





ORIGINAL ARTICLE

A novel model for predicting fatty liver disease by means of an artificial neural network

Yi-Shu Chen¹, Dan Chen¹, Chao Shen², Ming Chen ³, Chao-Hui Jin³, Cheng-Fu Xu¹, Chao-Hui Yu¹ and You-Ming Li ^{1,*}

¹Department of Gastroenterology, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, P. R. China; ²Health Management Center, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, P. R. China; ³Hithink Royal Flush Information Network Co., Ltd, Hangzhou, Zhejiang, P. R. China

*Corresponding author. Department of Gastroenterology, The First Affiliated Hospital, College of Medicine, Zhejiang University, 79 Qingchun Road, Hangzhou, Zhejiang 310003, P. R. China. Tel: +86-571-87236861; Fax: +86-571-87236861; Email: zlym@zju.edu.cn

Abstract

Background: The artificial neural network (ANN) emerged recently as a potent diagnostic tool, especially for complicated systemic diseases. This study aimed to establish a diagnostic model for the recognition of fatty liver disease (FLD) by virtue of the ANN.

Methods: A total of 7,396 pairs of gender- and age-matched subjects who underwent health check-ups at the First Affiliated Hospital, College of Medicine, Zhejiang University (Hangzhou, China) were enrolled to establish the ANN model. Indices available in health check-up reports were utilized as potential input variables. The performance of our model was evaluated through a receiver-operating characteristic (ROC) curve analysis. Other outcome measures included diagnostic accuracy, sensitivity, specificity, Cohen's κ coefficient, Brier score, and Hosmer-Lemeshow test. The Fatty Liver Index (FLI) and the Hepatic Steatosis Index (HSI), retrained using our training-group data with its original designated input variables, were used as comparisons in the capability of FLD diagnosis.

Results: Eight variables (age, gender, body mass index, alanine aminotransferase, aspartate aminotransferase, uric acid, total triglyceride, and fasting plasma glucose) were eventually adopted as input nodes of the ANN model. By applying a cut-off point of 0.51, the area under ROC curves of our ANN model in predicting FLD in the testing group was 0.908 [95% confidence interval (CI), 0.901–0.915]—significantly higher ($P < 0.05$) than that of the FLI model (0.881, 95% CI, 0.872–0.891) and that of the HSI model (0.885; 95% CI, 0.877–0.893). Our ANN model exhibited higher diagnostic accuracy, better concordance with ultrasonography results, and superior capability of calibration than the FLI model and the HSI model.

Conclusions: Our ANN system showed good capability in the diagnosis of FLD. It is anticipated that our ANN model will be of both clinical and epidemiological use in the future.

Key words: artificial neural network; diagnostic model; fatty liver disease; Fatty Liver Index; Hepatic Steatosis Index; uric acid

Submitted: 28 October 2019; Revised: 31 January 2020; Accepted: 19 June 2020

© The Author(s) 2020. Published by Oxford University Press and Sixth Affiliated Hospital of Sun Yat-sen University

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Fatty liver disease (FLD), a leading cause of end-stage liver disease, hepatocellular carcinoma, and liver transplantation worldwide [1–3], is characterized by the accumulation of fat droplets in hepatocytes [4]. The disease encompasses a spectrum of liver pathology with different clinical prognoses, ranging from simple steatosis to steatohepatitis, fibrosis, and cirrhosis. Despite a lack of long-term prospective evidence, we do have some insights into the natural history of this disease [5]. Simple steatosis is at the most clinically benign extreme, with a low risk of developing cirrhosis. However, the risk increases as steatosis becomes complicated by histologically conspicuous hepatocyte death and inflammation, known as steatohepatitis. Such facts, along with the healthcare costs and declining health-related quality of life associated with FLD, make it a disease worth extensive attention by the common people [6] and one that welcomes lifestyle modification and medical intervention from the early stage [7].

However, how to recognize the ‘early stage’ remains a problem. Scarcely can suspicion be raised by clinical manifestations, since most patients are asymptomatic. Even symptomatic patients present unspecific complaints such as fatigue, abdominal discomfort, and, only seldomly, manifestations of advanced liver disease [8]. Patients can rarely know whether they have FLD or not except on occasions of routine health check-ups. Considering the fact that a blood test is barely absent in health check-ups in China and its pervasive, convenient nature compared with radiographic or invasive examinations, this study aimed to establish a model for the recognition of FLD using solely blood tests.

