LEARN algorithm: a novel option for predicting non-alcoholic steatohepatitis

Short Title: LEARN algorithm for NASH prediction

Authors' names

Gang Li, MD[1#], Tian-Lei Zheng [2#], Xiao-Ling Chi, MD[3], Yong-Fen Zhu, MD[4], Jin-Jun Chen, MD[5], Liang Xu, MD[6], Jun-Ping Shi, MD, PhD [7], Xiao-Dong Wang, MD[8], Wei-Guo Zhao[2], Christopher D. Byrne, MD, PhD[9], Giovanni Targher, MD[10], Rafael S. Rios, MD[1], Ou-Yang Huang, MD[1], Liang-Jie Tang, MD[1], Shi-Jin Zhang[2], Shi Geng[2], Huan-Ming Xiao, MD[3], Sui-Dan Chen, MD[11], Rui Zhang, MD[12], Ming-Hua Zheng, MD, PhD[1,8,13]*

Affiliations

[1]NAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China;

[2]Artificial Intelligence Unit, Department of Medical Equipment, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, China;

[3]Department of Hepatology, Guangdong Provincial Hospital of Chinese Medicine, the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China;

[4]Department of Hepatology and Infection, Sir Run Run Shaw Hospital, Affiliated with School of Medicine, Zhejiang University, Hangzhou, China;

[5]Hepatology Unit, Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou, China; Hepatology Unit, Zengcheng Branch,

Nanfang Hospital, Southern Medical University, Guangzhou, China;

[6]Department of Hepatology, Tianjin Second People's Hospital, Tianjin, China,

[7]Department of Liver Diseases, Hangzhou Normal University Affiliated Hospital, Hangzhou, Zhejiang, China;

[8]Key Laboratory of Diagnosis and Treatment for the Development of Chronic Liver Disease in Zhejiang Province, Wenzhou, China;

[9]Southampton National Institute for Health Research Biomedical Research Centre, University Hospital Southampton, Southampton General Hospital, Southampton, UK;

[10]Section of Endocrinology, Diabetes and Metabolism, Department of Medicine, University and Azienda Ospedaliera Universitaria Integrata of Verona, Verona, Italy;

[11]Department of Pathology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China;

[12]Department of Nutrition, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China;

[13]Institute of Hepatology, Wenzhou Medical University, Wenzhou, China;

#Co-first author:

Gang Li and Tian-Lei Zheng

***Corresponding Author:**

Ming-Hua Zheng, MD, PhD

NAFLD Research Center, Department of Hepatology, the First Affiliated Hospital of Wenzhou Medical University; No. 2 Fuxue Lane, Wenzhou 325000, China.

E-mail: zhengmh@wmu.edu.cn; fax: (86) 577-55578522; tel: (86) 577-55579611.

Electronic word count: 4355

Number of tables/supplemental tables: 4/3

Number of figures/supplemental figures: 4/1

Abbreviation list:

ALT = alanine aminotransferase, AST = aspartate aminotransferase, AUROC = area under the receiver operator characteristic curve, BMI = body mass index, HOMA-IR = homeostasis model assessment of insulin resistance, NAFL = non-alcoholic fatty liver, NAFLD = non-alcoholic fatty liver disease, NAS = NAFLD activity score, NASH = non-alcoholic steatohepatitis, BIA = bioelectrical impedance analysis, LEARN = bioe**L**ectrical imp**E**dance **A**nalysis fo**R** **N**ash, WMBC= without missing data of body composition, PMBC = partial missing data of body composition, NPV = negative predictive value, OR = odds ratio, PPV = positive predictive value

Conflict of interest disclosure:

All authors have nothing to declare.

Author's contributions

Study concept and design: Gang Li and Ming-Hua Zheng

Acquisition of data: Gang Li, Xiao-Ling Chi, Yong-Fen Zhu, Jin-Jun Chen, Liang Xu, Jun-Ping Shi, Xiao-Dong Wang, Wei-Guo Zhao, Ou-Yang Huang, Liang-Jie Tang, Huan-Ming Xiao, Rui Zhang

Pathology analysis: Xiao-Dong Wang and Sui-Dan Chen

Drafting of the manuscript: Gang Li

Critical revision: Ming-Hua Zheng, Rafael S. Rios, Giovanni Targher and Christopher

D. Byrne

Statistical analysis: Gang Li

Artificial intelligence technology: Tian-lei Zheng, Wei-Guo Zhao, Shi-Jin Zhang, Shi Geng

Study supervision: Ming-Hua Zheng

All authors contributed to the manuscript for important intellectual content and approved the submission.

Ethical Statement:

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the local ethics board of our hospital (2016-246) and informed consent was taken from all individual participants.

Funding statement:

This work was supported by grants from the National Natural Science Foundation of China (82070588), High Level Creative Talents from Department of Public Health in Zhejiang Province (S2032102600032), Project of New Century 551 Talent Nurturing in Wenzhou. GT is supported in part by grants from the University School of Medicine of Verona, Verona, Italy. CDB is supported in part by the Southampton NIHR Biomedical Research Centre (IS-BRC-20004), UK.

Conflicts of interest

All authors: nothing to declare.

