, Mertens I, Hubens G, Van Marck E, Pelckmans P, Michielsens P, Van Gaal L. Noninvasive assessment of nonalcoholic fatty liver disease in obese or overweight patients. *Clin Gastroenterol Hepatol* 2012; **10**: 1162-1168; quiz e87 [PMID: [22796457](#) DOI: [10.1016/j.cgh.2012.06.019](#)]
- 32 **Hui JM**, Hodge A, Farrell GC, Kench JG, Kriketos A, George J. Beyond insulin resistance in NASH: TNF-alpha or adiponectin? *Hepatology* 2004; **40**: 46-54 [PMID: [15239085](#) DOI: [10.1002/hep.20280](#)]
- 33 **Sorino P**, Caruso MG, Misciagna G, Bonfiglio C, Campanella A, Mirizzi A, Franco I, Bianco A, Buongiorno C, Liuzzi R, Cisternino AM, Notarnicola M, Chiloio M, Pascoschi G, Osella AR; MICOL Group. Selecting the best machine learning algorithm to support the diagnosis of Non-Alcoholic Fatty Liver Disease: A meta learner study. *PLoS One* 2020; **15**: e0240867 [PMID: [33079971](#) DOI: [10.1371/journal.pone.0240867](#)]
- 34 **Cao W**, An X, Cong L, Lyu C, Zhou Q, Guo R. Application of Deep Learning in Quantitative Analysis of 2-Dimensional Ultrasound Imaging of Nonalcoholic Fatty Liver Disease. *J Ultrasound Med* 2020; **39**: 51-59 [PMID: [31222786](#) DOI: [10.1002/jum.15070](#)]
- 35 **Forlano R**, Mullish BH, Giannakeas N, Maurice JB, Angkathunyakul N, Lloyd J, Tzallas AT, Tsipouras M, Yee M, Thursz MR, Goldin RD, Manousou P. High-Throughput, Machine Learning-Based Quantification of Steatosis, Inflammation, Ballooning, and Fibrosis in Biopsies From Patients With Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol* 2020; **18**: 2081-2090.e9 [PMID: [31887451](#) DOI: [10.1016/j.cgh.2019.12.025](#)]
- 36 **Islam MM**, Wu CC, Poly TN, Yang HC, Li YJ. Applications of Machine Learning in Fatty Live Disease Prediction. *Stud Health Technol Inform* 2018; **247**: 166-170 [PMID: [29677944](#)]
- 37 **Birjandi M**, Ayatollahi SM, Pourahmad S, Safarpour AR. Prediction and Diagnosis of Non-Alcoholic Fatty Liver Disease (NAFLD) and Identification of Its Associated Factors Using the Classification Tree Method. *Iran Red Crescent Med J* 2016; **18**: e32858 [PMID: [28191344](#) DOI: [10.5812/ircmj.32858](#)]





Published by **Baishideng Publishing Group Inc**  
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

**Telephone:** +1-925-3991568

**E-mail:** [bpgoffice@wjgnet.com](mailto:bpgoffice@wjgnet.com)

**Help Desk:** <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