To begin with, as there is as yet no one widely accepted specific blood test for FLD, variables that have been proved or suspected to be associated with FLD were taken into consideration. Possible risk factors include increased body mass index (BMI), insulin resistance/type 2 diabetes mellitus (T2DM), and other parameters indicative of the metabolic syndrome (e.g. systemic hypertension, dyslipidemia, hyperuricemia/gout, cardiovascular disease) [2, 3]. On the other hand, a rapidly expanding body of clinical evidence supports the concept of FLD as a multisystem disease that affects extra-hepatic organs and regulatory pathways, increasing the risks of T2DM, cardiovascular and cardiac diseases, and chronic kidney disease [9]. The complex bidirectional relationship between FLD and the whole human-body system [10, 11] gives us the inspiration of seeking help from artificial intelligence, which may help us to make a good selection from candidate variables in an efficient and effective manner.

The artificial neural network (ANN) emerged in recent years as a potent diagnostic tool by virtue of its adaptability and excellent problem-solving-oriented architecture [12]. Like its biological counterpart, an ANN consists of a set of highly interconnected processing units (neurons) tied together with ‘weights (synapses),’ which indicate the strength of the connection [13]. The network usually consists of an input layer, an output layer, and one or more hidden layers [14]. During the training process, the association between the input and its corresponding output is explored by computer through the network, where the connection weights between the units are modified. As has been described by several studies [13, 15–18], the ANN is highly admired for its ability to learn through examples. The workhorse of learning in a neural network, the back-propagation algorithm [13, 19], helps to determine the weight between units. Initially, the weights are set randomly, under

which circumstance the output is far from ideal. The algorithm will make a comparison between the acquired output and the desired one, and generate an error value. The error value is then propagated backwards through the network, according to which the connection weights will be updated to make the model better agree with the ideal outcome. As learning proceeds, the overall error of the network decreases until a minimum is reached. In this way, the optimal network structure is established.

In this case, variables obtained from blood tests constituted the input layer and the output layer gave the diagnosis. The study aimed to assess the capability of our ANN model for the recognition of FLD on the strength of the blood-test variables. We also compared its performance with that of two other models: the Fatty Liver Index (FLI) and the Hepatic Steatosis Index (HSI), proposed by other researchers before.

Patients and methods

Subject inclusion

Participants aged between 18 and 70 years who underwent routine health check-ups at the First Affiliated Hospital, College of Medicine, Zhejiang University (Hangzhou, China) between January 2015 and February 2018 were retrospectively included in this study. A complete health check-up report required by our study, which included anthropometric assessment (height and body weight), laboratory results from blood samples, abdominal ultrasonography, and a summary of history taken by qualified doctors about personal information (age, gender, etc.). Participants with incomplete health check-up information were excluded.

According to the criteria proposed by the Chinese Liver Disease Association [20], the FLD diagnosis was made by the presence of at least two of the following three abnormal findings on abdominal ultrasonography: (i) diffusely increased liver near-field ultrasound echo (“bright liver”), liver echo greater than the kidney echo; (ii) vascular blurring; and (iii) gradual attenuation of far-field ultrasound echo. Participants whose ultrasonography results showed “mild fatty liver,” “heterogeneous fatty liver,” or “fatty liver tendency” were excluded.

Each subject diagnosed with FLD was randomly matched to a subject without FLD, who was of the same gender and the same age. A 1-year difference in age was acceptable. A total of 7,396 pairs of subjects with a mean age (standard deviation) of 49.35 (11.47) were then acquired (Supplementary Table 1). Those who failed to get into matches were excluded.

Data collection

Height and body weight were measured with regularly standardized digital scales. BMI was calculated using the formula: $BMI = \text{body weight (kg)} / \text{height squared (m}^2\text{)}$. Heart rate was measured at a resting state (staying still for at least 10 min).

Indices measured from blood samples are listed as the following: red blood cell (RBC), white blood cell (WBC), platelet (PLT), neutrophil (NEUT), eosinophil (EOS), monocyte (MO), lymphocyte (LY), hemoglobin (HGB), hematocrit (HCT), plateletcrit (PCT), fasting plasma glucose (FPG), total cholesterol (TC), total triglyceride (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), very-low-density lipoprotein cholesterol (VLDL-C), total protein (TP), albumin (ALB), globulin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl transpeptidase

(GGT), glycyproline dipeptidyl aminopeptidase (GPDA), alpha-L-fucosidase (AFU), creatinine (Cr), uric acid, and urea. All the biochemical examinations above were performed in the same laboratory using standard methods.

Abdominal ultrasound examination was conducted by well-trained, experienced doctors specializing in ultrasonography (with work experience of >5 years) who were blinded to clinical assessments and laboratory results.