Abstract

Background: There is an unmet need for accurate non-invasive methods to diagnose non-alcoholic steatohepatitis (NASH). Since impedance-based measurements of body composition are simple, repeatable and have a strong association with non-alcoholic fatty liver disease (NAFLD) severity, we aimed to develop a novel and fully automatic machine learning algorithm, consisting of a deep neural network based on impedance-based measurements of body composition to identify NASH (the LEARN algorithm).

Methods: A total of 1,259 consecutive subjects with suspected NAFLD were screened from six medical centers across China, of which 766 patients with biopsy-proven NAFLD were included in final analysis. These patients were randomly subdivided into the training and validation groups, in a ratio of 4:1. The LEARN algorithm was developed in the training group to identify NASH, and subsequently, tested in the validation group.

Results: The LEARN algorithm utilizing impedance-based measurements of body composition along with age, sex, pre-existing hypertension and diabetes, was able to predict the likelihood of having NASH. This algorithm showed good discriminatory ability for identifying NASH in both the training and validation groups (AUROC: 0.81, 95%CI 0.77-0.84 and 0.80, 0.73-0.87, respectively). This algorithm also performed better than serum cytokeratin-18 neoepitope M30 level or other non-invasive NASH scores (including HAIR, ION, NICE) for identifying NASH (p-value <0.001). Additionally, the LEARN algorithm performed well in identifying NASH in

different patient subgroups, as well as in subjects with partial missing body composition data.

Conclusion: The LEARN algorithm, utilizing simple easily obtained measures, provides a fully automated, simple, non-invasive method for identifying NASH.

Keywords: NAFLD, NASH, LEARN algorithm, body composition.

Introduction

Non-alcoholic steatohepatitis (NASH) is a major public health concern worldwide and, compared with hepatic steatosis alone, the annual incidence of hepatocellular carcinoma in patients with NASH-related cirrhosis is as high as 1-2%.(1,2) NASH is more likely to lead to advanced liver fibrosis, cirrhosis and eventually liver-related illness and death.(3-5) Therefore, due to its high prevalence and increased health risks, NASH is a significant economic and healthcare burden. The current definitive diagnosis of NASH is based not only on hepatocyte fat accumulation (steatosis), but also on histological evidence of hepatocyte ballooning and lobular inflammation.(6) Given that the majority of patients with NASH are asymptomatic, the acceptability of liver biopsy (i.e. the gold standard) is relatively low and, because of liver biopsy-associated morbidity and even mortality, developing screening strategies to identify those individuals at risk of progressive NASH, remains an unmet need. Furthermore, non-invasive tests that may accurately predict disease progression (as part of the natural history of NASH), or identify regression (in response to treatment), are urgently needed to decrease the reliance on repeat liver biopsies.(7-9)

Machine learning techniques require uploading a large amount of data to a computer program, and then selecting a model to "fit" these data for computer prediction, which creates new possibilities in medicine for diagnosing diseases.(10-12) In previous studies, machine learning has facilitated success in cancer diagnosis and diagnosis of liver fibrosis.(13,14) Recently, in Sagimet's NASH FASCINATE-2 Phase 2b Clinical

Trial, stain-free artificial intelligence (AI)-based digital pathology was incorporated as secondary and exploratory efficacy endpoints. These advances would have been unimaginable without machine learning. To date, however, there is no a validated, non-invasive, simple, machine learning-based algorithm (MLA) for diagnosing NASH.

Bioelectrical impedance analysis (BIA) is a simple, commonly used, non-invasive and inexpensive method for assessing body composition.(15) This method can provide >20 parameters on different dimensions of body composition, such as body fat content, muscle mass, bone mineral content and metabolic rate. Interestingly, there is evidence that body composition in NAFLD is different from that of non-steatotic control subjects.(16-19) However, the abundant body composition outputs from BIA have not yet been fully evaluated and exploited in the diagnosis and treatment of NAFLD.

Therefore, the main aim of our multicenter cross-sectional study was to establish and validate a novel MLA, referred to as a deep neural network algorithm for non-invasively identifying NASH by impedance-based measures of body composition (named as the LEARN [bioeLectrical impEdance Analysis foR Nash] algorithm).

Materials and Methods

Study subjects and design

A total of 1,259 consecutive subjects with suspected NAFLD were initially screened from six medical centers across China from September 2016 to April 2021. Inclusion criteria were as follows: (1) elevated serum aminotransferase concentrations and/or evidence of hepatic steatosis on imaging methods (irrespective of serum aminotransferase levels); (2) agreement to undergo a liver biopsy; (3) agreement to undergo BIA method within 1 month of liver biopsy; and (4) age range from 18 to 75 years. A total of 493 subjects were excluded due to the following criteria: (1) excessive alcohol consumption (more than 20 and 10 grams per day for men and women, respectively); (2) other coexisting chronic liver diseases, such as viral hepatitis, autoimmune hepatitis, or drug-induced liver injury; (3) absence of hepatic steatosis on histology (steatotic hepatocytes ≤5%); and (4) no BIA measurement. As a consequence of these exclusion criteria**,** 766 Chinese adults with biopsy-confirmed NAFLD were included in the final analysis **(Figure 1)**. The study was approved by the local ethics committee of each medical center. Written informed consent was obtained from each participant.