All procedures performed in the study involving human participants were in accordance with the ethical standards of the Ethics Committee of the First Affiliated Hospital, College of Medicine, Zhejiang University and according to the 1964 Helsinki Declaration and its later amendments. All participants were informed of the possible research purpose of their health check-up reports and gave verbal consent to their anonymized health data being used in our ANN model.

ANN design and assessment

The 14,792 subjects enrolled in our study were randomly assigned to a training group ($n = 10,354$; 70%) or a testing group ($n = 4,438$; 30%). Baseline characteristics of the two groups were compared by Student's *t*-test. Within the training group, 7,396 subjects (50% of the entire study group) were randomly selected to train the network and the remaining 2,958 (20%) was used for cross validation. The training group was used to build our ANN model while the testing group was used to evaluate its diagnostic capability. The validation group was designed to prevent the network from being overtrained, which means deviation from the general predictive characteristics due to specific cases.

The performance of our network was evaluated through a receiver-operating characteristic (ROC) curve analysis. The area under the ROC curve (AUROC) with 95% confidence interval (CI) was calculated as a major indicator of the model's diagnostic performance in this binary classifier system.

Indices available in the health check-up reports were sorted out as potential input variables for our ANN model. Our selection of variables worked in a stepwise way. Assume that we have a model with N input variables (for short, we name it "Nmodel"). Then we randomly took away one input variable from it, producing N possibilities. In each case of possibility, the remaining ($N - 1$) input variables will be trained to create a new network model with its corresponding AUROC calculated. For the number N , we obtain N new network models with N values of AUROC. The model with the AUROC value most approximate to that of the Nmodel will be reserved; in other words, the corresponding input variable is then eliminated. The procedure will be repeated until we find a statistically significant difference between the newest Nmodel and the last one, which is the model we actually need. In this way, our ANN model was eventually established.

The FLI and the HSI were used as comparisons with our ANN model in capability of FLD diagnosis. Considering that the two models were derived from foreign populations, we retrained them using our training-group data with its original designated input variables before comparison.

Statistical analysis

Categorical variables are presented as numbers of subjects and percentages; continuous variables are presented as mean values and standard deviations. In particular, ALT, AST, GGT, TBA, TG, FPG, VLDL, and EOS showed a highly skewed distribution and were Log transformed (Log *e*) after which normal

distribution was achieved. All the variables were then normalized. Output values would range from 0 (FLD absent) to 1 (FLD present).

The relationship between variables and FLD diagnosis was explored through univariate analysis by the Wald test and confirmed by multivariate analysis through principal-component analysis. Variables that showed no statistically significant relationship were excluded, whereas the remaining ones were used as potential input nodes to build the ANN.

Comparison of the three ROC curves was conducted using the Hanley–McNeil method. Apart from AUROC, other evaluation indicators, including overall accuracy (correct predictions divided by total predictions), sensitivity, specificity, positive predictive value, and negative predictive value, were calculated for different cut-off points of outputs. The best cut-off point, determined by Youden's Index (sensitivity + specificity - 1), was used for classification in the testing group.

The Brier score—a function that measures the average squared deviation between predicted probabilities for events and their actual outcomes—was calculated in the testing group; a lower Brier score represents a higher accuracy.

Agreement between the predictions of the three models and ultrasonography results in the testing group was reported as Cohen's *k* coefficient using the formula: $[\text{Pr}(a) - \text{Pr}(e)]/[1 - \text{Pr}(e)]$, where $\text{Pr}(a)$ is the relative observed agreement and $\text{Pr}(e)$ is the proportion of agreement expected to occur by chance alone. Agreement is considered excellent if *k* is >0.80, good if *k* ranges from 0.60 to 0.80, fair if *k* ranges from 0.40 to 0.60, and poor if *k* is <0.40.

The degree of calibration was evaluated by the Hosmer-Lemeshow test; a lower Hosmer-Lemeshow statistic indicates a better calibration capability of the model.

All analyses were conducted through programs compiled on Python with its *scipy*, *sklearn* libraries by qualified technicians. A *P*-value of <0.05 was considered significant.

Results

Development of the ANN diagnostic model by the training group

Subjects who met the enrollment criteria stated above were randomly assigned to the training group (10,354 subjects) and the testing group (4,438 subjects). No significant difference was found between these two groups in terms of baseline characteristics (Table 1).

In the training group of 10,354 subjects, univariate analysis showed that 27 of our original 29 variables were significantly associated with FLD (Table 2), which was also confirmed by multivariate analysis. Though serum TG showed a *P*-value of 0.076 in multivariate analysis, given the commonly accepted importance of serum TG in FLD diagnosis, we did not exclude it. The 27 variables as well as age and gender were deemed as candidates for stepwise elimination as described above, after which 8 variables (age, gender, BMI, ALT, AST, uric acid, TG, and FPG) were eventually reserved to be nodes that constituted the ANN input layer.