Clinical and laboratory data

For every patient, demographic data, anthropometry, clinical biochemical parameters and concomitant diseases were measured and collected during liver biopsy, at each center, within 48 hours from liver biopsy. Hypertension was defined as blood pressure

≥130/85 mmHg or current use any of anti-hypertensive drugs. Presence of diabetes was defined as self-reported physician diagnosis of diabetes, use of anti-hyperglycemic drugs, fasting glucose levels ≥7 mmol/L or hemoglobin A1c (HbA1c) ≥6.5% (≥48 mmol/mol). Homeostasis model assessment (HOMA-IR) was used to estimate insulin resistance, and body mass index (BMI) ≥25 kg/m^2 was diagnosed as overweight/obese. The specific methods for assessing HOMA-IR and BMI have been described in our previous study.(20)

Measurement of cytokeratin-18 neoepitope M30 (CK-18 M30)

Serum CK-18 M30 level was measured only in the Wenzhou cohort. Serum CK-18 M30 level was determined by a commercially ELISA kit (Herui Biomed Company Limited, Suzhou, China), according to the manufacturer's recommendation. The specific detection details have been described in our previous study.(21)

Body Composition Measurement

Each patient was examined for body composition (within 48 hours of the liver biopsy) by professionally trained personnel at each center in accordance with uniform operating instructions. Specifically, BIA (InBody 720; Biospace, land Seoul, Korea) was employed to measure body composition. According to operating instructions, the subjects took off their shoes and removed their belongings and coats and stood on the designated electrodes. The thumb of both hands was placed on the thumb electrode button, and the other four fingers were all placed on the electrode, under the handle,

with arms straightened. The impedance of left arm, right arm, trunk, right leg and left leg were measured at six frequencies (1, 5, 50, 250, 500, and 1000 kHz). Based on the aforementioned impedances, the system automatically produced information of body composition, which included 20 parameters, such as intracellular water, extracellular water, total body water, soft lean mass, fat free mass, weight, skeletal muscle mass, body fat mass, percent body fat, waist-hip ratio, right arm, left arm, trunk, as well as right leg, left leg, visceral fat area, body cell mass, bone mineral content, basal metabolic rate, arm circumference and arm muscle circumference.

Liver biopsy

Liver biopsies were performed using a 16-gauge needle under ultrasound guidance as previous described.(21) All liver biopsy specimens were interpreted by an experienced pathologist from each center. Diagnostic criteria for NAFLD were the evidence of steatotic hepatocytes >5% on histology. NASH was diagnosed only when the NAFLD activity score (NAS) was ≥4 and each component of its three histological features (i.e. steatosis, ballooning and lobular inflammation) was ≥1.(22) Liver fibrosis stage was graded from 0 to four, according to the Brunt's histologic criteria.

Development of the LEARN algorithm

An experienced AI team used neural network algorithms to build a prediction model that provided clinicians with an individual's probability of having NASH. As shown in the **Figure 2**, the data from our 766 patients with biopsy-proven NAFLD were first

subdivided randomly into a training set (613 patients) and a validation set (153 patients), in a 4:1 ratio. The data included in the LEARN algorithm included age, sex, diabetes and hypertension status, as well as 20 body composition parameters obtained by BIA. The MLA process is currently the subject of a patent application. In particular, we normalized processing, and inputted these data to the input layer composed of the full connection network. In this layer, we further analyzed the 20 body composition parameters. Each parameter was automatically assigned to a different weight in the neural network model, and the best choices of the first six body composition information parameters (i.e., arm circumference, body fat percentage, bone mineral content, basal metabolic rate, body cell mass and visceral fat area) for prediction of NASH were selected after the method of exhaustion (as shown in **Figure 2**). The six body composition parameters along with age, sex, and prior history of diabetes or hypertension were re-sent into the input layer, to extract the feature matrix. The feature matrix was then inputted into the residual network layer, composed of four residual modules, which can also be called the hidden layer. Each fully connected residual module consists of three fully connected modules and residual structure. The first two fully connected modules include the fully connected layer, the batch normalization layer and the Tanh activation function, and the last module removes the activation function compared with the previous two modules. A single fully connected residual module may be expressed as:

$$x_m = F(x, \{W_1, W_2, W_3\})$$

$$y = Dropout(Tanh(x_m + x))$$

where x is the input feature, y is the output feature, W indicates the weight of the fully connected module and F indicates the combination of three fully connected modules. For the feature x of the input fully connected residual module, the module firstly uses three fully connected modules to extract feature successively to generate intermediate feature x_m. Then add x as residuals to x_m, and use the Dropout function to generate y after Tanh activation. Four fully connected residual modules can increase the depth of the model while suppressing the disappearance of gradient, thus improving the performance of the model. Finally, the extracted features are inputted into the output layer, composed of the fully connected network and softmax activation function, to calculate the probability of having NASH.

Other widely used non-invasive NASH scores

As liver biopsy can be fraught with major acute complications, there are some widely used non-invasive scores for diagnosing NASH, such as ION, HAIR, NICE and model, which are based on combinations of laboratory indicators and metabolic factors.(23-25) In particular, these three non-invasive NASH scores can be calculated as follows:

The index of NASH (ION) =1.33 waist-to hip ratio +0.03 × triglycerides (mg/dl) + 0.18 × ALT (U/L) +8.53 ×HOMA-IR – 13.93 in men; 0.02 × triglycerides (mg/dl) + 0.24 × ALT (U/L) + 9.61 × HOMA-IR – 13.99 in women.