The performance of our ANN model in predicting FLD in the training group was excellent, with an AUROC of 0.906 (95% CI, 0.902–0.911)—significantly higher ($P < 0.05$) than that of the FLI model (0.871; 95% CI, 0.864–0.878) and that of the HSI model (0.876; 95% CI, 0.871–0.881).

Diagnostic-accuracy details of the ANN model for different cut-off points in the training group are outlined in Table 3. The

Table 1. Baseline characteristics of the study population stratified by ANN groups

Variable	Training group (n = 10,354)	Testing group (n = 4,438)	P-value
Heart rate (/min)	76.32 ± 11.09	76.49 ± 11.05	0.4
BMI (kg/m ²)	24.83 ± 3.33	24.77 ± 3.32	0.293
TP (g/L)	73.53 ± 3.90	73.60 ± 3.80	0.644
ALB (g/L)	47.56 ± 2.69	47.58 ± 2.65	0.524
Globulin (g/L)	25.97 ± 3.23	25.98 ± 3.24	0.987
ALT (IU/L)	27.81 ± 22.98	27.49 ± 21.82	0.43
AST (IU/L)	23.88 ± 12.41	23.74 ± 11.38	0.506
GGT (IU/L)	42.42 ± 55.34	42.07 ± 49.71	0.72
Cr (μmol/L)	75.41 ± 14.92	75.10 ± 14.52	0.24
Urea (mmol/L)	5.45 ± 1.24	5.43 ± 1.22	0.229
Uric acid (μmol/L)	356.84 ± 87.56	357.29 ± 86.99	0.778
TG (mmol/L)	1.84 ± 1.50	1.84 ± 1.48	0.783
TC (mmol/L)	4.84 ± 0.91	4.83 ± 0.92	0.444
HDL-C (mmol/L)	1.24 ± 0.34	1.23 ± 0.34	0.672
LDL-C (mmol/L)	2.81 ± 0.76	2.81 ± 0.78	0.856
VLDL-C (mmol/L)	0.80 ± 0.56	0.79 ± 0.54	0.441
FPG (mmol/L)	5.25 ± 1.33	5.23 ± 1.27	0.495
AFU (IU/L)	28.65 ± 7.64	28.59 ± 7.58	0.691
GPDA (IU/L)	77.00 ± 17.56	76.38 ± 17.58	0.051
WBC (×10 ⁹ /L)	6.25 ± 1.56	6.25 ± 1.65	0.944
NEUT (×10 ⁸ /L)	56.54 ± 8.13	56.43 ± 8.04	0.469
LY (×10 ⁹ /L)	34.18 ± 7.56	34.27 ± 7.49	0.469
MO (×10 ⁸ /L)	6.45 ± 1.76	6.45 ± 1.78	0.813
EOS (×10 ⁸ /L)	2.40 ± 1.88	2.41 ± 1.92	0.848
HGB (g/L)	150.23 ± 14.60	150.34 ± 14.65	0.676
PLT (×10 ⁹ /L)	221.10 ± 53.07	221.19 ± 54.59	0.93
RBC (×10 ¹² /L)	4.95 ± 0.46	4.96 ± 0.46	0.601
HCT (%)	44.74 ± 3.86	44.77 ± 3.88	0.669
PCT (%)	0.24 ± 0.05	0.24 ± 0.05	0.825
Age (years)	49.40 ± 11.47	49.22 ± 11.23	0.357
Male gender	7,403 (71.50%)	3,178 (71.6%)	0.923

Continuous variables are presented as mean values and standard deviations. Categorical variables are presented as numbers of subjects and percentages.

ANN, Artificial Neural Network; BMI, body mass index; TP, total protein; ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transpeptidase; Cr, creatinine; TG, total triglyceride; TC, total cholesterol; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; VLDL-C, very-low-density lipoprotein cholesterol; FPG, fasting plasma glucose; AFU, alpha-L-fucosidase; GPDA, glycyproline dipeptidyl aminopeptidase; WBC, white blood cell; NEUT, neutrophil; LY, lymphocyte; MO, monocyte; EOS, eosinophil; HGB, hemoglobin; PLT, platelet; RBC, red blood cell; HCT, hematocrit; PCT, plateletcrit.

best cut-off point turned out to be 0.51 according to Youden's Index. Applying a cut-off point of 0.51, the ANN prediction of FLD showed a sensitivity of 83.3% and a specificity of 81.4%. Likewise, the cut-off points for FLI and HSI were also determined, as 45 and 30, respectively.

Performance of the ANN model in FLD diagnosis in the testing group

The ANN model, as well as the other two retrained linear models, was finally evaluated on the testing group of 4,438 patients for the diagnosis of FLD. AUROC with 95% CI and other outcome measures were calculated (Table 4). The AUROC of our ANN model in predicting FLD in the testing group was 0.908 (95% CI, 0.901–0.915)—significantly higher ($P < 0.05$) than that of the FLI model (0.881; 95% CI, 0.872–0.891) and that of the HSI model

(0.885; 95% CI, 0.877–0.893). The sensitivity and specificity of ANN prediction in the testing group were 83.7% and 80.4%, respectively—both higher than those of the FLI model (81.2% and 78.2%, respectively) and the HSI model (81.7% and 78.6%, respectively). Particularly, when we tried to eliminate uric acid from our eight-variable model, the AUROC dropped to 0.901 (95% CI, 0.897–0.906)—significantly lower ($P < 0.05$) than that of the eight-variable one, which might illustrate the importance of uric acid in predicting FLD.

The degree of concordance between ANN, FLI, HSI predictions, and the ultrasonography results are shown in Table 5. The ANN model correctly identified 82.1% of the subjects; the k -statistic was 0.642, reflecting good agreement. The FLI model showed a lower accuracy (79.6%); the k -statistic was 0.592, indicating fair agreement. The HSI model showed a relatively higher accuracy of 80.2%; the k -statistic was 0.604. Additionally, the Brier score of the ANN model was 0.107—lower than that of the FLI model (0.118) and the HSI model (0.114), suggesting a higher diagnostic accuracy of the ANN in predicting FLD.

The Hosmer-Lemeshow test was performed to analyse the degree of calibration. The ANN model showed a relatively lower value of 4.85, whereas the values of the FLI model and the HSI model were 5.07 and 5.01, respectively, indicating a better calibration capability of the ANN model.

Discussion

Recent studies have demonstrated that ANN analysis is potentially superior to traditional statistical approaches, especially when the function of given variables remains unknown or when the impact of a variable is influenced by other variables in a complex multidimensional system [17, 19, 21]. Given FLD is a complex systemic disease, where the ANN could give full play to its unique strengths, we established a diagnostic model based on the risk factors of FLD with the aid of the ANN.

After a deliberate training process, eight variables (age, gender, BMI, ALT, AST, uric acid, TG, and FPG) were eventually adopted as the input nodes of our ANN diagnostic model. Our ANN model exhibited a good performance in predicting FLD in the testing group, as illustrated by high AUROC (0.908; 95% CI, 0.901–0.915), high accuracy (82.1%; Brier score 0.1073), good concordance (k -statistic 0.642), and good calibration (Hosmer-Lemeshow statistic 4.85).

The utilization of FLD risk factors has been pursued by clinicians and various statistical models have been established to predict FLD following this line of reasoning. The FLI model and the HSI model, two among the well-known FLD models, were retrained using our data for comparison with the ANN model.

Bedogni et al. [22] first proposed FLI in 2006. Derived from an Italian population, the index, ranging from 0 to 100, is calculated through an algorithm incorporating BMI, waist circumference, TG, and GGT. The model showed an AUROC of 0.84 (95% CI, 0.81–0.87) in detecting FLD in its original study and 0.881 (95% CI, 0.872–0.891) in our study; it has been validated in several other populations [23–26]. Lee et al. [27] proposed HSI in 2010 from a Korean population. The index is calculated through an algorithm taking in BMI, ALT/AST ratio, and the presence/absence of diabetes mellitus. When being tested in the original study, HSI had an AUROC of 0.812 (95% CI, 0.801–0.824), while, in our study, the AUROC was 0.885 (95% CI, 0.877–0.893). Compared with the two previous FLD models based on linear statistical analysis, our ANN model showed higher AUROC, better diagnostic accuracy, greater concordance, and superior capacity of calibration in the testing group.

Table 2. Relationship between variables and FLD diagnosis by univariate and multivariate analysis in the training group