The NICE model = -5.654 + 3.780E-02 × ALT (IU／L) + 2.215E-03× CK18 fragment (IU／L) + 1.825 × (presence of metabolic syndrome = 1)

The HAIR score was calculated by adding hypertension =1, ALT > 40 U/L = 1, and HOMA-IR index > 5.0 = 1 for each patient (0–3).

Statistical analysis

Continuous and categorical data were expressed as means ± standard deviations, and medians (1st quartile, 3rd quartile), or proportions, respectively. For the purpose of determining statistical differences between the training and the validation groups, the unpaired Student's t-test (for normally distributed continuous data), the Mann-Whitney U-test (for non-normally distributed continuous data) and the chi-square test (for categorical variables) were used. PASS15 was used to estimate the sample size. The area under ROC curve was 0.80, α=0.05(bilateral), β=0.1 (test efficiency was 0.9), and the ratio between groups was 3:2. It was found that a minimum of 42 subjects, including 25 patients and 17 controls, needed to be enrolled. A sample size of 776 is completely sufficient. The process of establishing the deep neural network algorithm (the LEARN algorithm) was summarized in **Figure 2**. The area under the receiver operating characteristic (AUROC) curve was calculated to evaluate the discrimination of the machine learning intelligent diagnostic model. Cut-off values for the diagnosis of NASH were identified in the training group, corresponding to 90% sensitivity and 90% specificity, respectively. At the same time, the specificity, sensitivity, negative predictive value (NPV), positive predictive value (PPV) and the gray zone were also calculated corresponding to each cut-off value in the training and validation groups. Statistical significance was two sided, set at p-value less than 0.05.

All statistical tests were performed by R (http://www.r-project.org) and SPSS version 22.0 (SPSS Inc).

Results

Patient characteristics

A total of 766 Chinese adult patients with biopsy-confirmed NAFLD from six hospitals were randomly assigned to the training (n = 613) and validation groups (n = 153) in a ratio of 4:1. As shown in **Table 1**, there were no statistical differences in demographic and biochemical measurements, body composition data, as well as other widely used non-invasive NASH scores (ION, HAIR, NICE model) and individual features of liver histology between the two groups. It is worth noting that a complete body composition examination report included 20 parameters, involving also skeletal muscle mass and abdominal fat area. However, due to the loss of data during the collection process, some patient's body composition examination reports were incomplete and reported between 12 and 19 measurements, which we referred to as subjects with partial missing data on body composition (PMBC). The baseline clinical, biochemical and BIA characteristics of patients stratified by those without missing BIA data (WMBC) (n=690) and those with partial missing BIA data (PMBC) (n=76) are summarized in **supplementary Table 1**. **Supplementary Table 2** shows the baseline characteristics of patients with biopsy-proven NAFLD, stratified by NASH or non-NASH, both in the training and validation groups. Compared with those with non-NASH, patients with NASH differed in terms of age, BMI, percent

353 body fat, visceral fat area, fasting glucose, plasma lipid profile and serum
354 transaminases. Notably, LEARN algorithm, ION, HAIR, and NICE model, as well as
355 histological stages of fibrosis also significantly differed between NASH or non-
356 NASH patients, both in the training or validation groups.
357
358 **LEARN algorithm development**
359 As shown in **Figure 2**, we have developed a novel and automatic machine learning
360 model, called the LEARN algorithm, which included age, sex, prior hypertension,
361 prior diabetes, as well as six body composition parameters (namely arm
362 circumference, percent body fat, bone mineral content, basal metabolic rate, body cell
363 mass and visceral fat area). Biochemical parameters were not included in the LEARN
364 algorithm. The individual probability of having NASH was calculated via the LEARN
365 algorithm. For example, for a patient who has undergone BIA-based measurements of
366 body composition and who provides data on age, sex, and prior history of
367 hypertension or diabetes, it is possible to calculate her/his probability of having
368 NASH.
369
370 **Diagnostic performance of LEARN algorithm in the training and validation**
371 **groups**
372 The AUROC for LEARN algorithm in the training and validation groups were 0.81
373 (95%CI: 0.77-0.84) (**Figure 4a**), and 0.80 (95%CI: 0.73-0.87) (**Figure 4c**),
374 respectively. In both patient groups, the LEARN algorithm performed well for

diagnosing NASH. To more accurately identify NASH, we chose 0.492 (sensitivity = 0.90) and 0.531(specificity = 0.91), as dual cut-off values in the training group (**Figure 4b**). As shown in **Table 2**, when we chose these two cut-off values obtained by the LEARN algorithm, there was a NPV of 0.70 to rule out NASH and a PPV of 0.93 to rule in NASH in the training group, respectively. Similarly, in **Figure 4d** and **Table 2**, when we used the same dual cut-off values in the validation group, the cut-off values of 0.492 (sensitivity = 0.91) and 0.531 (specificity = 0.89) gave a NPV of 0.71 to rule out NASH and a PPV of 0.90 to rule in NASH in the training group, respectively. The diagnostic efficiency in the validation group also showed the same level of discrimination as that in the training group. Also, **Figure 3** shows the boxplots of the LEARN algorithm vs. histopathological severity of NAFLD in the training group. We observed that the prediction probability, as calculated by the LEARN algorithm, increased progressively with the histological severity of lobular inflammation, ballooning, steatosis and presence of definite NASH.

Subgroup analyses

We tested the diagnostic performance of the LEARN algorithm in different patient subgroups, in both training and validation groups. As shown in **Table 3,** the LEARN algorithm performed well in all subgroups both in the training group and in the validation group, regardless of sex, age, serum ALT levels and the presence or absence of liver fibrosis, hypertension, diabetes, or obesity (AUROCs ranging from 0.77 to 0.82). In particular, it should be noted that the LEARN algorithm performed

well among patients with or without liver fibrosis (AUROCs ranging from 0.77 to 0.83), in both the training and validation groups.

Diagnostic performance of the LEARN algorithm vs. other widely used non-invasive scores or biomarkers for NASH

As shown in **Table 4**, the LEARN algorithm showed a better diagnostic performance for identifying NASH (AUROC 0.80, 95%CI: 0.77-0.84) compared with other non-invasive NASH scores and the biomarker CK-18 M30. The AUROCs of these non-invasive scores or biomarkers were all less than 0.75 in the whole cohort; in particular, serum CK-18 M30 had an AUROC 0.73 (95%CI: 0.69-0.77); HAIR, 0.63 (95%CI: 0.59-0.67); ION, 0.67 (95%CI: 0.63-0.72); and the NICE model, 0.73 (95%CI: 0.69-0.77).

Diagnostic performance of the LEARN algorithm in the PMBC group

As shown in **supplementary Table 1**, there were 76 NAFLD patients with PMBC in the whole cohort. In order to improve the applicability of the LEARN algorithm in the real world where incomplete BIA data may occur, we adopted the strategy of replacing partially missing data of BIA values by mean values for the group. **Supplementary Table 3** shows that the diagnostic performance of the LEARN algorithm in the PMBC group, and the AUROC was 0.82 (95%CI: 0.72-0.92). As shown in **Supplementary Figure 1**, in the PMBC group, the LEARN algorithm showed a greater AUROC compared with ION and HAIR for predicting NASH. The

diagnostic performance of the LEARN algorithm in the PMBC group also performed well despite partially missing data.

Discussion

In this large cross-sectional multicenter study, we have developed a novel, fully automatic MLA, referred to as the LEARN algorithm (patent-pending, 2021110501603) to non-invasively diagnose NASH. For patients with biopsy-proven NAFLD who undergo impedance-based measures of body composition and provide simple information on age, sex, diabetes status and hypertension, it is possible to predict their probability of having NASH on histology with acceptable certainty. Our newly developed LEARN algorithm performed well in both the training and validation groups, and across a range of clinically relevant subgroups of patients. To our knowledge, this is the first multicenter study to develop a prediction model based on body composition, for non-invasively identifying NASH.

Besides clinical features of the metabolic syndrome (including hypertension and diabetes), there are other risk factors for a faster progression of NAFLD to NASH. A large number of studies suggested that body composition may be different in NAFLD from that in people without NAFLD, and that there is metabolic dysfunction in NAFLD.(16-19) Increased dietary calorie intake and lack of physical exercise may increase the amount of adipose tissue, and accumulation of fat mass may induce insulin resistance and exacerbate liver damage in NAFLD.(26-28) Otgonsuren et al.

showed that anthropometric measures, such as arm circumference and body fat percentage, were significantly higher in NAFLD than in non-steatotic controls.(29) Ko et al. found that ultrasound-detected NAFLD was associated with higher BMI, larger waist circumference, and greater body fat mass, through a large sample analysis involving 2,759 participants.(30) Idilman et al. showed that visceral adipose tissue alone could be a modest risk factor for predicting NASH (AUROC, 0.64).(31) In addition, arm circumference, percent body fat, BMI, waist circumference, visceral adipose tissue, skeletal muscle mass (sarcopenia) may be also risk factors for greater NAFLD severity.(32) Filip et al. reported that osteoporosis (as measured by bone mineral content) may also increase the risk of NAFLD.(33) In our study, the LEARN algorithm highlights the utility of body composition measurements for the diagnosis of NASH and this algorithm may help in reducing the number of unnecessary liver biopsies for diagnosing NASH. The value of impedance-based measurements of body composition may also be even greater if the full cost of liver biopsies is to be taken into account (allowing for biopsy-associated complications).

Prediction models lacking transparency and predictability have the potential to cause harm. Our research overcomes this shortcoming. Choosing machine-learning models with high transparency rather than black box models, with high decision-making risk is preferred. Our study uses the classical algorithm of deep learning to develop a new deep neural network for processing big data that includes impedance-based measurements of body composition. In the LEARN algorithm, residual networks were

used to prevent gradient disappearance and strengthen the ability of the deep neural network to extract features, thus improving the classification performance of the deep neural network; the dropout method was used to address the over-fitting issue and improve the generalization ability of the deep neural network; fully connected layers were added to analyze the relationship between features more clearly and intuitively, and reduce the influence of feature position on classification. Finally, we re-cycled the residual network to further improve the non-linear expression ability and complexity of the deep neural network. By following this design approach, the LEARN algorithm was optimized.