Variable	Training group (n = 10,354)	P-value by univariate analysis	P-value by multivariate analysis
Heart rate (/min)	76.32 ± 11.09	<0.001	<0.001
BMI (kg/m ²)	24.83 ± 3.33	<0.001	<0.001
TP (g/L)	73.53 ± 3.90	<0.001	<0.001
ALB (g/L)	47.56 ± 2.69	<0.001	<0.001
Globulin (g/L)	25.97 ± 3.23	<0.001	<0.001
ALT (IU/L)	27.81 ± 22.98	<0.001	<0.001
AST (IU/L)	23.88 ± 12.41	<0.001	<0.001
GGT (IU/L)	42.42 ± 55.34	<0.001	<0.001
Cr (μmol/L)	75.41 ± 14.92	0.224	0.342
Urea (mmol/L)	5.45 ± 1.24	0.659	0.564
Uric acid (μmol/L)	356.84 ± 87.56	<0.001	<0.001
TG (mmol/L)	1.84 ± 1.50	<0.001	0.076
TC (mmol/L)	4.84 ± 0.91	<0.001	0.013
HDL-C (mmol/L)	1.24 ± 0.34	<0.001	<0.001
LDL-C (mmol/L)	2.81 ± 0.76	<0.001	<0.001
VLDL-C (mmol/L)	0.80 ± 0.56	<0.001	<0.001
FPG (mmol/L)	5.25 ± 1.33	<0.001	<0.001
AFU (IU/L)	28.65 ± 7.64	<0.001	<0.001
GPDA (IU/L)	77.00 ± 17.56	<0.001	<0.001
WBC (×10 ⁹ /L)	6.25 ± 1.56	<0.001	<0.001
NEUT (×10 ⁸ /L)	56.54 ± 8.13	0.003	0.005
LY (×10 ⁸ /L)	34.18 ± 7.56	<0.001	<0.001
MO (×10 ⁸ /L)	6.45 ± 1.76	<0.001	<0.001
EOS (×10 ⁸ /L)	2.40 ± 1.88	0.010	0.004
HGB (g/L)	150.23 ± 14.60	<0.001	<0.001
PLT (×10 ⁹ /L)	221.10 ± 53.07	<0.001	<0.001
RBC (×10 ¹² /L)	4.95 ± 0.46	<0.001	<0.001
HCT (%)	44.74 ± 3.86	<0.001	<0.001
PCT (%)	0.24 ± 0.05	<0.001	<0.001

FLD, fatty liver disease; BMI, body mass index; TP, total protein; ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transpeptidase; Cr, creatinine; TG, total triglyceride; TC, total cholesterol; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; VLDL-C, very-low-density lipoprotein cholesterol; FPG, fasting plasma glucose; AFU, alpha-L-fucosidase; GPDA, glycyproline dipeptidyl aminopeptidase; WBC, white blood cell; NEUT, neutrophil; LY, lymphocyte; MO, monocyte; EOS, eosinophil; HGB, hemoglobin; PLT, platelet; RBC, red blood cell; HCT, hematocrit; PCT, plateletcrit.

Table 3. Diagnostic accuracy at different cut-off points of ANN output in the training group

ANN output cut-off point	Sensitivity	Specificity	PPV	NPV	Accuracy
0.1	0.985	0.411	0.626	0.964	0.698
0.2	0.965	0.553	0.683	0.94	0.759
0.3	0.934	0.66	0.733	0.909	0.797
0.4	0.893	0.734	0.77	0.873	0.813
0.5	0.839	0.808	0.814	0.834	0.823
0.6	0.768	0.868	0.853	0.789	0.818
0.7	0.672	0.92	0.893	0.737	0.796
0.8	0.543	0.954	0.922	0.676	0.749
0.9	0.348	0.983	0.954	0.601	0.666

The HSI study recruited a total number of 10,724 individuals—more than 20 times that of the FLI study. In general, a larger cohort study tends to formulate a more convincing model. In our study, we recruited even more participants, which helped us to obtain a model with better performance. Of note, in our model, a relatively new variable, namely serum uric acid, was included as one of the input nodes in an attempt to improve the diagnostic efficacy. As researchers probe further into the relationship between FLD and metabolic disorders, the value of uric acid

Table 4. Performance of ANN, FLI, and HSI in terms of AUROC with 95% CI in both the training group and the testing group

Model	AUROC	95% CI
Training group		
ANN model	0.906	0.902–0.911
FLI model	0.871*	0.864–0.878
HSI model	0.876*	0.871–0.881
Testing group		
ANN model	0.908	0.900–0.915
FLI model	0.881*	0.872–0.891
HSI model	0.885*	0.877–0.893

*Significant difference compared with ANN model.

as a predictor of FLD has come into the spotlight. A large cohort study of 8,925 participants previously conducted by our research group clarified that an elevated serum uric acid level could be an independent risk factor for FLD [28]. The association has also been observed by large population-based cross-sectional studies conducted in Western populations [29, 30]. Though the underlying mechanisms stay unclear and await further research [31, 32], the variable is anticipated to be of potential diagnostic and therapeutic value in the future [33, 34].