It is important to underline that in our study we included not only NAFLD patients with elevated serum transaminase levels, but also those without normal serum transaminases who had evidence of hepatic steatosis at recruitment as diagnosed by imaging techniques. For the LEARN algorithm, we used double cut-off points to identify NASH, as shown in **Figure 4** and **Table 2**. For the purpose of excluding NASH, a lower cut-off value was chosen. For diagnosing NASH, a higher cut-off point was selected. For the LEARN algorithm, the lower cut-off value of 0.492 showed a high sensitivity (90%) and a NPV of 0.70, while the upper cut-off value of 0.536 showed similar specificity (90%) and a PPV of 0.93 in the training and validation groups.

The choice of cut-off values conducive to optimum sensitivity or specificity depends

485 on the purpose of detection. Shown in **Table 2,** there is a gray zone of 39% when dual
486 cut-offs were used to identify NASH. However, it should be noted that there is always
487 a “gray zone” for all non-invasive tests that use two cutoff thresholds.(34) On the
488 other hand, approximately 60% of NAFLD patients were able to avoid liver biopsies
489 using our LEARN algorithm when dual cut-offs were chosen.
490
491 Currently, treatment of NASH is a major focus of drug development
492 worldwide.(35,36) Early, non-invasive identification of NASH for possible drug
493 treatment will be an important medical challenge in the next few years. However,
494 patients with NAFLD, especially those with normal serum ALT levels, and those who
495 are nonobese or do not have diabetes are often ignored in further assessment of
496 NAFLD severity. Therefore, we have also analyzed the diagnostic performance of our
497 newly proposed LEARN algorithm in identifying NASH in different patient
498 subgroups, stratified by obesity, diabetes or serum ALT levels (**Table 3**). Interestingly,
499 our LEARN algorithm performed well in the non-obese, non-diabetic, or serum
500 normal or abnormal ALT (ALT > 40 U/L) subgroups, in both the training and
501 validation groups. Importantly, NASH patients with or without fibrosis did not
502 influence the diagnostic performance of the "LEARN" algorithm. Both in the training
503 and the validation groups, the diagnostic performance of the "LEARN" algorithm was
504 above 0.75 among patients with or without liver fibrosis. In the whole cohort, we also
505 compared the diagnostic performance of LEARN algorithm, serum CK-18 M30 level,
506 HAIR, ION and NICE models in identifying NASH, and found that these latter non-

invasive scores had moderate accuracy in our cohort, although this finding might be partially affected by differences in the prevalence of NASH among different study populations.(24,25,37,38) As shown in **Table 4**, the diagnostic performance of the LEARN algorithm in identifying NASH had an AUROC of 0.80, which is significantly better than other non-invasive NASH scores mentioned above.

In our study, PASS15 was used to estimate the sample size. Each patient was examined for body composition by professionally trained personnel at each center in accordance with uniform operating instructions, and the data were extracted at each of the 6 participating sites by trained data collectors and compiled into spread sheets. Then an experienced AI team used neural network algorithms to build a prediction model that provided clinicians with an individual's probability of having NASH as described above.

Our BIA data were extracted at each of the 6 participating sites by trained data collectors and compiled into spread sheets. The data collection process was checked repeatedly, to reduce the chance of document errors. However, some of our patient's body composition examination reports were incomplete and reported between 12 and 19 measurements, which we referred to as subjects with 'partial missing BIA data'. We did not exclude these subjects from the analysis in order to improve the utility of our LEARN algorithm in the real world. In the process of building the LEARN algorithm, the artificial intelligence algorithm has used an average value to replace

missing data for those patients who had ‘partial missing BIA data’. Specially, we adopted the strategy of replacing PBMC group values by mean values, as this can improve the utility of the LEARN algorithm in the real world where missing data is relatively common. In the PMBC group, the diagnostic performance for identifying NASH performed well with an AUROC of 0.82.

There are some important limitations that should be mentioned. Firstly, the participants were all Chinese of Han ethnicity, so our results might not be applicable to other ethnic groups. Secondly, when comparing results from our cohort with other published studies that used non-invasive scores or biomarkers for diagnosing NASH, the heterogeneity between studies might at least in part contribute to the different diagnostic performances of these non-invasive tests for NASH (e.g., serum CK-18 M30 level, HAIR, ION, NICE models). Specially, the HAIR score system is a non-invasive score for predicting NASH based on hypertension, ALT levels and insulin resistance. When the score is ≥2, the AUROC for predicting NASH is 0.9, and the sensitivity and specificity are 80% and 89%, respectively. However, this model is currently only applicable to patients with BMI>35 kg/m^2.(23) So the applicability is not widespread. In our study, the Han- population was mainly included, and BMI generally concentrated in 24-29 kg/m^2.Therefore, HAIR had a low AUROC in this study. Finally, there was a “gray zone” and 39% patients couldn’t be identified with NASH or NAFL. This latter problem is a common limitation for all non-invasive tests where two cut-off thresholds are used.(39) In addition, a two-step approach has been

also recently reported. By using this two-step approach, patients in the “gray zone” were re-evaluated in combination with other non-invasive diagnostic tests and the need for liver biopsy was reduced significantly without much effect on the percentage of misclassifications.(40) In future studies, we will evaluate whether the combination of our LEARN algorithm with other non-invasive NASH scores contributes to the improved stratification of severity of NAFLD.

In conclusion, we have developed a fully automatic LEARN algorithm utilizing impedance-based measurements of body composition along with age, sex, and prior history of hypertension or diabetes, which shows good predictive ability for non-invasively identifying NASH in a large multi-center study across China. Our results suggest that routine measurement of body composition for the assessment of patients with NAFLD may be helpful in staging severity of liver disease and identification of NASH.

Reference

1. Younossi Z, Tacke F, Arrese M, et al. Global Perspectives on Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. Hepatology (Baltimore, Md) 2019;69:2672-82.
2. Chalasani N, Younossi Z, Lavine J, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology (Baltimore, Md) 2018;67:328-57.
3. Brunt E, Janney C, Di Bisceglie A, et al. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. The American journal of gastroenterology 1999;94:2467-74.
4. Diehl A, Day C. Cause, Pathogenesis, and Treatment of Nonalcoholic Steatohepatitis. The New England journal of medicine 2017;377:2063-72.
5. Alkhouri N, Tincopa M, Loomba R, et al. What Does the Future Hold for Patients With Nonalcoholic Steatohepatitis: Diagnostic Strategies and Treatment Options in 2021 and Beyond? Hepatology communications 2021.
6. Younossi Z, Stepanova M, Rafiq N, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. Hepatology (Baltimore, Md) 2011;53:1874-82.
7. Zhou Y, Zheng K, Targher G, et al. Non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis. The lancet Gastroenterology & hepatology 2021;6:9-10.
8. Zhou Y, Wong V, Zheng M. Consensus scoring systems for nonalcoholic fatty liver disease: an unmet clinical need. Hepatobiliary surgery and nutrition 2021;10:388-90.
9. Rios R, Zheng K, Targher G, et al. Non-invasive fibrosis assessment in non-alcoholic fatty liver disease. Chinese medical journal 2020;133:2743-5.
10. Obermeyer Z, Emanuel E. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. The New England journal of medicine 2016;375:1216-9.
11. Beam A, Kohane I. Big Data and Machine Learning in Health Care. JAMA 2018;319:1317-8.
12. Fralick M, Colak E, Mamdani M. Machine Learning in Medicine. The New England journal of medicine 2019;380:2588-9.
13. Chen K, Nie Y, Park S, et al. Development and Validation of Machine Learning-based Model for the Prediction of Malignancy in Multiple Pulmonary Nodules: Analysis from Multicentric Cohorts. Clinical cancer research : an official journal of the American Association for Cancer Research 2021;27:2255-65.
14. Feng G, Zheng K, Li Y, et al. Machine learning algorithm outperforms fibrosis markers in predicting significant fibrosis in biopsy-confirmed NAFLD. Journal of hepato-biliary-pancreatic sciences 2021;28:593-603.
15. Pietrobelli A, Rubiano F, St-Onge M, et al. New bioimpedance analysis system: improved phenotyping with whole-body analysis. European journal of clinical nutrition 2004;58:1479-84.
16. Ariya M, Koohpayeh F, Ghaemi A, et al. Assessment of the association between body composition and risk of non-alcoholic fatty liver. PloS one 2021;16:e0249223.
17. Miyake T, Miyazaki M, Yoshida O, et al. Relationship between body composition and the histology of non-alcoholic fatty liver disease: a cross-sectional study. BMC gastroenterology 2021;21:170.
18. Schmitz S, Schooren L, Kroh A, et al. Association of Body Composition and Sarcopenia with NASH in Obese Patients. Journal of clinical medicine 2021;10.

19. Samala N, Desai A, Vilar-Gomez E, et al. Decreased Quality of Life Is Significantly Associated With Body Composition in Patients With Nonalcoholic Fatty Liver Disease. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2020;18:2980-8.e4.
20. Li G, Rios R, Wang X-X, et al. Sex influences the association between appendicular skeletal muscle mass to visceral fat area ratio and nonalcoholic steatohepatitis in patients with biopsy-proven NAFLD. British Journal of Nutrition 2021.
21. Zhou Y, Ye F, Li Y, et al. Individualized risk prediction of significant fibrosis in non-alcoholic fatty liver disease using a novel nomogram. United European gastroenterology journal 2019;7:1124-34.
22. Newsome P, Sasso M, Deeks J, 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. The lancet Gastroenterology & hepatology 2020;5:362-73.
23. Dixon J, Bhathal P, O'Brien P. Nonalcoholic fatty liver disease: predictors of nonalcoholic steatohepatitis and liver fibrosis in the severely obese. Gastroenterology 2001;121:91-100.
24. Anty R, Iannelli A, Patouraux S, 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. Alimentary pharmacology & therapeutics 2010;32:1315-22.
25. Younes R, Rosso C, Petta S, et al. Usefulness of the index of NASH - ION for the diagnosis of steatohepatitis in patients with non-alcoholic fatty liver: An external validation study. Liver international : official journal of the International Association for the Study of the Liver 2018;38:715-23.
26. Després J, Lemieux I. Abdominal obesity and metabolic syndrome. Nature 2006;444:881-7.
27. Kuk J, Katzmarzyk P, Nichaman M, et al. Visceral fat is an independent predictor of all-cause mortality in men. Obesity (Silver Spring, Md) 2006;14:336-41.
28. Nobarani S, Alaei-Shahmiri F, Aghili R, et al. Visceral Adipose Tissue and Non-alcoholic Fatty Liver Disease in Patients with Type 2 Diabetes. Digestive diseases and sciences 2021.
29. Otgonsuren M, Stepanova M, Gerber L, et al. Anthropometric and clinical factors associated with mortality in subjects with nonalcoholic fatty liver disease. Digestive diseases and sciences 2013;58:1132-40.
30. Ko Y, Wong T, Hsu Y, et al. The Correlation Between Body Fat, Visceral Fat, and Nonalcoholic Fatty Liver Disease. Metabolic syndrome and related disorders 2017;15:304-11.
31. Idilman I, Low H, Gidener T, et al. Association between Visceral Adipose Tissue and Non-Alcoholic Steatohepatitis Histology in Patients with Known or Suspected Non-Alcoholic Fatty Liver Disease. Journal of clinical medicine 2021;10.
32. Habig G, Smaltz C, Halegoua-DeMarzio D. Presence and Implications of Sarcopenia in Non-alcoholic Steatohepatitis. Metabolites 2021;11.
33. Filip R, Radzki R, Bieńko M. Novel insights into the relationship between nonalcoholic fatty liver disease and osteoporosis. Clinical interventions in aging 2018;13:1879-91.
34. Boursier J, Guillaume M, Leroy V, et al. New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD. Journal of hepatology 2019;71:389-96.
35. Newsome P, Buchholtz K, Cusi K, et al. A Placebo-Controlled Trial of Subcutaneous