Table 5. Concordance between predictions of FLD based on ANN, FLI, and HSI vs ultrasonography in the testing group

Model	Participants undergoing ultrasonography			Accuracy	Cohen's k coefficient
	FLD present	FLD absent	Total		
ANN					
FLD present	1,857	435	2,292	0.821	0.642
FLD absent	361	1,783	2,144		
Total	2,218	2,218	4,436		
FLI					
FLD present	1,255	363	1,618	0.796	0.592
FLD absent	291	1,302	1,593		
Total	1,546	1,665	3,211		
HSI					
FLD present	1,813	474	2,287	0.802	0.604
FLD absent	405	1,744	2,149		
Total	2,218	2,218	4,436		

For ANN, a cut-off of 0.51 was applied for classification; for FLI, a cut-off of 45 was applied; for HSI, a cut-off of 30 was applied.
ANN: artificial neural network; FLI: Fatty Liver Index; HSI: Hepatic Steatosis Index.

Potential clinical uses of our ANN model include the selection of subjects for further examination and the identification of patients for lifestyle counseling. Recognizing underlying chronic disease and promoting a healthier lifestyle constitute a pivotal component of the 'Healthy China 2030' blueprint proposed by the Chinese government [35]. Our ANN model, based on potential risks, may provide hospitals with an effective and economical strategy to detect FLD from mere blood tests—an ordinary and convenient part of routine health check-ups. Furthermore, from the viewpoint of epidemiological research, our model can be used to select subjects at a greater risk of FLD for the design of observational or interventional studies [22].

There are some limitations in the present study. First, a potential criticism could be the use of ultrasonography as our standard for FLD diagnosis. As the most common choice in clinical practice for the diagnosis of hepatic steatosis, ultrasonography is non-invasive, safe, widely available, inexpensive, sensitive ($\leq 94\%$), and specific ($\leq 95\%$) [8, 36] in detecting fatty liver as demonstrated in previous studies. However, it cannot perform ideally when fatty infiltration is below a threshold of 30% [8, 37], which means that an undefined number of FLD cases might be missed in our study. Second, histological features, which are closely associated with disease progression, such as inflammation and fibrosis, are not discussed in our study. Considering our data resource as well as the invasive nature and potential risks of liver biopsy, it is neither feasible nor ethically reasonable to ask for liver biopsy from a health check-up population [38]. Though patients with steatohepatitis require a closer follow-up due to their worse prognostic implications, all FLD patients should receive interventions in lifestyles and corrections in metabolic disturbance. Third, our ANN model was built and tested on an internal cohort and it could thus be argued that data from other populations might lead to a decrease in its diagnostic ability. Nonetheless, as has been described above, we believe that the distinctive learning ability of the ANN will make it feasible to give a diagnosis on data sets that it has never seen before [13]. We sincerely welcome any further validation of this model from external cohorts.

In conclusion, the ANN helped us to present an effective diagnostic model for FLD, based on easily obtainable clinical data. The ANN is superior to conventional statistical linear approaches and it could be of both clinical and research value in

tackling the global health problem of FLD. The performance of the ANN could be further improved by including new cases from other populations [13].

Supplementary data

Supplementary data is available at *Gastroenterology Report* online.

Authors' contributions

C.Y.S., C.D., S.C., C.M., J.C.H., X.C.F., Y.C.H., and L.Y.M. contributed to this work. In detail, C.Y.S., C.M., J.C.H., and X.C.F. analysed the data. C.Y.S. and X.C.F. wrote the paper. C.D. and S.C. provided the data. Y.C.H. and L.Y.M. critically revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key R&D Program of China [2017YFC0908900].

Conflicts of interest

The authors declared that they have no conflicts of interest in this work.

References

- Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: clinical prediction rules and blood-based biomarkers. *J Hepatol* 2018;**68**:305–15.
- Younossi Z, Henry L. Contribution of alcoholic and nonalcoholic fatty liver disease to the burden of liver-related morbidity and mortality. *Gastroenterology* 2016;**150**:1778–85.
- Younossi ZM, Blissett D, Blissett R et al. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatology* 2016;**64**:1577–86.
- Sertoglu E, Ercin CN, Celebi G et al. The relationship of serum uric acid with non-alcoholic fatty liver disease. *Clin Biochem* 2014;**47**:383–8.

5. Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. *Hepatology* 2010;51:1820–32.
6. Perumpail BJ, Khan MA, Yoo ER et al. Clinical epidemiology and disease burden of nonalcoholic fatty liver disease. *World J Gastroenterol* 2017;23:8263–76.
7. Younes R, Bugianesi E. Should we undertake surveillance for HCC in patients with NAFLD? *J Hepatol* 2018;68:326–34.
8. Machado MV, Cortez-Pinto H. Non-invasive diagnosis of non-alcoholic fatty liver disease: a critical appraisal. *J Hepatol* 2013; 58:1007–19.
9. Byrne CD, Targher G. NAFLD: a multisystem disease. *J Hepatol* 2015;62:S47–64.
10. Chalasani N, Younossi Z, Lavine JE et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2018;67:328–57.
11. Younossi ZM, Koenig AB, Abdelatif D et al. Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73–84.
12. Strebba CT, Ionescu M, Gheonea DI et al. Contrast-enhanced ultrasonography parameters in neural network diagnosis of liver tumors. *World J Gastroenterol* 2012;18:4427–34.
13. Cucchetti A, Piscaglia F, Grigioni AD et al. Preoperative prediction of hepatocellular carcinoma tumour grade and microvascular invasion by means of artificial neural network: a pilot study. *J Hepatol* 2010;52:880–8.
14. Amato F, López A, Peña-Méndez EM et al. Artificial neural networks in medical diagnosis. *J Appl Biomed* 2013;11:47–58.
15. Komeda Y, Handa H, Watanabe T et al. Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience. *Oncology* 2017;93:30–4.
16. Grossi E, Mancini A, Buscema M. International experience on the use of artificial neural networks in gastroenterology. *Dig Liver Dis* 2007;39:278–85.
17. Cucchetti A, Vivarelli M, Heaton ND et al. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. *Gut* 2007;56:253–8.
18. Abedi V, Goyal N, Tsvigoulis G et al. Novel screening tool for stroke using artificial neural network. *Stroke* 2017;48:1678–81.
19. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;346:1075–9.
20. Fan JG, Jia JD, Li YM et al. Guidelines for the diagnosis and management of nonalcoholic fatty liver disease: update 2010: (published in Chinese on Chinese Journal of Hepatology 2010; 18:163–166). *J Dig Dis* 2011;12:38–44.
21. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995;346:1135–8.
22. Bedogni G, Bellentani S, Miglioli L et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol* 2006;6:33.
23. Calori G, Lattuada G, Ragona F et al. Fatty liver index and mortality: The Cremona study in the 15th year of follow-up. *Hepatology* 2011;54:145–52.
24. Kim JH, Kwon SY, Lee SW et al. Validation of fatty liver index and lipid accumulation product for predicting fatty liver in Korean population. *Liver Int* 2011;31:1600–1.
25. Zelber-Sagi S, Webb M, Assy N et al. Comparison of fatty liver index with noninvasive methods for steatosis detection and quantification. *World J Gastroenterol* 2013;19:57–64.
26. Ruhl CE, Everhart JE. Fatty liver indices in the multiethnic United States National Health and Nutrition Examination Survey. *Aliment Pharmacol Ther* 2015;41:65–76.
27. Lee J-H, Kim D, Kim HJ et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis* 2010;42:503–8.
28. Li Y, Xu C, Yu C et al. Association of serum uric acid level with non-alcoholic fatty liver disease: a cross-sectional study. *J Hepatol* 2009;50:1029–34.
29. Shih MH, Lazo M, Liu SH et al. Association between serum uric acid and nonalcoholic fatty liver disease in the US population. *J Formos Med Assoc* 2015;114:314–20.
30. Sirota JC, McFann K, Targher G et al. Elevated serum uric acid levels are associated with non-alcoholic fatty liver disease independently of metabolic syndrome features in the United States: liver ultrasound data from the National Health and Nutrition Examination Survey. *Metabolism* 2013; 62:392–9.
31. Xu C. Hyperuricemia and nonalcoholic fatty liver disease: from bedside to bench and back. *Hepatol Int* 2016;10: 286–93.
32. Xu CF, Yu CH, Xu L et al. Hypouricemic therapy: a novel potential therapeutic option for nonalcoholic fatty liver disease. *Hepatology* 2010;52:1865–6.
33. Xu CF, Yu CH, Ma H et al. Prevalence and risk factors for the development of nonalcoholic fatty liver disease in a Nonobese Chinese Population: the Zhejiang Zhenhai Study. *Am J Gastroenterol* 2013;108:1299–304.
34. Ryu S, Chang Y, Kim SG et al. Serum uric acid levels predict incident nonalcoholic fatty liver disease in healthy Korean men. *Metabolism* 2011;60:860–6.
35. Fan JG, Kim SU, Wong V. New trends on obesity and NAFLD in Asia. *J Hepatol* 2017;67:862–73.
36. Schwenzer NF, Springer F, Schraml C et al. Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance. *J Hepatol* 2009;51:433–45.
37. Saadeh S, Younossi ZM, Remer EM et al. The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology* 2002;123:745–50.
38. Bedogni G, Miglioli L, Masutti F et al. Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. *Hepatology* 2005;42:44–52.