Semaglutide in Nonalcoholic Steatohepatitis. The New England journal of medicine 2021;384:1113-24.
36. Dufour J, Caussy C, Loomba R. Combination therapy for non-alcoholic steatohepatitis: rationale, opportunities and challenges. Gut 2020;69:1877-84.
37. Wieckowska A, Zein N, Yerian L, et al. In vivo assessment of liver cell apoptosis as a novel biomarker of disease severity in nonalcoholic fatty liver disease. Hepatology (Baltimore, Md) 2006;44:27-33.
38. Poynard T, Ratziu V, Charlotte F, et al. Diagnostic value of biochemical markers (NashTest) for the prediction of non alcoholo steato hepatitis in patients with non-alcoholic fatty liver disease. BMC gastroenterology 2006;6:34.
39. Zhou Y, Gao F, Liu W, et al. Screening for compensated advanced chronic liver disease using refined Baveno VI elastography cutoffs in Asian patients with nonalcoholic fatty liver disease. Alimentary pharmacology & therapeutics 2021;54:470-80.
40. Gao F, Huang J, Zheng K, et al. Development and validation of a novel non-invasive test for diagnosing fibrotic non-alcoholic steatohepatitis in patients with biopsy-proven non-alcoholic fatty liver disease. Journal of gastroenterology and hepatology 2020;35:1804-12.

692 **Table Legends**

693 **Table 1.** Baseline characteristics of patients with biopsy-proven NAFLD.

694 **Table 2.** Diagnostic performance of the LEARN algorithm.

695 **Table 3.** Diagnostic performance of LEARN algorithm in different patient subgroups.

696 **Table 4.** Pairwise comparisons between AUROCs for the LEARN algorithm and

697 other non-invasive NASH scores or biomarkers for identifying NASH.

698 **Supplementary Table 1.** Baseline characteristics of patients with or without partial

699 missing data of body composition.

700 **Supplementary Table 2.** Baseline characteristics of patients, stratified by NASH or

701 non-NASH on histology in the training or validation groups.

702 **Supplementary Table 3.** Diagnostic performance of LEARN algorithm in groups of

703 patients with or without partial missing data of body composition.

704

705

706

707

708

709

710

711

712

713

Figure Legends

Figure 1. Flowchart of the study.

Figure 2. Flowchart of the deep neural network algorithm for prediction of NASH (namely the LEARN algorithm).

Input layer: input the normalized data into this layer consisting of four modules. First, a Full Connected (FC) layer, to synthesize the features extracted from the previous section. Second, a Batch Normalization (BN) layer, to simplify the calculation and make the data retain its original expression ability as far as possible after the normalization processing. Third, Tanh function, a nonlinear function to help machines learn complex mappings. Last, the Dropout layer, reducing overfitting, extract a matrix containing 1536 features. Hidden layer: the matrix containing 1536 features is input into this layer, and the data needs to be looped four times through the residual module. Output layer: the output layer is consisted of a FC layer and a Softmax function. Through the Softmax function, we can map the output values to the interval (0, 1) for the final classification.

Figure 3. Boxplot of the LEARN algorithm versus histopathological severity of NAFLD in the training group: (a) steatosis grade, (b) ballooning grade, (c) lobular inflammation grade, and (d) presence of definite NASH.

Figure 4. Diagnostic performance of LEARN algorithm and sensitivity, specificity of the dual cut-off values in the training and validation groups. (a, b) training group;(c, d) validation group. *Abbreviations*: LEARN algorithm: deep neural network model for identifying nonalcoholic steatohepatitis.

Supplementary Figure 1. Pairwise comparison of ROC curves between the deep neural network model for identifying NASH (LEARN algorithm) and ION and HAIR models in the group of patients with partial missing data of body composition.